imtc14

JUNE 3 & 4, 2014

International Multi Track Conference on Sciences, Engineering & Technical Innovations

VOL - 1

Edited By
Dr Manoj Kumar

SPONSORED BY

Goverment of India
Department of Science & Technology
Ministry of Science & Technology

Punjab Technical University
Jalandhar

ORGANISED BY

CT GROUP OF INSTITUTIONS
www.ctgroup.in

CT INSTITUTE OF ENGINEERING MANAGEMENT & TECHNOLOGY

U.E.II-Pratappura Road, Shahpur, Jalandhar -144020 Punjab, INDIA
Ph. : 0181- 5055127, 5055128, Fax : 0181-5055130 I M. +91 9316326906
Conference Website:www.imtc.ctgroup.co.in I Email ID: imtc14@ctgroup.in

Proceedings of

International Multi Track Conference on Sciences, Engineering & Technical Innovations

imtc14

JUNE 3 & 4, 2014

Edited By

Dr Manoj Kumar

Organised By

CT INSTITUTE OF ENGINEERING MANAGEMENT & TECHNOLOGY

JALANDHAR - 144020 (PUNJAB)

Sponsored By

Government of India
Department of Science & Technology
Ministry of Science & Technology

PUNJAB TECHNICAL UNIVERSITY
JALANDHAR

Published by

CT EDUCATIONAL SOCIETY

CT INSTITUTE OF ENGINEERING MANAGEMENT & TECHNOLOGY

U.E.II-Pratappura Road, Shahpur, Jalandhar -144020 Punjab, INDIA

First Impression : 2014

Sciences, Engineering & Technical Innvations

ISBN : 978-81-929077-0-3

DISCLAIMER

The authors are solely responsible for the contents of the papers compiled in this volume. The publishers or editors do not take any responsibility for the same in any manner.

Published By

CT EDUCATIONAL SOCIETY
CT INSTITUTE OF ENGINEERING
MANAGEMENT & TECHNOLOGY

ISBN : 978-81-929077-0-3

Typset by

Aryana Software Develoment Cell
C/o. CT Institutions
U.E.II-Pratappura Road, Shahpur,
Jalandhar -144020 Punjab, INDIA

Printed by

AN Enterprises
198-fateh pura,
Behind singh steel
Jalandhar

PREFACE

This book contains the proceedings of the International Multi Track Conference on Sciences, Engineering & Technical Innovations (IMTC-14) which was organized by CT Institute of Engineering, Management & Technology (CTIEMT), Jalandhar on June 3-4, 2014.This conference aims to disseminate the latest research in Wireless Networks & Mobile Computing, Optical Communication, Software Engineering/Cloud Computing/Biomedical Signal Processing, Image/Speech Processing, Neural Network/Fuzzy Logic, Analog/Digital/VLSI/Antenna, Civil Engineering, Electrical Engineering, Mechanical Engineering, Pharmaceutical & Biotechnology, Applied Sciences, Management & Education and other relevant topics and applications. This conference is co-sponsored by Department of Science and Technology, India and Punjab Technical University, Jalandhar.

The friendliness and candidness of IMTC-14 shall enhance the ability of all the delegates to grow by interacting with the experienced resource persons and young researchers.

My gratitude is due to S. Charanjit Singh Channi, Chairman, CT Group of Institutions Jalandhar, who has been really very encouraging and supportive.

This conference and the proceedings have been possible due to the unstinted support of the Management of CT group.

I am very thankful to my colleagues and whole staff rendering their services as co-editors in compilation of this volume. Finally, I very cordially thank all the people of CT group for their efforts to maintain the high scientific level of conference and proceedings.

Dr. Manoj Kumar
Convener
imtc14

Conference Committee

Chief Patron

- **Dr. Rajneesh Arora,** Vice Chancellor, Punjab Technical University

Conference Host

- **S. Charanjit Singh Channi,** Chairman, CT Group of Institutions

Conference Convener(s)

- **Dr. Manoj Kumar, Group Director**, CT Group of Institutions

Conference Co-Convener(s)

- **Dr. Anil Sharma**
- **Dr. Rahul Malhotra**
- **Mr. Pankaj Thakur**
- **Dr. Manju Bala**
- **Mr. Vikrant Sharma**

Conference Secretaries

- **Dr. Sumit Jain**
- **Dr. Harjit Pal Singh**
- **Dr. Vaneet Thakur**
- **Mr. Anurag Sharma**
- **Mr. S. S. Matharu**
- **Ms. Kamalpreet Kaur**
- **Dr. Neha Sharma**
- **Dr. Rajeev Kharb**

Conference Organizing Committee

- **Mr. S.S. Matharu**
- **Ms. Surbhi Sharma**
- **Ms. Isha Garg**
- **Mr. Jagdeep**
- **Mr. Gurjeet Singh**
- **Ms. Sukhvir Kaur**
- **Ms. Shivani Khurana**
- **Ms. Sukhvir Kaur**
- **Mr. Prince Verma**

Publications Committee

- **Dr. Manju Bala**
- **Mr. Navdeep Singh Thind**
- **Dr. Harjit Pal Singh**
- **Mr. Anurag Sharma**

Head, Media and Publicity

- **Mr. Abhishek Soni**
- **Mr. Gaurav Khanna**
- **Mr. Pratik Mahendru**

Sponsorship & Treasurer

- **Dr. Anil Sharma**
- **Mr. Vikrant Sharma**
- **Mr. Vishav Kapoor**

Board & Lodging

- **Dr. Rajeev Kharb**
- **Mr. Gurpreet Singh Virk**
- **Er. Amandeep Singh Sehmby**

Transport Committee

- **Mr. Kuldeep Singh**
- **Mr. Sangram Singh,**

Registration Committee

- **Mr. Varun Sharma**
- **Mr. Sangram Singh**
- **Ms. Navneet Gil**

Hospitality Committee

- **Mr. Rohit Sarin**
- **Mr. Rahul Sharma**

Contents

Preface

Committee

Track-1 Technical Session 1 (A & B):
WIRELESS NETWORK & MOBILE COMPUTING

Track-1 Technical Session 2: OPTICAL COMMUNICATION

Track-1 Technical Session 3-(A & B): SOFTWARE ENGINEERING/CLOUD COMPUTING/BIOMEDICAL SIGNAL PROCESSING

Track-1 Technical Session 4-(A & B): IMAGE/SPEECH PROCESSING

Track-1 Technical Session 5: NEURAL NETWORK/FUZZY LOGIC

Track-1 Technical Session 6: ANALOG/DIGITAL/VLSI/ANTENNA

Track-1

Technical Session 1-(A & B)

WIRELESS NETWORK & MOBILE COMPUTING

PAPR Reduction Using the Huffman Coding Combined with the Modified Companding Function Technique for OFDM Systems

Vinit Bansal
UCOE, Department of Electronics and Communication
Punjabi University, Patiala (Punjab),India
vinitbansal5@gmail.com

Manjeet Singh Patterh
Department of Electronics and Communication
Punjabi University, Patiala(Punjab),India

Abstract–**Orthogonal frequency division multiplexing has some disadvantage. One of them is its high PAPR (peak to average power ratio).The high value of PAPR degrades the performance when non linear power amplifiers are employed. This paper proposes a technique to reduce PAPR using Huffman coding combined with modified Companding function.Via computer simulations,It is verified that the proposed technique has better PAPR reduction than using them separately.We compare the proposed technique to the previously suggested original companding method.Using the propose technique about 9 dB PAPR reduction is achieved.Simulation result also show that the Power spectral density(PSD) of the Modified Companding function is better than the original companding function.**

Keywords—Modified Companding Function, Huffman Coding, Peak-to-average power ratio (PAPR), Orthogonal frequency division multiplexing (OFDM), Power Spectral density.

I. INTRODUCTION

Orthogonal Frequency division multiplexing is a multicarrier modulation scheme which can be used to transmit the broadband data over wireless channel.In OFDM,the incoming bit stream is divided into the parallel bit stream,lower rate substreams and then transmit using orthogonal sub carriers[1].So,the bandwidth of entire channel is divided between the orthogonal subcarriers,so that the bandwidth of each subcarrier is much smaller than the channel coherence bandwidth and hence multipath fading is reduced.So,OFDM is more bandwidth efficient system than single carrier conventional system[2].OFDM is used in many wireless standards such as Terrestrial digital video broadcasting (DVB-T),digital audio broadcasting (DAB), wireless local area network standards, such as, IEEE 802.11a, ETSI HYPERLAN/2, wireless metropolitan area network standard IEEE 802.16d[3].

The main drawback of OFDM is its high PAPR.If the number N of carriers is large,the peak to average power ratio can also be large.The high PAPR causes degradation in the performance when non linear high power amplifier(HPA) is used,because HPA limits the output with certain value and reduces the power efficiency of amplifier. So HPA distorts the signal with high PAPR.The non linear distortion causes both in-band and out-of-band radiation. The band interference increases the bit error rate of received signal.In literature, various techniques has been proposed to reduce PAPR include clipping[4], coding[5], Selective level mapping(SLM)[6],tone-reservation[7],activeconstellation Mapping[8]. These techniques are complex in nature excluding the clipping technique.Coding method has BER performance at the expense of decreasing bandwidth efficiency. The objective of this paper is to reduce the PAPR.The Power spectral density of the proposed technique is also investigated.

In this paper,we propose a new technique in which PAPR reduction is done by combining Huffman coding with the modified companding function.Also the power spectral density of the proposed function is better compared to Huffman coding combined with original companding function. The proposed technique keeps the average power level the same as the original signal's power level.

The paper is organized as follows. Section II discuss about concept of PAPR,OFDM.In Section III,we introduce the PAPR reduction technique using Huffman coding combined with modified comapanding function and original companding function simulation results in section IV.Also the effects of the proposed method on the out of band (OOB) power emission are also investigated. Finally conclusions are presented in section V.

II. OFDM SYSTEMS & PAPR FORMULATION

In OFDM systems,total bandwidth is divided into many overlapping orthogonal subcarrier channels,which are transmitted in parallel.

The complex envelope of the transmitted OFDM signals can be given as[9]

$$x(n)=\frac{1}{\sqrt{N}}\sum_{k=0}^{N-1}S_k e^{j\frac{2\pi}{N}kn} \text{ ,n=0,1….N-1} \qquad (1)$$

Where k=0…N-1 represent a sequence of N symbols to be transmitted.

Inverse Fast Fourier transform(IFFT) and Fast Fourier Transform(FFT) are helpful in OFDM implementation. With the use of IFFT at the transmitter and FFT at the receiver, the bank of modulators and demodulators is not required.

The PAPR of the discrete OFDM signal may be expressed as

$$PAPR=\frac{\max\{|x(n)|^2\}}{E\{|x(n)|^2\}} \qquad (2)$$

Where E{.} denotes expectation.

The complementary cumulative distribution function(CCDF=1-CDF) of the PAPR is used and CCDF for $x(n)$ is defined as

$$P_{x[n]}(x[n_0]) = \mathrm{Prob}(x[n] > x[n_0]) \qquad (3)$$

Where $x(n_0)$ is the threshold level.CCDF is used to measure the performance of PAPR reduction methods.For large values of N(N 64),the OFDM symbols will have Gaussian distribution. Then CCDF turns out to be[10]

$$P_{x[n]}(x[n_0]) = 1 - (1 - e^{-x[n_0]})^N \qquad (4)$$

III. Proposed System

The proposed system combines the Huffman coding and modified companding function which is shown as in Fig.2.The input data stream is converted from serial to parallel stream among M symbols(each of n bits),then Huffman coding will generate a serial bit stream that is again converted to parallel among M symbols.Then N point IFFT is performed and then companding transform isapplied to the OFDM signal.

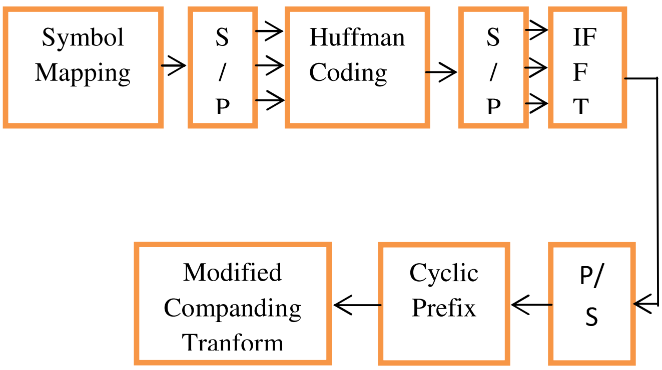


Fig. 1. OFDM Transmitterwith Huffman coding and Modified Companding transform

A. *Huffman Encoding*

Huffman coding will assign fewer bits to symbols which have more probability of occurrence and more bits to seldom occurred symbols[11]. Therefore,the addition of the same phase symbols will be reduced.

B. *Companding transform*

In the companding function,the compression of the OFDM signal at the transmitter and expansion at the receiver is done. μ law was first companding function proposed to decrease PAPR for OFDM systems[12].In μ -law companding technique,the compression of OFDM signals at the transmitter is given by

$$y = X_m \frac{\log\left(1 + \mu \frac{|x|}{X_m}\right)}{\log(1+\mu)} sign(x) \qquad (5)$$

Where X_m is the peak amplitude of the input signal and it is given as $X_m = \max(|x|)$ and x is the instantaneous value of theHuffman coded OFDM signal.Although μ-law companding function decreases the PAPR but in this Companding function the average power of the companded signal and original signal is not same, hence it increases the power spectral density (psd) of the transmitted signal.Another undesired effect of μ -law function, involves the requisite expansion of the received companded signal which

amplifies the receiver noise. In this section, we first propose a modified μ law companding function. Then the modified companding function is combined with the Huffman coded OFDM signal which results in lower psd as compared to combine with μ -law companding function.

C. *Modified Companding Function*

In modified Companding Function[13], a new parameter β is multiplied with the Original Companding function.The modified μ -law companding function is given as

$$Z = \beta X_m \frac{\log\left(1 + \mu \frac{|x|}{X_m}\right)}{\log(1+\mu)} sign(x) \qquad (6)$$

Where β is

$$\beta = \sqrt{\left(\frac{E(|x|^2)}{E(|y|^2)}\right)} \qquad (7)$$

Thus the average power of the Huffman coded OFDM signal combined with modified companding function will become equal to the Huffman coded OFDM signal.

IV. SIMULATION RESULTS

To evaluate the performance of the PAPR reduction and Power spectral density,the stimulated system employs an OFDM signal with QPSK modulated,N=52 subcarriers and IFFT length=64.The average length of Huffman coding 1.9 is used for simulation purpose. The value of parameterμ=16 is used for original companding function and μ=32 is used for the modified companding function. As the value of μ increases the PAPR reduces at the expense of increase of power spectral density.Figure 2 and figure 3 plot the CCDF,which is defined as the probability that the PAPR exceeds certain PAPR_0 for Huffman coded OFDM signal combined with Original companding and modified companding function respectively. The Huffman coding reduces the PAPR about 2 dB, companding function reduces the PAPR about 6dB where as the proposed method reduces the PAPR significantly about 9dB If the compression ratio in Huffman coding becomes higher, PAPR reduction will also increases.

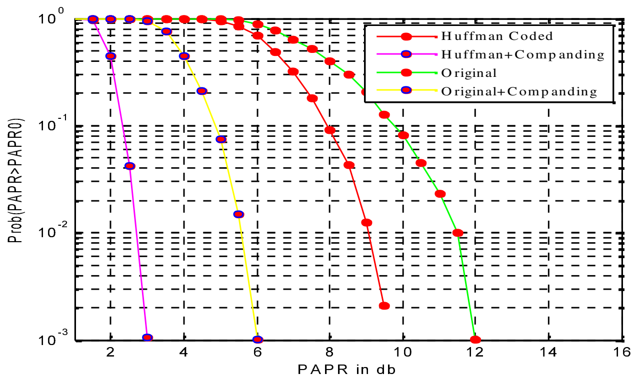


Fig.2. Comparison of CCDF For Original Companding Function

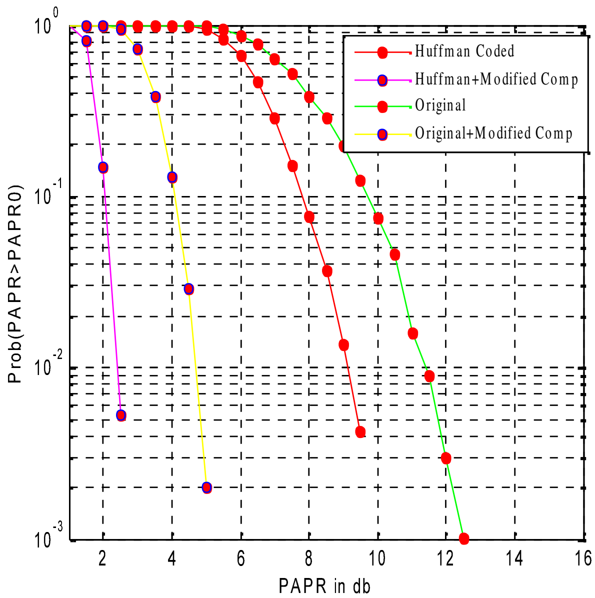


Fig.3. Comparison of CCDF For Modified Companding Function

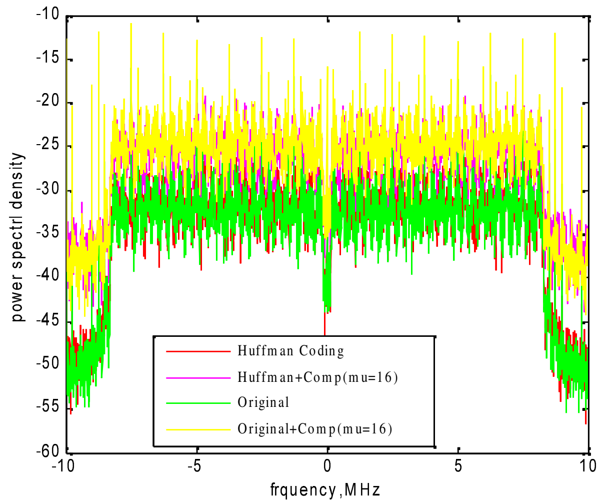


Fig. 4. PSD comparison For original Companding Function(=16)

Power spectral density comparison are depicted in Figure 4-7 where it is seen that the modified companding function has better psd than the original companding function. Figure 3-4 shows psd of original companding function. The results in figure 3-6 shows that the bigµparameter (µ=255) generates a higher power level in comparison with that of using a small µ parameter.As the value of µ increases Out of band emissions (OOB) increases.Figure 5-6 shows the psd of modified companding function in which normalized power is considered and it shows that the modified Companding function has smaller OOB emissions than that of the original companding function.

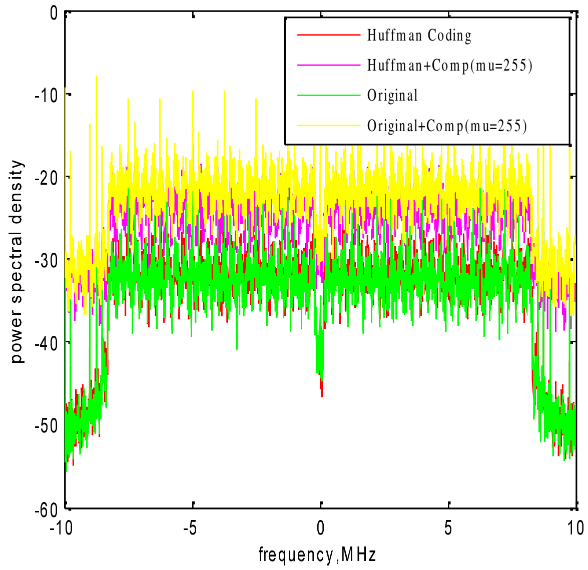


Fig.5. PSD comparison For original Companding Function(=255)

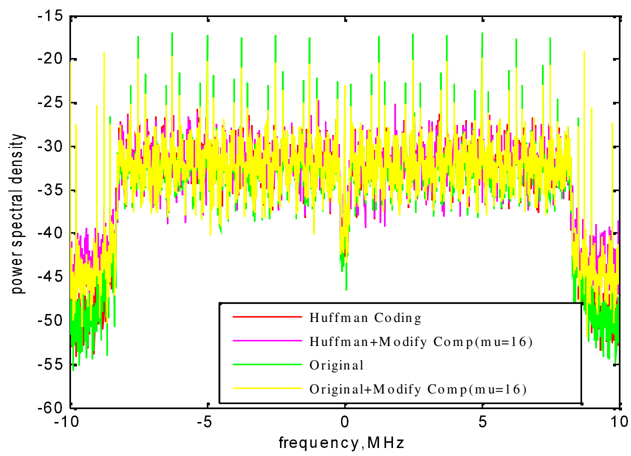


Fig.6.PSD comparison For Modified Companding Function(=16)

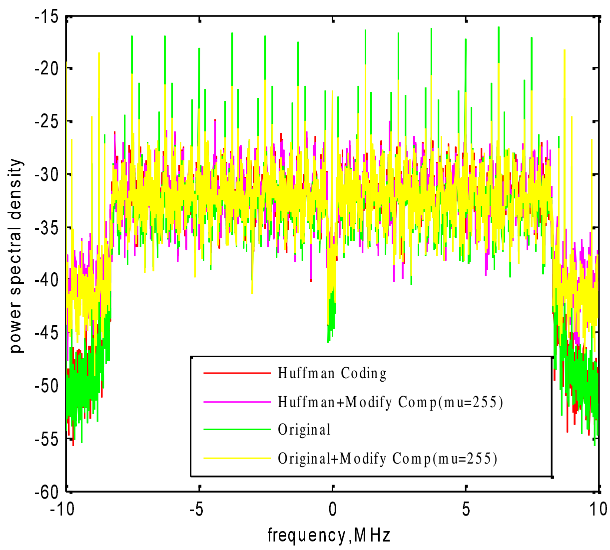


Fig.7.PSD comparison For Modified Companding Function(=255)

V. CONCLUSION

The proposed approach that combines the advantage of Huffman coding and modified companding methodreduces PAPR significantly about 9 dBThe out of band emissions increases with the increase in value of parameter μ.The original companding function reduces PAPR more significantly that that of the modified companding function but psd of modified companding function is better than that of original companding function.Although,the same amount of PAPR reduction can be achieved by slightly higher value of μ of modified companding function as compared to the original companding function.

REFERENCES

[1] R.V.Nee and Ramjee Prasad, "OFDM for MultimediaCommunication."Artech House, 2000, pp: 134-139.

[2] Ramjee Prasad,"OFDM for Wireless Communication System,"Artech House,2004,pp:189-193.

[3] M. I. Rahman, S. S. Das, and F. H. P. Fitzek, "Ofdm based wlan systems," Aalborg University, Centre for TeleInFrastruktur (CTiF), Den-mark, Tech. Rep. R-04-1002; v1.2, Feb. 2005,pp:1090-1095.

[4] H. Ochiai and H. Imai, "Performance of the deliberate clipping with adaptive symbol selection for strictly band-limited ofdm systems," IEEEJ. Select. Areas Commun., vol. 18, no. 11, pp. 2270–2277, Nov. 2000A. E. Jones, T. A. Wilkinson, and S. K. Barton, "Block coding scheme for reduction of peak-to-average envelope power ratio of multicarrier transmission systems," IEEE Electronics Letters, vol. 30, no. 8, Dec. 1994, pp. 2098– 2099,

[5] C. L. Wang and Q. Y. Yuan, "Low-complexity selected mapping schemes for peak-to-average power ratio reduction in ofdm systems," IEEE Trans.Signal Processing, vol. 53, no. 12, Dec. 2005, pp. 4652–4660,

[6] J. Tellado, "Peak to average power ratio reduction for multicarrier modulation," Ph.D. dissertation, Univ. of Stanford, Stanford, 1999.

[7] B. S. Krongold and D. L. Jones, "Papr reduction in Ofdm via active constellation extension," IEEE Trans. Broadcast., vol. 49, no. 3, Feb. 2003, , pp. 258–268,

[8] Erik Dahlman,et. Al, "3G Evolution:HSPA and LTE for Mobile Broadband, "Academic Press,2007.

[9] S. H. Han and J. H. Lee, "An overview of peak-to-average power ratio reduction techniques for multicarrier transmission," IEEE WirelessCommun. Mag., vol. 12, no. 2, , Apr. 2005 pp. 56–65

[10] Ashraf A.Eltholth,et. Al, "Peak-to-Average Power Ratio Reduction in OFDM Systems using Huffman Coding," Proceedings of World Academy of Science,Engineering and Technology Vol.33,Sept .2008.

[11] T. Jiang and G. Zhu, "Nonlinear companding transform for reducing peak-to-average power ratio of ofdm systems," IEEE Trans. Broadcast., vol. 50, no. 3, Sep. 2004, pp. 342–346,

[12] Gazi, O., "A new companding technique for PAPR reduction in OFDM communication systems," Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2011 3rd International Congress on , vol., no.2, Oct. 2011, pp.1-7.

Performance Comparison of DYMO and LANMAR by Varying Nodes and Velocity in MANET using Random Waypoint Mobility Model

Shruti Bajaj
CSE Department, CTIEMT, Jalandhar
shrutibajaj826@gmail.com

Lakhwinder Kaur
CSE Department, CTIEMT, Jalandhar
lakhwinder kaur.24@gmail.com

***Abstract*— A Mobile Ad-hoc Network (MANET) is a self configuring, infrastructure-less network of mobile devices (nodes) where nodes communicate with each other using multi-hop wireless links. It may connect hundreds to thousands of mobile nodes. These nodes form an arbitrary topology, where the routers are free to move randomly and arrange themselves as required. This paper presents the effects of velocity and varying nodes on the performance of two MANET routing protocols i.e. Dynamic MANET On Demand (DYMO reactive routing protocol) and Landmark Ad-hoc routing (LANMAR proactive routing protocol).The performance is analyzed using varying number of nodes and maximum velocity based on Packet Delivery Ratio (PDR), Throughput, Average end to end delay, Average Jitter. These simulations are carried out using QualNet Simulator.**

Keywords*— *Mobile Ad-hoc Network (MANETs) Routing Protocol, DYMO, LANMAR, Random Waypoint Mobility Model, CBR,* QualNet5.1 *(Simulator).

I. INTRODUCTION

Wireless communication networks are either infrastructure based networks or ad-hoc networks. Ad hoc network consists of mobile nodes which communicate with each other through wireless medium without any fixed infrastructure. A mobile ad-hoc network (MANET) is a collection of nodes, which have the possibility to connect on a wireless medium and form an arbitrary and dynamic network with wireless links. That means that links between nodes can change during time, new nodes can join the network, and the other nodes can leave the network [1]. The main vision of MANET is to support robust and efficient operation in wireless networks by incorporating routing functionalities at each mobile node. In MANETs in order to support mobile computing the routing protocols will have to perform necessary functions such as determination of network topology, maintaining network connectivity, transmission scheduling and channel assignment and packet routing. Routing protocols in MANETs are based on the design goals of minimal control overhead ,minimal processing overhead, dynamic topology maintenance and loop prevention .So the performance analysis of MANET routing protocols is an essential task to know its behavior and work in that environment[2].

Now we briefly describe the key characteristics of MANET Routing Protocols DYMO and LANMAR which are used in our simulations .We also describe the particular parameters that we choose when implementing each protocol .In reactive or On Demand Routing routes are only discovered when they are actually needed i.e. they invoke a route determination procedure only on demand [2]. In contrast, in proactive routing (table driven routing) each node must continuously maintain route between pair of nodes.

The Dynamic MANET On-Demand (DYMO) protocol [3] is a simple and fast routing protocol for multi-hop networks. It discovers uni-cast routes among DYMO routers within the network in an on-demand fashion, offering improved convergence in dynamic topologies. To ensure the correctness of this protocol, digital signatures and hash chains are used. The basic operations of the DYMO protocol are route discovery and route management. During route discovery Route request (RREQ) message is broadcasted to the network. Every intermediate node participates in hop-by-hop dissemination of this message and records a route to the originator [4]. When a destination node receives this RREQ message, it responds with a Route reply (RREP) message uni-cast towards the originating node. Every node receiving this message creates a route to the destination node and finally this RREP message arrives at the originator of the RREQ message. When a change occurs in the network, topology nodes maintain their routes and monitor their links. When a data packet is received for a route or link that is no longer available the sources of packet generates a Route error (RERR) message and send this RERR message to the packet source to indicate the current route is broken .

LANMAR protocol combines properties of link state and distance vector algorithm and builds subnets of groups of nodes which are likely to move together .A Landmark node is elected in each subnet. During the packet forwarding process, the destination is checked to see if it is within the forwarding node’s neighbor scope. The packet is directly forwarded to the address obtained from the routing table. On the other hand, if the packet's destination node is much farther. The packet is first routed to its nearest landmark node. As the packet gets closer to its destination, it acquires more accurate routing information, thus in some cases it may bypass the landmark node and routed directly to its destination.. A distance vector, which is calculated based on the number of landmarks, is added to each update packet. [3].

Mobility models represent the movement of mobile users, and how their location, velocity and acceleration change over time. Such models are frequently used for simulation purposes when new communication or navigation techniques are investigated [5].

Random Waypoint Mobility was first proposed by Johnson and Maltz is a random mode designed for the movement of mobile users, and how their location, velocity and acceleration change over time [5]. It includes pause times between changes in direction and/or speed. A mobile node begins by staying in one location for a certain period of time (i.e., a pause time). Once this time expires, the mobile node chooses a random destination in the simulation area and a speed that is uniformly distributed between [min-speed, max-speed].

The main focus of this work is to analyze and investigate the effect of mobility on two MANET routing protocol DYMO (reactive) and LANMAR (proactive) using Random Waypoint mobility model for wireless ad-hoc.

II. SIMULATION ENVIRONMENT

In this paper, an attempt has been made to investigate and analyze the effect of mobility on two MANET routing protocols named DYMO and LANMAR. For this random waypoint mobility model is used. Random mobility models the random movement of mobile nodes. The simulations are successfully carried out using Qual-Net simulator [6]. Qualnet is a comprehensive suite of tools for modeling large wired and wireless networks. It uses simulation to predict the behavior and performance of networks to improve their design, operation and management. It is a network simulation tool that simulates wireless and wired packet mode communication networks [6].

For the entire simulations traffic source used is CBR, item size is 512 bytes, simulation time is 300 seconds, and transmission range is 250m within an area of 1000m*1000m.

TABLE I SIMULATION PARAMETERS

Simulation parameters	***Value***
Dimension of space	1000*1000
Network Type	Mobile
Node Placement Strategy	Random
Traffic source	C B R
Network size (Number of nodes)	40,60,80,100,120
Velocity	30mps, 60mps, 90mps, 120mps, 150mps
Protocols	DYMO and LANMAR
Mobility Model	Random Waypoint Mobility Model
Transmission range	250m
Simulation time	300 seconds

The performance metrics used to measure the performance of routing protocols are: packet delivery ratio, throughput, average end to end delay, and jitter. Packet delivery ratio (PDR) is the ratio between number of received packets by destination node and the number of data packets sent by source node. It should be close to unity for better performance

PDR= Number of packets successfully received at application layer / Number of packet send from source application layer.

Throughput is measured by the total amount of packets which is received by a destination node. It is measured by *byte/sec* or *bit/sec*. High throughput is always expected for any routing protocol.

Throughput= Total amount of data received from sender/Time takes for the receiver to get the last packet (bits/sec).

Average end-to-end delays of data packets includes all possible delays caused by buffering during route discovery latency, queuing at the interface queue, retransmission delays at the MAC, and propagation and transfer times.

Average jitter is the time variation between subsequent packet arrivals. For an efficient routing, it should be as low as possible.

III. RESULTS AND DISCUSSIONS

This section discusses the behavior of the routing protocols considering the number of nodes and maximum velocity using random waypoint mobility model.

A. Case-I VARYING NUMBER OF NODES

The packet delivery ratio graph for MANET routing protocols DYMO and LANMAR for number of nodes using Random Waypoint Mobility model is shown in Fig.1. From the graph, we can see that DYMO performs better than LANMAR protocol in terms of delivery ratio. DYMO has more delivery ratio at every point for nodes due to its reactive nature. From nodes 100 to 120 both protocols perform parallel operations. For efficient routing PDR should be high. Thus, we can conclude that DYMO shows better performance in terms of delivery for number of nodes using Random Waypoint Mobility model.

The throughput for DYMO and LANMAR protocols for varying nodes shown in Fig.2. Our simulation results revealed that LANMAR has lesser value in comparison DYMO protocol. The reason for lesser value of throughput for LANMAR is that its value declines when the number of nodes is 80. When node density is 120, both protocols give almost same operation. Thus, it is concluded that DYMO shows better performance in terms for throughput.

The average End-End Delay is shown in Fig.3. According to our simulation results, both the protocols are showing similar performance. When the numbers of nodes are 40 the value of DYMO rises to great extent which in turn lowers its performance. When node density is 60, both protocols give exact values of delay. There is a great fluctuation in value of protocols when nodes are100. In average DYMO has greater value than LANMAR, but for efficient routing delay should be low. Thus, LANMAR is found to posses' minimum end to end delay with maximum value hence shows better performance for average delay under mobility.

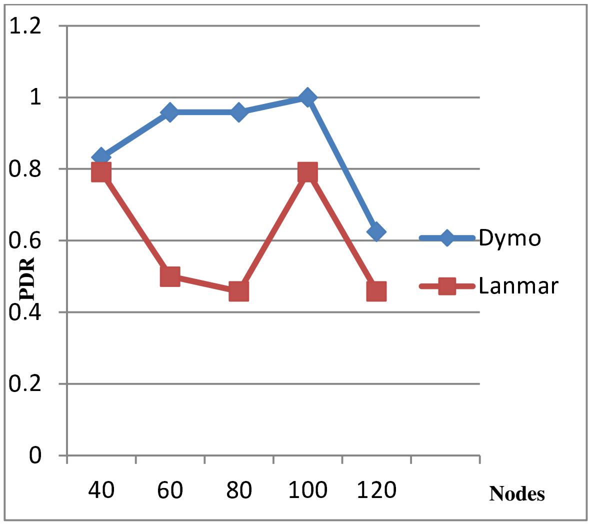


Fig.1. Number of Nodes v/s Packet Delivery Ratio (PDR)

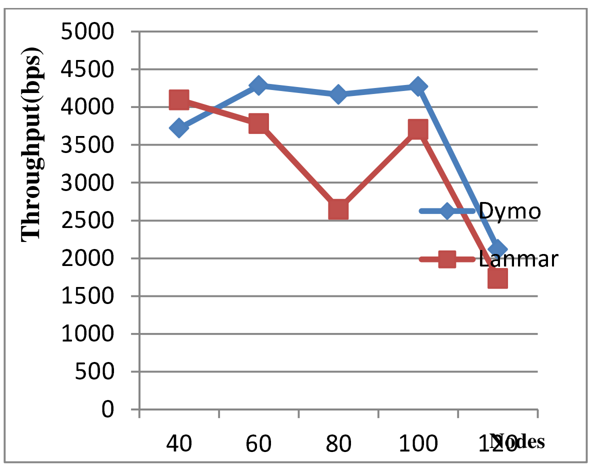


Fig.2. Number of Nodes v/s Throughput

Fig.4 shows average jitter. In our simulation results we clearly observe that DYMO has higher value than LANMAR protocol at almost every .This is due to the high packet collision probability more chances to loss packet between subsequent packet transmissions For efficient routing Jitter should be low. So, we can conclude that LANMAR gives better performance for average jitter under random waypoint mobility

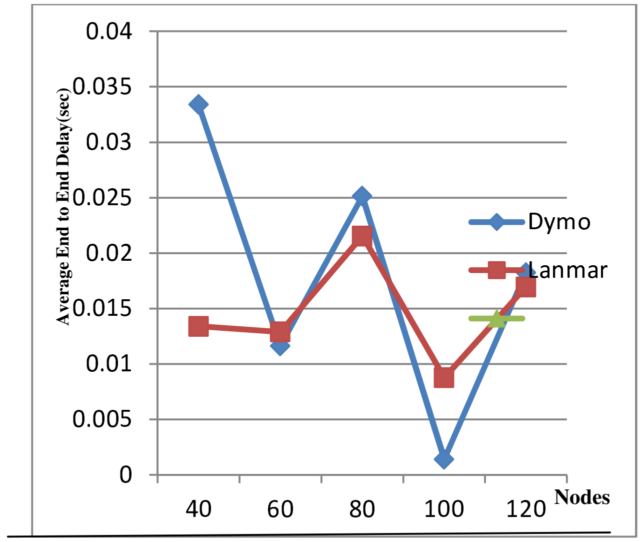


Fig.3. Number of Nodes v/s Average End-to-End Delay

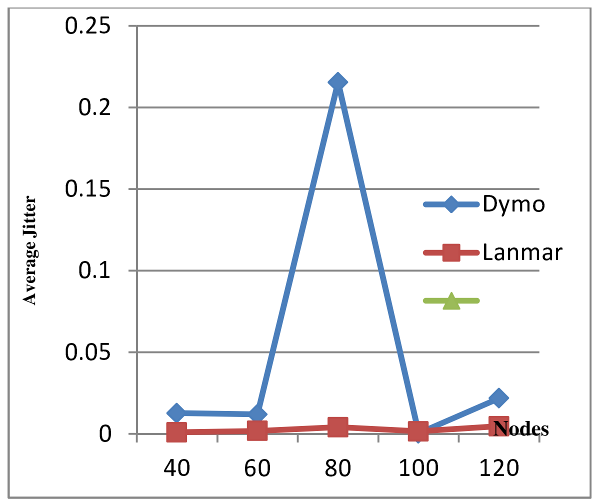


Fig.4. Number of Nodes v/s Average Jitter

B. Case-II VARYING MAXIMUM VELOCITY

The packet delivery ratio graph for MANET routing protocols DYMO and LANMAR for maximum variable velocity using random waypoint mobility is shown in Fig .5. From the graph, it can be clearly shown that can see that DYMO has higher value at all points. At point of 30 mps both protocols have same delivery ratio. As the velocity increases, DYMO posses higher value than LANMAR. Hence, we can finally conclude that DYMO shows the better performance in terms of delivery for variable velocity under random waypoint mobility. The graph for throughput with respect to maximum velocity is shown in Fig .6. Our results show that LANMAR shows inferior and lower throughput as compared to DYMO. With velocity increase, throughput of DYMO is found to be parallel to LANMAR. We also find that there is a great decline in the throughput of LANMAR when the velocity increases from 60 to 90 mps, then again increases minor. So, overall LANMAR performance degrades. Hence, we can conclude that DYMO shows the better performance in terms of throughput with variable maximum velocity under mobility.

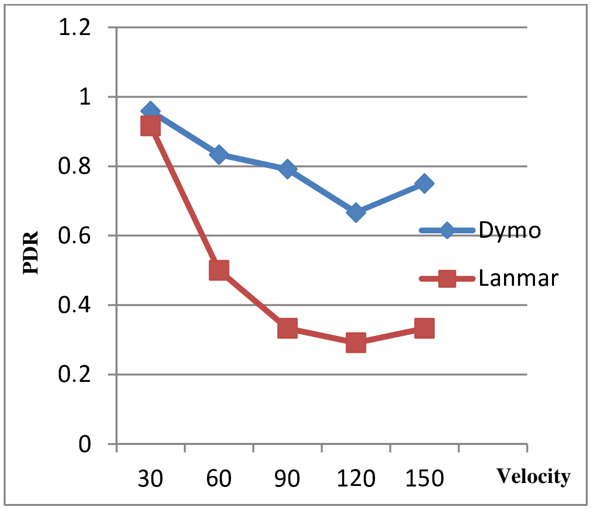


Fig.5. Maximum Velocity v/s Packet Delivery Ratio (PDR)

The graph for average end-to–end delay for DYMO and LANMAR routing protocols is shown in Fig..7. From the graph it is clear that LANMAR protocol has minimum en-to-end delay throughout the velocity variation thereby giving the best performance. It is also observed that DYMO shows the worst performance in case of delay having maximum value of delay. For an efficient routing delay should be low. Hence, we conclude that LANMAR is better protocol than DYMO for varying velocity under mobility.

The graph for average jitter for maximum velocity can be shown in Fig.8. Our simulation results show that LANMAR shows minimum jitter with a maximum value giving best performance. DYMO here also demonstrated the worst performance with maximum value. Its performance decreases to a large extent when velocity is 60mps, then increases slowly. The value of LANMAR protocol remains almost same i.e almost a straight line throughout the increase in velocity. So, average jitter has no effect on LANMAR protocol for velocity variation under mobility. Hence, LANMAR performs better for variable velocity for jitter using Random Waypoint mobility model**.**

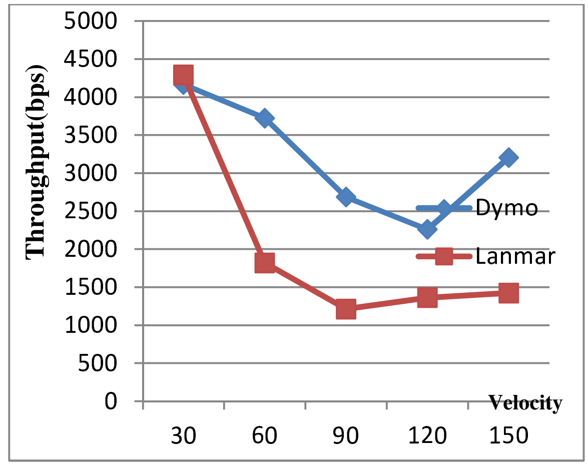


Fig.6. Maximum Velocity v/s Throughput

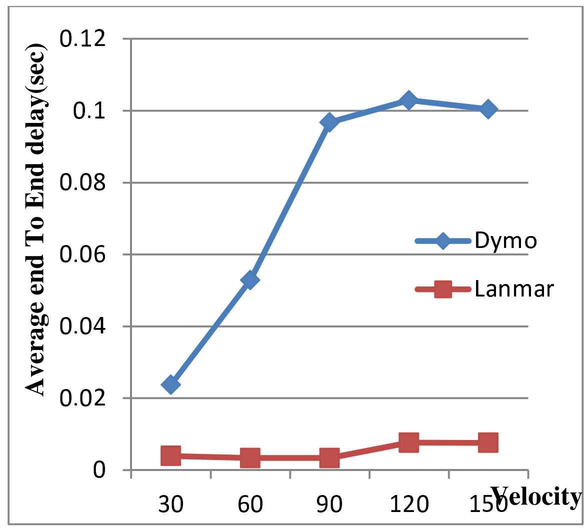


Fig.7. Maximum Velocity v/s Average end-to-end Delay

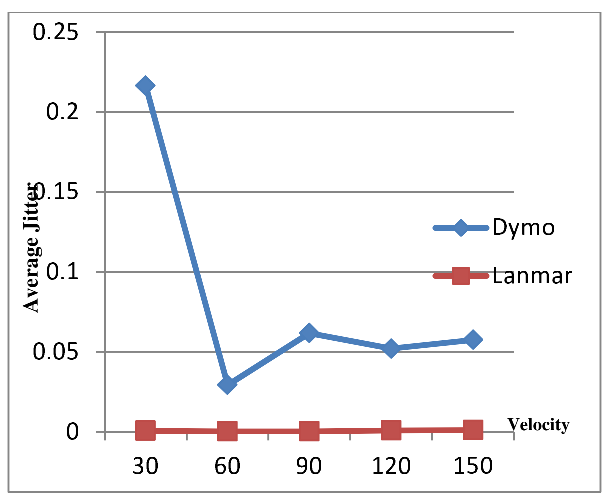


Fig.8. Maximum Velocity v/s Average Jitter

IV. CONCLUSION AND FUTURE WORK

In this paper we have analyzed and investigated the effects of mobility on DYMO and LANMAR MANET routing protocols using Random Waypoint Mobility model, by varying nodes and maximum velocity for ad-hoc networks using Qualnet simulator. For the packet delivery ratio, both for varying nodes and varying maximum velocity DYMO protocol proved better as compared to LANMAR protocol. In case of throughput, also for both varying nodes and maximum velocity DYMO showed better performance than LANMAR protocol.

For the end-to-end delay LANMAR protocol proved better as compared to DYMO protocol under mobility. For the case of average jitter also, LANMAR protocol showed better performance compared to DYMO protocol under mobility. In this work different simulation scenarios are generated with varying nodes, and varying maximum velocity.

Future work of MANET routing protocol involves the implementation of new routing protocols. Other protocols can also be taken into account such as TORA, ZRP and LDR with respect to different metrics including node density, pause time and energy. Future work can also be geared using another simulator i.e. Qualnet Cyber Xeta which supports additional features of MANET routing protocol such as energy models, battery models etc.

REFERENCES

[1] Mrs. Geeta V, Dr. Sridhar Aithal, and Dr. K. ChandraSekaran "Effect of Mobility over Performance of the Ad hoc Networks ", IEEE INFOCOM 2006.

[2] R. Biradar1, Hiren H D Sarma, and Kalpana Sharma ," Performance Comparison of Reactive Routing Protocols of MANETs using Group Mobility Model" International Conference on Signal Processing Systems 978-0-7695-3654-5/09 $25.00 IEEE 2009.

[3] Harish Shakywar, Sanjeev Sharma, Santoh Sahu," Performance Analysis of DYMO and STAR Routing Protocols for Grid Placement model with varying network Size IJCTA ISSN: 2229:6093.

[4] MIAO Quan-xing , XU Lei," DYMO Routing Protocol Research and Simulation Based on NS2 " 2010 International Conference on Computer Application and System Modeling (ICCASM 2010).

[5] Homepage on Mobility. [Online]. Available: http://en.wikipedia.org/wiki/Mobility_model.

[6] Qualnet 5.2.Available. [Online].http://Scalable-networks.com.

Bandwidth Reservation Scheme for Keeping the Call-Dropping Probability Low in Cellular Networks

Dipanjan Saha
JIS College of Engineering
India
dipanjan.saha.1989@gmail.com

Avishek Gupta
JIS College of Engineering
India
avig2005@gmail.com

Sumanta Chatterjee
JIS College of Engineering
India
jiscesumanta@gmail.com

Abstract—**In the present paper the bandwidth reservation techniques have been widely studied in cellular networks in order to full fill the desired quality-of-service (QoS) requirements. The goal is to decrease the call-dropping probability and to increase the bandwidth utilization. Proper system design and operation involve keeping the call drop rate as low as possible. There are numerous call drop causes in cellular networks with most of them occurring due to lack of radio resources created by electromagnetic causes and handover. Another important factor of call drop rate is the traffic load in which, the call arrival rate and holding time play significant roles. In this paper, we also present an overview of call-dropping probability factors in cellular networks. In addition, a bandwidth reservation adaptation algorithm is also proposed. The algorithm adjusts the amount of bandwidth reserved for handover sessions according to the current network conditions thus creating a balance between new session blocking rate and handover session dropping rate.**

Keywords*: *Call drop probability, quality of service (QoS), handover, traffic load, radio channels, bandwidth, cellular network, channel reservation.

I. INTRODUCTION

In recent years, there has been a tremendous growth in wireless networking industry. The cellular networks are now expected to support multimedia traffic with a wide range of bandwidth requirements. To allow more applications to share the radio bandwidth, there is a trend toward reducing the cell size. With a smaller cell size, calls are more likely to be forced to terminate due to insufficient bandwidth when handoff takes place. Bandwidth is shared between handoff and new calls. Handoff calls are usually given higher priority than new calls when allocating network channels. This can be done by reserving a number of channels in every cell to be used by handoff calls only. A well designed scheme can achieve the desired Quality of Service (QoS) requirements and efficiently utilize network resources by avoiding unnecessary reservation.

Many channel-reservation algorithms have been proposed to reduce the handoff-call forced-termination rate. Dynamic reservation algorithm is proposed to overcome the Call dropping probability.

Call dropping is caused by lack of available radio channels which in turn may be caused by propagation factors such as distance losses, path loss, and multipath fading.

Another important fact of call dropping is Handover. Handover is the mechanism that transfers an ongoing call from one cell to another cell as a user moves through the coverage area of a cellular system. When a Mobile Terminal (MT) moves, it is quite possible that the currently serving Base Station Subsystem (BSS) may no longer be able to provide reasonable quality of service as compared to some other BSS. The currently serving Mobile Switching Centre (MSC) may decide to hand over this service to some other better serving BSS or in some cases to another MSC. Occasionally, this handover process fails and the call drops. The minimization of call drop rate needs efficient algorithm for making handover requests at the right place at the right time based on the propagation environment.

Another important mechanism of call drop rate is the traffic load in which, the call arrival rate and holding time play significant roles.

II. MATHEMATICAL MODEL

A. Number of Channels:

The number of channels is calculated using the Erlang B formula for loss probability. The formula is given as follows:

$$B = \frac{\frac{A^N}{N!}}{\sum_{k=0}^{N} \frac{A^k}{k!}} \qquad (1)$$

Where B = loss probability, A = offered traffic intensity in Erlangs, N = available number of channels

This shows that call-dropping probability decreases as the number of channels increases.

B. Channel Utilization:

The channel utilization depends on the total traffic. The formula is given as follows:

ρ = (Traffic intensity) / (Number of channels) (2)

C. Call Arrival Rate:

$$\lambda_t = \frac{\text{Number of ca}}{\text{14400 se}} \quad (3)$$

D. Call duration:

Call duration is another parameter that can affect the quality of service in a cellular network.

$$h = \frac{A}{\lambda} \quad (4)$$

E. Drop-Call Probability:

$$P(Y = n) = \frac{(v_d t)^n}{n!} e^{-v_d t}, \quad n \geq 0 \quad (5)$$

Here, V_d= is the drop-call rate, t= the call duration, Y = a random variable that counts the number of drops and n= the confirmed calls dropped. Where A = traffic intensity in Erlangs, λ = call arrival rate.

The number of dropped calls is calculated from the relation:

Drop call rate =

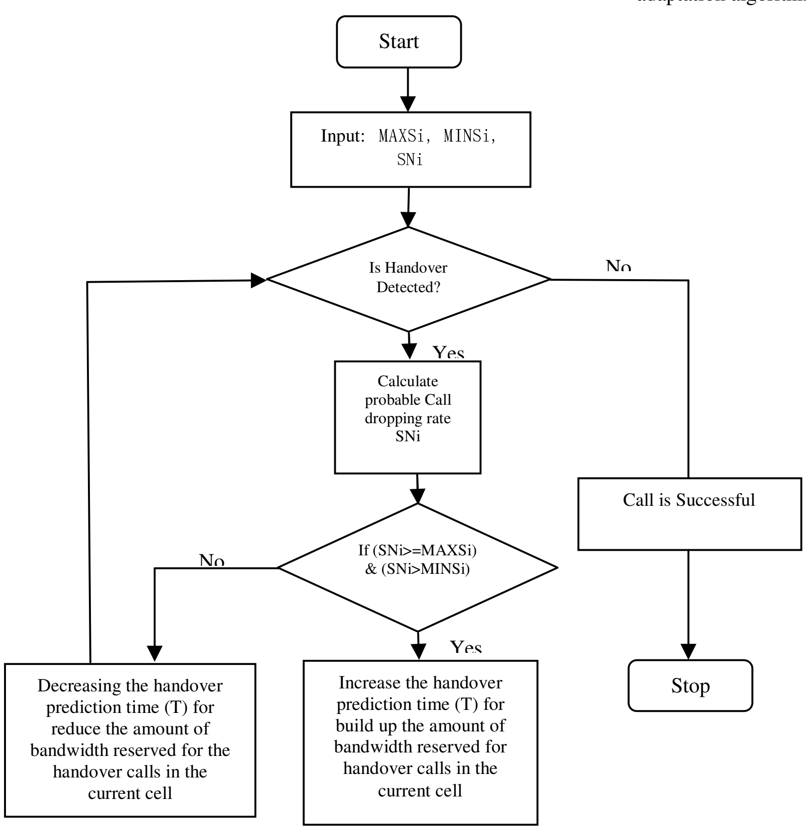
III. PROPOSED DYNAMIC BANDWIDTH RESERVATION ADAPTATION ALGORITHM

Call arrival rate, λt, refers to the traffic offered expressed as the number of call attempts per unit time. The formula is given as follows:

MAXSi represents the acceptable upper bound of the handover call dropping rate for Class i service flows , MINSi represents the corresponding lower of the handover call dropping rate for Class i service flows, Ri represents the number of Class i handover calls in each period in its cell , RXi represents the number of Class i handover calls dropped due to the lack of bandwidth in each period in its cell , SNi represents the handover call dropping rate of Class i service flows and T represents the handover prediction time.

Step 1. Step1: Take Inputs as MAXSi, MINSi, Ri, RXi and SNi and T=0;
Step 2. Record R1, R2, RX1, RX2 in this period of time;
Step 3. Call dropping rate of two services, i.e.; SN1= RX1/ R1; SN2= RX2/ R2;
Step 4. if (SN1>=MAXS1 && SN2<MINS2)
Step 5. then T = T + SNi;Step5: else if (SN1<MINS1 && SN2>=MAXS2)
Step 6. then T = T – SNi;
Step 7. else if (SN1>=MAXS1 && SN2>=MINS2 || SN1>=MINS1 && SN2>=MAXS2)
Step 8. : then notify the neighbour cells to increase T;
Step 9. else if (SN1<MINS1 && SN2<MAXS2 || SN1<MAXS1 && SN2<MINS2)
Step 10. then notify the neighbour cells to decrease T;
Step 11. Stop.

Fig. 1. Flow chart of the proposed dynamic bandwidth reservation adaptation algorithm



IV. RESULTS AND DISCUSSIONS

SNi (the handover call dropping rate of Class i service flows) calculated by formula (1).

SN1= RX1/ R1; SN2= RX2/ R2 (1)

When a BS (BS of cell j, for example) has received bandwidth reservation requests from all of its neighbor cells, it then calculates the amount of resources that need to be reserved in its cell for different types of handover service flows in the next period by formula (2) below.

$$G_{j,k} = \sum_{k \in F, i \in I} B_{i,j,k} + T_{j,k} \quad (2)$$

Gj,k represents the total amount of bandwidth reserved in cell j for type k service flows .It is composed of two parts. the first part is the sum of the bandwidth requests from all the neighbor cells, while the second part Tk is a parameter used to adjust the amount of bandwidth reserved for type k service flows according to the current handover call dropping rate. I denote the set of all the neighbor cells of cell j.

Finally, estimate Tj, k in formula (2) by formula (3)-(4):

$$T_{j,1} = T \; T_{j,2} = -T \quad (3)$$

$$T_{j,1.i} = \frac{G_{j,1,i}}{\sum_{q=1}^{3} G_{j,1,q}} \times T_{j,1} \quad (4)$$

There are two different methods to adjust the amount of bandwidth reservation:

a.) If one of the two classes of handover service flows has a very high call dropping rate (>=MAXS), while the call dropping rate of the handover service flows of the other class is at a low level (<MINS), then it could increase or decrease T between the bandwidth reservation of these two classes of service flows.

b.) If one of the two classes of handover service flows has a very high call dropping rate (>=MAXS), while the call dropping rate of the handover service flows of the other class is also at a relatively high level (>MINS), then it could increase the handover prediction time for the neighbor cells to augment the amount of bandwidth reserved for handover calls in the current cell. On the contrary, decreasing the handover prediction time for the neighbor cells would reduce the amount of bandwidth reserved for the handover calls in the current cell.

Table I. NETWORK DATA

MSC	**Number of Active Base Stations**	**Period**	**Traffic Intensity**	**Drop Call Rate (%)**	**Drop call rate with Handover (%)**	**Number Of call Attempts in Busy Hour**
MSC1	60	January	2610	3.09	0.06	148698
		February	2450	2.59	0.06	149880
		March	2596	2.82	0.07	151597
		April	2537	2.98	0.08	157018
MSC2	54	January	1836	3.16	0.04	116560
		February	2070	2.68	0.05	129990
		March	1918	2.83	0.05	113816
		April	1849	2.35	0.07	134855
MSC3	51	January	1390	1.51	0.07	95212
		February	1438	1.71	0.08	94388
		March	1295	1.79	0.07	91888
		April	1319	1.92	0.08	90865

Table II. MEAN VALUES OF COMPUTED TRAFFIC/CHANNEL PARAMETERS AND DROP-CALL PROBABILITY

MSC	**Call Arrival Rate (calls/s)**	**Call Duration (s)**	**Number of Channels**	**Channel Utilization Factor**	**Drop-call Probability y (%)**
MSC1	0.18	245	47	0.92	3.9
MSC2	0.16	223	41	0.91	4.5
MSC3	0.15	216	37	0.87	5.3

In the present paper the call drop probability analysis with experimental data has taken from an operative GSM network. The information were collected over a period of six months for each of six MSCs and comprised of the traffic intensity, drop-call rate and number of call attempts as shown in Table 1. From this the Number of channels, channel utilization factor, call arrival rate and call duration were estimated as shown in Table 2.

The results indicate that the call-drop probability decreases as the number of radio channels, utilization factor, call arrival rate and call duration increase as shown in Table 2. The optimized results were graphically represented as shown in Fig: 1, 2, 3 and 4. The curves above those four figures are not unexpected because of the constantly varying and unpredictable propagation characteristics of the radio channel.

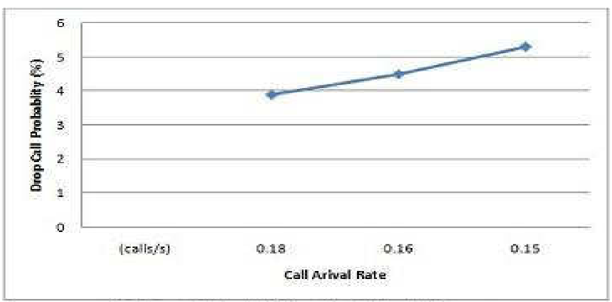


Fig1:Drop Call Probability with Call Arrival Rate

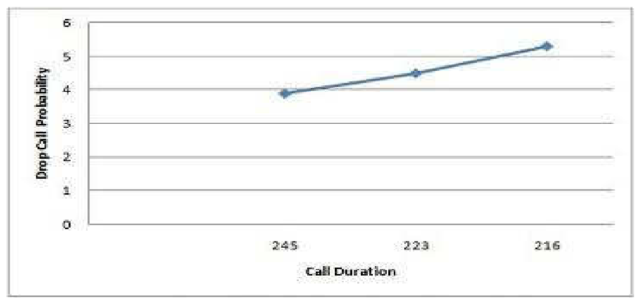


Fig2. Drop Call Probabalility With Call Duration

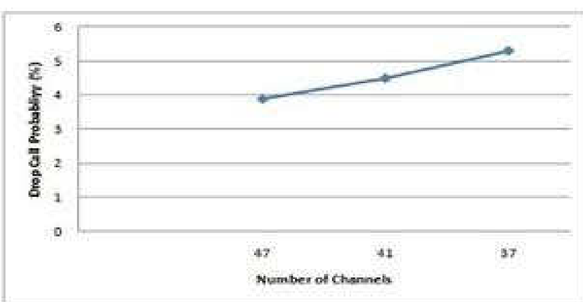


Fig3:Drop Call Probability with Number of Channels

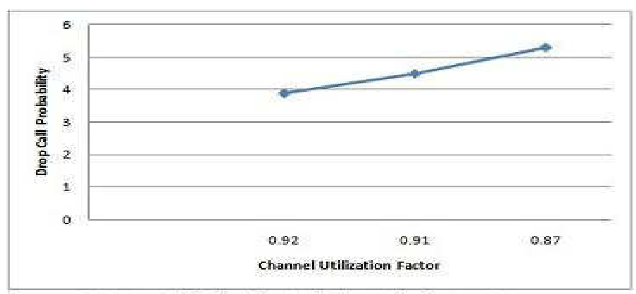


Fig4:Drop Call Probability With Channel Utilization Factor

V. CONCLUSION

In this present paper proposed a dynamic, prediction-based adaptive bandwidth reservation algorithm for the different types of service flows in the networks. It first introduced a handover prediction algorithm, then according to the prediction results, our algorithm adopts a multi-class bandwidth reservation and call admission control policies to ensure that different types of service access the network and utilize the system resources. In addition, this paper also demonstrated an algorithm to dynamically adjust the amount of bandwidth reserved based on the current handover call dropping rate. According to adjust the bandwidth we achieve a balance between the new call blocking rate and the handover call dropping rate.

REFERENCES

[1] G. Boggia, P. Camarda, and A. D'Alconzo, "Modelling of Call Dropping in Well-Established Cellular Networks", EURASIP Journal on Wireless Communications and Networking, Vol. 2007, Article ID 17826, October, 2007.

[2] B.L. Cragin, "Prediction of Seasonal Trends in Cellular Dropped Call Probability", IEEE, Capital City Applied Research, Research Paper, 2375 Club Mredian Dr. #B7, Okemos, MI 48864, USA, 2006.

[3] A. R. Mishra, "Fundamentals of Cellular Network Planning & Optimization", 2004 John Wiley & Sons, Ltd. ISBN: 0-470-86267-X, PP 21-29.

[4] S. Rapp port. The Multiple-Call Hand-Off Problem in High-Capacity Cellular Communications Systems. In IEEE Trans. On Vehicular Technology, Vol. 40, No. 3, August 1991.

[5] P. Orlik and S. Rappaport. Traffic Performance and Mobility Modelling of Cellular Communications with Mixed Platforms and Highly Variable Motilities. In Proceedings of VTC97. Phoenix, USA, May 1997.

[6] D. Hong and S. S. Rappapot, "Traffic model and performance analysis for cellular mobile radio telephone system with prioritized and nonprioritized handoff procedures," IEEE Trans. Veh. Technol., vol. VT-35, pp. 77–92, Aug. 1986.

[7] N. D. Tripathi, J. H. Reed, and H. F. VanLandinoham, "Handoff in cellular systems," IEEE Pers. Commun., vol. 5, pp. 26–37, Dec. 1998.

[8] D. A. Levine, I. F. Akyildiz, and M. Naghshimeh, "Aresource estimation and call admission algorithm for wireless multimedia networks using the shadow cluster concept," IEEE/ACM Trans. Networking, vol. 5, pp. 1–12, Feb. 1997.

Network Coding for Wireless Communication

Iqbal Singh
Dept. of CSE
GNDU RC, Gurdaspur

S.B. Rana
GNDU RC, Gurdaspur

Rasneet Kaur Chauhan
Dept. of ECE
GNDU RC, Gurdaspur

***Abstract*—Wireless networks faces number of problems such as low throughput, dead spots, mobility issues etc. However, their properties like broadcast nature of the medium and spatial diversity open windows address such issues. Recently much interest has been shown in network coding in wireless communications. This paper explores the network coding for wireless networks, by describing how it faces problems of throughput, reliability, mobility, and management. We also discuss the practical challenges that it faces.**

Keywords—BER, Redundancy, Coding, Linear combination

I. INTRODUCTION

The design of wireless networks is based on the basic model of wired architecture. It abstracts the wireless channel as a point-to-point link, and bring-in the protocols used in wired network into the wireless environments. For example, we use shortest path protocols, routers have nothing to do with the modification but only forwarding the data, and reliability of network during packet losses relies on retransmissions. It proved to be good for wired media but in case of wireless medium it is rather unpredictable and unreliable [1].

The fundamental nature of wireless medium is totally different. BER is high and it may vary over short time scales. Wired networks are comparatively predictable. Further, wired links are unicast links, but the majority of wireless links (with omni-directional antennas) are broadcast links. There is no interference between transmissions of wired networks, while it is the most common problem in wireless communications. Wired nodes are static while the wireless networks were designed to meet the fundamental needs of mobility and portability.[5]

These characteristics may seem to be disadvantageous at first glance but they are not. These can be used as advantages. Unreliability can be dealt with the broadcast nature of wireless; when a node broadcasts a packet, it is likely that at least one nearby node receives it, which can then function as the next-hop and forward the packet. In conventional approach there is a single designated next-hop, and when the reception fails retransmission is required from previous hop. Data redundancy is also one of the features of wireless networks. It occurs because packet contents are known to multiple nodes due to multiple hops. Also as the same packet is broadcasted to multiple nodes within the coverage, redundancy is amplified. There is need to of an alternative design to utilize intrinsic properties of wireless networks, such as spatial diversity and data redundancy, rather than foisting an artificial wired abstraction.

In this paper we discuss how network coding can be used to improve the throughput, reliability, fairness, and management of wireless networks. We also discuss the practical limitations and challenges it faces to cope with existing design.

II. NETWORK CODING

In network coding contents from two nodes are mixed and coded at some intermediate node and then forwarded. Prior work shows that network coding achieves the multicast capacity of the network [1]. Two results have practical implications. First, linear combinations of the packets, c_ip_i, (where c_i is coefficient and p_i is packet) achieve the maximum capacity bounds, and coding and decoding can be done in polynomial time. Second, Ho et al. show that the above is true even when the routers pick random coefficients [2]. This enables distributed network coding, where consultation between routers is not needed. Researchers have applied network coding to many areas including wireless networks, energy, secrecy [3], content distribution, and distributed storage. Most work on network coding, however, is theoretical.

III. POTENTIAL BENEFITS

The following simple examples will explain the potential benefits of building future wireless networks around network coding.

A. Throughput

We discuss here how a network coding helps in increasing the throughput. It is so because it allows the routers to compress and modify the data (as in the scenario in Fig. 1). Router generates packets by matching it with neighbor. This type of coding is called inter-flow network coding because the coding is done over packets that differ in their next hop, and thus from different flows.

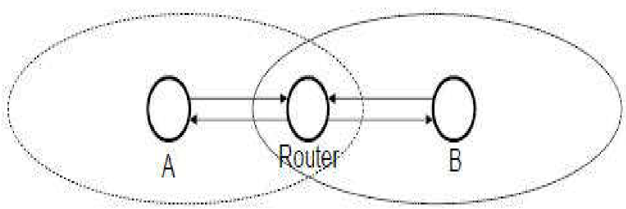


Fig. 1. A simple example showing improve in throughput

1) Redundancy

In the Fig. 1 we consider the scenario that A wants to transmit data to B and B want to transmit data to A. They cannot communicate directly due to coverage problem. So they use router for their communication. In conventional communication A and B will send their data to router. Then router will send data sequentially to both A and B. In this way 4 transmissions are needed.

In network coding, router will perform the X-OR operation on data received from A and B and will send the X-ORed result in single transmissions to both A and B. Decoding will be done at node A and node B as explained in following TABLEs.

TABLE I. X-OR OPERATION AT THE ROUTER

A	B	C=A⊕B
0	0	0
0	1	1
1	0	1
1	1	0

TABLE II. DECODING AT NODE A

A	C	A⊕C=B
0	0	0
0	1	1
1	1	0
1	0	1

TABLE III. DECODING AT NODE B

B	C	B⊕C=A
0	0	0
1	1	0
0	1	1
1	0	1

Assume the data streams to be sent as shown in TABLE I. Router performs the X-OR operation as shown in TABLE I. Let it be C. This data stream is sent to both A and B in a single stream. Decoding at node A to recover B is done by performing X-OR operation between A and C. Similarly decoding to recover A at node B is done by performing X-OR operation. These two operations are shown in TABLE II and TABLE III. This exploits redundancy by compressing two signals into one thus increasing throughput.

2) COPE: An inter-flow network coding protocol

Considering again the example shown in fig. 1,COPE exploits the nature that nodes share data for free, through broadcast in its neighborhood. This scenario is very useful in coding since each node have overlapping reservoir of packets it can use to decode. This is explained in example in Fig. 2.

In this scenario S1 and S2 needs to communicate with D1 and D2 respectively. Due to broadcast nature of wireless media D1 can hear S2 and D2 can hear S1. Router can exploit this by broadcasting X-ORed version to both D1 and D2. This X-topology is very useful in real environment. In real environment there may be lesser nodes which use reverse path for communication as shown in topology of fig.1. But many flows intersect at relay and thus can be coded together. Increase in throughput can be seen in the Fig.3.

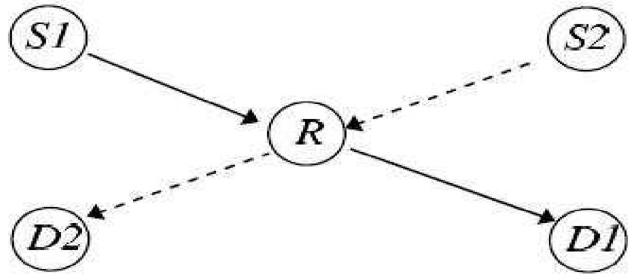


Fig. 2. X Topology

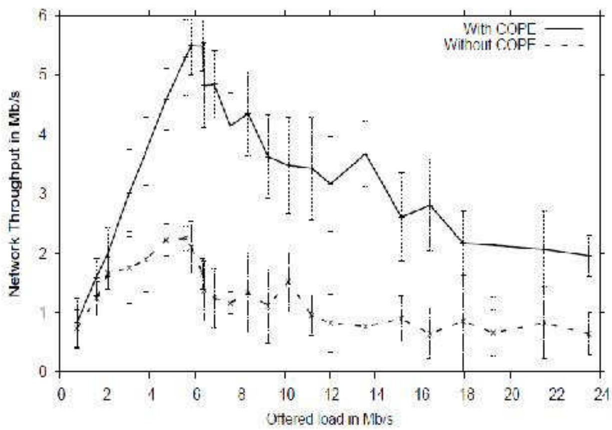


Fig. 3. Several folds increase in throughput

For more general topologies, COPE leads to larger bandwidth savings than are apparent from the above example. It can XOR more than a pair of packets and produce a multifolds increase in the throughput. To summarize, COPE is a MAC extension that has two components:

- Opportunistic Listening: COPE exploits the shared nature of the wireless medium which, for free, broadcasts each packet in a small neighborhood around its path. Each node stores the packets it overhears for a limited period. It also tells its neighbors which packets it has heard by annotating the packets it sends. This creates an environment conducive to coding because nodes in each area have a large and partially overlapping reservoir of packets they can use for decoding.

- Opportunistic Coding: when a node transmits a packet, it uses its knowledge of what its neighbors have heard to deliver multiple packets in a single transmission. The node XORs multiple packets when each intended next hop has enough information to decode the encoded packet.

B. Reliability

The primary means of ensuring reliability in the present architecture is retransmission of lost packets. This works well in wired networks where the bit error rate is very low, but is inefficient over error prone wireless channels. The examples in this section employ intraflow network coding i.e., routers mix packets heading to the same destination. Without coding, a transmitter needs to know which exact packets the destination (or the next-hop) misses so that it can retransmit them. When the network is unreliable, communicating this feedback reliably consumes significant bandwidth. In the presence of coding, no specific packet is indispensable and as a result a transmitter does not need to learn which particular packet the destination misses, it only needs to get feedback from the destination once it has received enough packets to decode the whole file [4].

In network coding R transmits random linear combination of the packets. In particular, R transmits to the destination packets of the form

$$p_i = c_{j1}p_i + \ldots + c_{jn}p_n \qquad (1)$$

$$\begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix} = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \\ c_{n1} & \cdots & c_{nn} \end{pmatrix}^{-1} \begin{pmatrix} p'_1 \\ \vdots \\ p'_n \end{pmatrix} \qquad (2)$$

Once the destination has decoded the packets, it immediately sends R one acknowledgement for the whole file. Since the ack on average needs two trials, and assuming R continues transmitting during that period, R will deliver the file to the destination in 2n + 2 transmissions, on average.

C. Dead Spots

Network coding also helps in dealing with dead spots. Consider the scenario in Fig. 4-(b), where Alice would like to transfer a file to Bob. Bob is not within Alice's radio range, and thus Alice needs the help of an intermediate node. There are 5 nearby wireless nodes which could relay Alice's packets to Bob.

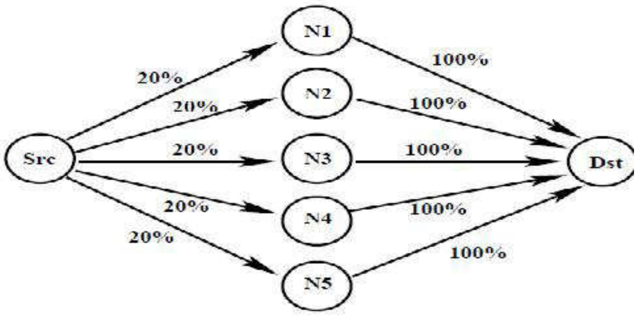


Fig. 4. Need of coordination between intermediate nodes is removed

Unfortunately, Alice is in a dead spot with 80% loss rate to every nearby wireless node. In today's 802.11 networks, Alice will pick the best path to Bob for her transfer. But since all paths are lossy, each packet will have to be transmitted 6 times (5 times form Alice to the relay and once from the relay to Bob).

A better approach would make use of spatial diversity to improve Alice's throughput. Consider scenario as shown in Fig. 4.

Alice broadcasts her packets, and any relay that hears a packet can forward it to Bob. In this case, the probability of delivering a packet to the relay increases from 20% to (1 − 0.85) × 100   67.2%. Thus, on average a packet is transmitted 2.5 times (1.5 times from Alice to the relays and once from the receiving relay to Bob.) This increases Alice's throughput by 2.4x. But using diversity without coding can create a different problem. Multiple relay nodes can hear the same packet and thus attempt to transmit it to Bob[5]. This creates spurious transmissions and wastes the wireless bandwidth.

The combination of spatial diversity and network coding solves the problem. Alice broadcasts her packets. Any relay can participate in forwarding Alice's packets to Bob. To do so, the relay creates random linear combinations of the packets it has received from Alice so far. Specifically, a relay transmits to Bob packets of the form p'i= cijpj is one of Alice's packets, and cj's are random coefficients. If the file contains n packets, then Bob can decode the whole file after receiving any n coded packets, using the matrix inversion in 1. Once Bob decodes the file, he immediately broadcasts an acknowledgment for the whole transfer, causing the relays to stop their transmissions.

D. Mobility

In mobile environments, where the network configuration changes quickly, routing updates are costly. Network coding reduces this cost.

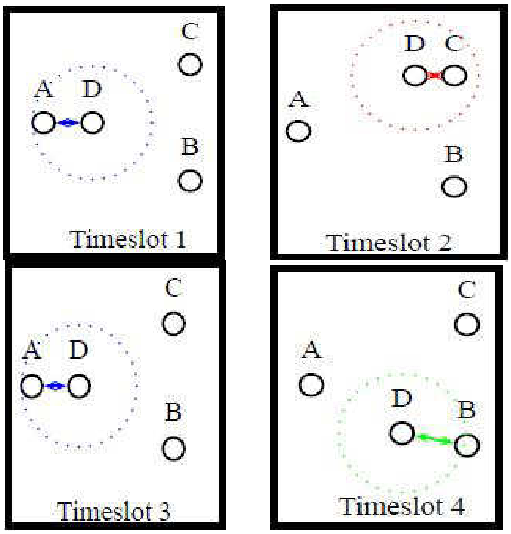


Fig. 5. Node D broadcasts information without knowing which nodes are within its range

A possible random mobility pattern is depicted in Fig. 5. Without coding, when D comes in range with node A, it needs to detect the existence of this new neighbor through some packet exchange. It also need to learn which packets A has already received during its prior interactions with D.[9]

E. Monitoring

The highly stochastic nature of wireless environments makes it desirable to monitor the state of wireless links. In this section, we show how network coding can be exploited to better monitor the link loss rates in a multi-hop wireless network. As a concrete example, let us consider the simple sensor network shown in Fig. 6 and try to estimate the loss rate on all links using only end-to end measurements. This approach is often used in wireless sensor networks [6] to exploit the data aggregation paradigm: data collecting nodes can infer loss rates on internal links by exploiting whether packets from various sensors are received, without the need to actively inject probes. In the example of Fig. 6, let nodes A, B, C, and D be sensors, and nodes E and F be sinks that can easily communicate with each other (e.g., connected through a high-bandwidth link).

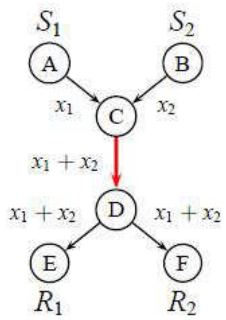


Fig. 6. Example of monitoring

Let nodes A and B send probes x1 and x2 respectively in one time slot; let C XOR the incoming packets and forward x1 + x2 downstream to D; let D broadcast x1 + x2 to both E and F. Each of the nodes E and F may receive x1, x2, x1 + x2, or nothing, depending on the success/failure of the transmission on each link. This is one round of probes and can be repeated several times. By sending several rounds of probes from A and B and by observing the pairs received by E and F at each round, we can infer the link loss rates on all five links. This approach is essentially link loss tomography enhanced with network coding operations. It that the network coding approach has several advantages. Compared to unicast and multicast tomographic techniques, it is able to identify a larger number of links. E.g., in Fig. 6, we are able to identify the loss rate on all five links; in contrast, a multicast probe from A to E and F would not be able to distinguish links A − C and C − D that are connected in tandem; similarly, a unicast probe from A to F would not be able to distinguish links A – C, C – D, D – F on its path.

IV. CHALLENGES

There are some challenges with this network coding technique which are summarized as follows.

A. Challenges of broadcast networks

It is a side product of relying on the broadcast channel, which has implications on MAC, routing, and transport protocols.

The standard access mode of 802.11 and similar MACs is a Distributed Coordination Function (DCF) combining carrier sense multiple-access (CSMA) with collision avoidance (CA). A node sending only if it detects that the channel is clear[7]. It then waits for an acknowledgment from the receiver. If the sender does not receive an acknowledgement within a specific period, it assumes that there was a collision and selects a random back off timer uniformly distributed within a contention window. The contention window doubles for every failed transmission in order to reduce the probability of collisions. In contrast, there is no acknowledgment for broadcast packets. Further, it is unclear how to add this functionality without creating significant complexity and the potential for ack implosion. The lack of acknowledgments for broadcast packets implies that the MAC does not have any congestion avoidance function in such a mode. Hence, the use of an unmodified 802.11 style MAC leads to an increased number of collisions, and a drop in the throughput of the system.

B. Routing

Traditional routing protocols impose a point-to-point shortest path link. However, with broadcast, multiple nodes could simultaneously receive a packet and one or more of them might choose to transmit as a result. This changes the notion of routing from a single shortest path to a multipath problem, where decisions are made after packet reception, not at the time of transmission.

C. Transport

The 802.11 MAC used in broadcast mode does not perform the usual link layer functionality of contention resolution and retransmission of lost packets. For inter-flow coding, the resulting high loss rate needs to be addressed; otherwise it would be mistaken as a sign of congestion by transport protocols like TCP[8], causing them to reduce their sending rate unnecessarily. For intra-flow coding, reliability comes for free, but a different problem surfaces. Coding and decoding involve linear operations over batches of packets. Batching does work well with window-based transport protocols like TCP. Hence, transport protocols would need to be redesigned to be effective in the presence of these changed forwarding semantics.[10]

D. Coding challenges

The coding challenges arise from the desire to combine several attractive properties, such as low complexity, delay and memory requirements, high achievable rates, and adaptability to unknown channel conditions. In general, there is a trade-off between these properties. Network coding requires intermediate nodes in the network to perform operations over finite fields in real-time. While the cost of inter-flow coding is usually low, the general linear codes used in intra-flow coding could be computationally expensive.

E. Real-Time Traffic

Intra-flow coding across n packets implies that a receiver needs in general to collect all n of them before extracting the data. In real-time applications, such as audio and video, the associated delay might be prohibitive for large values of n, which seems to indicate that small values of n would be desirable. Using small values of n on the other hand may not allow us to optimally mix the information and realize the theoretically promised benefits. Thus, there is a tension in balancing these two opposing requirements.

V. CONCLUSION

Network coding enables more efficient, scalable, and reliable wireless network. These opportunities come with a need for rethinking our MAC, routing, and transport protocols. We believe, however, that future research will overcome these challenges and integrate network coding into the wireless network design.

REFERENCES

[1] R. Ahlswede, N. Cai, S. R. Li, and R.W. Yeung. Network Information Flow. In IEEE Transactions on Information Theory, July 2000.

[2] T. Ho, R. Koetter, and M. M´edard.

[3] S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, M. Medard, and M. Effros. Resilient network coding in the presence of byzantine adversaries. IEEE Transactions on Information Theory, 2006.

[4] Y. Mao, F. R. Kschischang, B. Li, and S. Pasupathy. A factor graph approach to link loss monitoring in wireless sensor networks. IEEE SAC, 23.

[5] S. R. Li, R. W. Yeung, and N. Cai. Linear network coding. In IEEE Transactions on Information Theory, 2003. S. R. Li, R. W. Yeung, and N. Cai. Linear network coding.

[6] C. Fragouli, J. WIdmer, and J. LeBoudec. On the benefits of network coding for wireless applications. In NetCod 2006.

[7] S. N. Diggavi, N. Al-Dhahir, A. Stamoulis, and A. R. Calderbank. Great expectations: The value of spatial diversity to wireless networks, 2004.

[8] M. Ghaderi, D. Towsley, and J. Kurose. Reliability benefit of network coding In University of Massachusetts Amherst Technical Report 07-08, 2007.

[9] Wireless Network Information Flow: A determistic approach. By Avestimehr, A.S. Sch. of Electr. & Comput. Eng., Cornell Univ., Ithaca, NY, USA and Diggavi, S.N. ; Tse, D.N.C.Published in Information Theory, IEEE Transactions on April 2011

[10] Practical methods for wireless network coding with multiple unicast transmissions T Aktas, O Yilmaz, E Aktas – published in Communications, IEEE Transactions on march 2013.

Secure Routing Protocol in Heterogeneous Wireless Sensor Network

P. Madhumathy
Anna University
Chennai, India
sakthi999@gmail.com

D. Sivakumar
Dept. of CSE
Easwari Engineering College,
Chennai, India.

Abstract— **Multipath routing is an effective mechanism for fault and intrusion tolerance to improve data delivery in heterogeneous wireless sensor networks. This mechanism will fail, if there is no efficient redundancy management to answer user queries in the presence of unreliable and malicious nodes. Redundancy management means exploiting the tradeoff between energy consumption vs. gain in reliability, timeliness and security maximize the system useful lifetime. In this paper we have explored general security threats in wireless sensor network and made an extensive study to categorize available data gathering protocols and analyze possible security threats on them. Due to limitations of sensor devices, the networks exposed to various kinds of attacks and conventional defenses against these attacks are not suitable due to the resource constrained nature of these kinds of networks. Therefore, security in WSNs is a challenging task due to inheritance limitations of sensors. We represent their problems and methodologies which are used in order to address the problems.**

Keywords— Heterogeneous Wireless Sensor Networks, Multipath Routing, Intrusion Detection, Reliability, Security, Energy Conservation.

I. INTRODUCTION

WSN must not only satisfy the application specific QoS requirements such as reliability, timeliness and security, but also minimize energy consumption to prolong the system useful lifetime. The tradeoff between energy consumption vs reliability gain with the goal to maximize the WSN system lifetime has been well explored in the literature. However, no prior work exists to consider the tradeoff in the presence of malicious attackers.

Using homogeneous nodes which rotate among themselves (CHs) and sensor nodes (SNs) leveraging CH election protocols such as HEED [1] for lifetime maximization has been considered [2]. Using heterogeneous nodes can further enhance performance and prolong the system lifetime. In the latter case, nodes with superior resources serve as CHs performing computationally intensive tasks while inexpensive less capable SNs are utilized mainly for sensing the environment. The tradeoff issue between energy consumption vs. QoS gain becomes much more complicated when inside attackers are present as a path may be broken when a malicious node is on the path. This is especially the case in heterogeneous WSN (HWSN) environments in which CH nodes may take a more critical role in gathering and routing sensing data. Thus, very likely the system would employ an intrusion detection system (IDS) with the goal to detect and remove malicious node. Multipath routing is considered an effective mechanism for fault and intrusion tolerance to improve data delivery in WSNs. It satisfies the energy consumption through turn off the sensor nodes for period of time to save energy. The basic idea is that the probability of at least one path reaching the sink node or base station increases as we have more paths doing data delivery. While most prior research focused on using multipath routing to improve reliability [3], some attention has been paid to using multipath routing to tolerate insider attacks. These studies, however, largely ignored the tradeoff between QoS gain vs. energy consumption which can adversely shorten the system lifetime.

II. ROUTING PROTOCOL

Routing techniques in wireless sensor networks are influenced by two factors. First, it has to deal with hardware and resource constraints. The routing algorithm has to be energy aware, thus minimize the control information flows and communication. Routing table maintenance is limited by memory capacity [1]. Second, the nature of sensor network applications defines traffic patterns, which are different from the traditional ones. In sensor networks, it is not necessary to support communication between any pair of nodes [2],[3], the dominant traffic is one-to-many (base station multicast), many-to-one (data sent to the base station) and local communication between neighbors. As the resources are limited and the number of nodes is large, wireless sensor network usually does not support global addressing, that brings high overhead. It often trade on its data centric character instead and deploys attribute-based addressing. This means the base station sends queries for data with specific properties. However routing technique is Routing is based on Geographic routing. No path information maintained by individual SNs. Location of neighboring node known to a sending node. All nodes receive and maintain location of CH (through voting process).

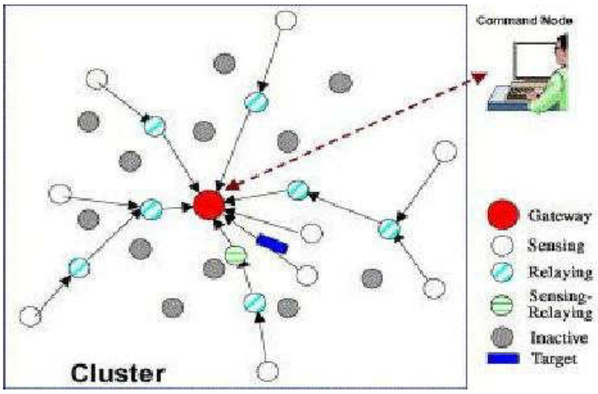


Fig.1. Source and Path Redundancy in HWSN

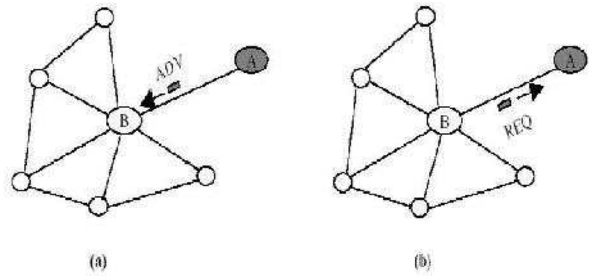


Fig.2.(a) &(b) denotes Node B sends a REQ listing all of the data it would like to acquire.

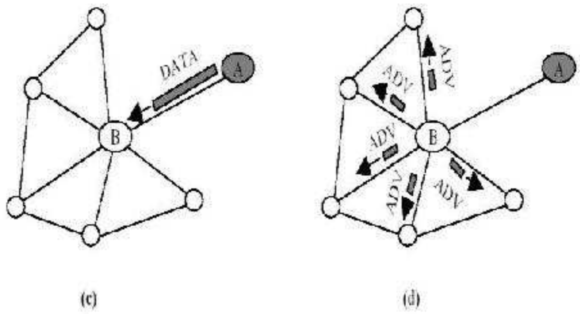


Fig.3. (c)& (d) If node B had its own data, it could aggregate this with the data of node A and advertise.

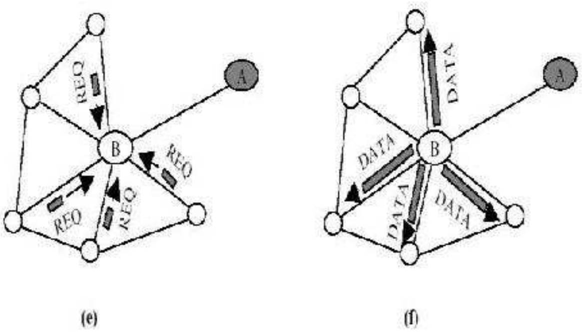


Fig.4.(e)&(f) Nodes need not respond to every message

III. SECURE MULTIPATH PROTOCOL IN WSN

To this category are included geographic routing algorithms like Geographic and Energy Aware Routing (GEAR) or Geographic Forwarding (GF).

Attacks on routing since the concept of sensor networks originates from the wireless ad-hoc networks, many attacks on wireless ad- hoc networks can be adapted for sensor networks. Sybil attack is such an example [4]. Karlof and Wagner [3] show another types of attacks and furthermore they propose two novel attacks – HELLO floods and sinkholes. Denial of Service attacks on sensor networks are studied by Stankovic and Wood [5].We present a brief summary of major attack classes here. We presented selected findings about secure multipath routing protocols which have been proposed for wireless sensor network in this subsection. Problems that protocols want to be solved are represented. And the methodology of particular protocol has been mentioned in order to know how that secure multipath routing protocol addresses the problem. In this subsection, six selected approaches will be reviewed and it is organized based on the problem and methods which try to address the problem. The problem of lack of security in SPIN protocol S-SPIN is a kind of negotiation base routing protocol which uses MAC for integrity check and correctness of the packets. As confidentiality is application dependent, the founder of S-SPIN did not identify the method of message confidentially. There are three stages in S-SPIN: ADV, REQ and DATA. When a node gains new data, it will send ADV message to its neighbors. Once a node receives ADV message from its neighbor, it checks and verifies the ADV message through the MAC. If verification process has been achieved, then receiver will send the REQ message to source. REQ will be verified at the source again through the MAC and in the case of success, DATA message which contains encrypted data will be sent to destination. To prove security of S-SPIN, its founders use the formal security framework proposed by Acs et al [3]. They prove that S-SPIN is secure if MAC scheme is secure against existential forgery attack. The basic concept of mentioned mechanism is simple: if node A forwarded the message, it saves it in its buffer and waits for the acknowledgment. If

node A can hear the acknowledgment from destination in the certain amount of time, then the receiver node is secure, otherwise it will be a malicious node. Therefore in order to make a secure path, node A selects another node (route) and forwards the packet to

it and waits for acknowledgment. This process continues until the packet reach to destinations. Many of the new flooding and gossiping routing protocol that proposed for wireless sensor network do not consider the security issue. How to have secured flooding or gossiping r The probability model to estimate the MTTF of a HWSN using multipath data forwarding to answer queries issued from a mobile user roaming in the HWSN area. To find best redundancy level (m_s, m_p) that maximizes MTTF, while satisfying query reliability (R_{req}) and timeliness (T_{req}) , requirements .

$$R_q > R_{req}$$

Implicitly satisfies timeliness (T_{req}) requirement. Maximum Number of queries that can be answered before queries. The probability that the first i queries are successful but the (i+1) the query failure. The expected number of queries that the system can answer without experiencing the failure with R_q Finally MTTF as the probability weighted average of the Number of queries can handle without experiencing a without experiencing any deadline, transmission, or security failure

$$\text{MTTF}= \sum_{i=1}^{N_q-1} iR_q^i\left(1-R_q\right)+N_qR_q^{N_q} \quad (1)$$

MTTF formulation is that to deduce the maximum number of queries,N_a are processed successfully without any failure for which the system will have the longest lifetime span. Energy consumption is to estimate through amount of energy consumed by transmission and reception over wireless link.

Where, E_{elec} Energy to run the transmitter and receiver circuitry (J/bit). E $_{amp}$ is Energy used by the transmit tamplifier to achieve an acceptable signal to noise ratio (J/bit/m2). r^2 is energy loss due to channel transmission, Energy to receive a message .For transmission and reception energy consumption of sensors, adopt energy model in CH and SN. Lastly, for intrusion detection every node is evaluated by an *m* voters and the knowledge of N_q to calculate the system MTTF given by equation.

The three-tier security scheme provides better network resilience against mobile sink replication attack compared to the single polynomial pool approach. This scheme delivers the same security performance as the single polynomial pool approach when the network is under a stationary access node replication attack. In both schemes, for any sensor node u that needs to authenticate and establish a pair-wise key with a stationary access node A, the two nodes must share at least a common polynomial in their polynomial rings. To perform a stationary access node replication attack on a network, the adversary needs to compromise at least a single polynomial from the static pool. This can be obtained easily by capturing arbitrary sensor nodes in the network.

Then, the adversary can make use of this compromised polynomial by a replicated stationary access node to enable insecure access to the network. When successful access to the network has been obtained through the compromised static polynomial, the replicated stationary access node transmits recorded mobile sink data request messages. Next, the sensor nodes that have the compromised polynomial in their rings will insecurely authenticate and establish a pair-wise key with the replicated node and thus deliver their data to the replicated node.

V. PERFORMANCE EVALUATION FOREXISTING SYSTEM

In exiting system, the optimum communication range and communication mode were derived to maximize the network lifetime. Voting based distributed intrusion detection is applied to remove malicious nodes from the HWSN .In some clustering algorithm, unsatisfactory cluster formations, which may cause the network to suffer from load imbalance or extra energy consumption.

Voting-based Clustering Algorithm (VCA) for energy-efficient data dissemination in wireless sensor networks. This new approach lets sensors vote for their neighbors to elect suitable cluster heads. VCA is completely distributed, location unaware and independent of network size and topology. It combines load balancing, energy and topology information together by using very simple voting mechanisms. Simulation results show that VCA can reduce the number of clusters by 5-25% and prolong the lifetime of a sensor network by 10-30% over that of existing energy-efficient clustering protocols

IV. PROPOSED SYSTEM

To explore more extensive malicious attacks in addition to packet dropping and bad mouthing attacks and investigate intrusion detection and multipath routing based tolerance protocols to react to these attacks The Effects of T_{IDS} on MTTF under low capture rate and high capture rate is shown in Fig (2) Another direction is to consider targeted attacks, capture certain strategic nodes with higher probability and malicious behavior and collide with other attackers to avoid intrusion detection

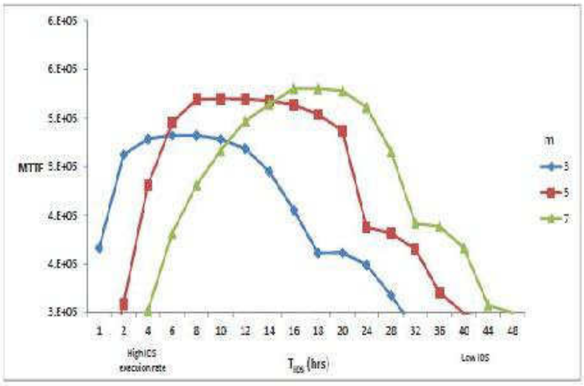


Fig.2 Effect of T_{IDS} on MTTF under low capture rate

REFERENCES

[1] F.Bao, I. R. Chen, M. Chang, and J. Cho, "Hierarchical trust management for wireless sensor networks and its applications to trust-based routing and intrusion detection," IEEE Trans. Netw. Service Manag.,vol. 9, no. 2, pp. 161–183, 2012.

[2] E. Stavrou and A. Pitsillides, "A survey on secure multipath routing protocols in WSNs," Comput. Netw., vol. 54, no. 13, pp. 2215–2238,2010

[3] T.shu, M. Krunz, and S. Liu, "Secure data collection in wireless sensor networks using randomized dispersive routes," IEEE Trans. Mobile Comput., vol. 9, no. 7, pp. 941–954, 2010.

[4] I. R. Chen, A. P. Speer, and M. Eltoweissy, "Adaptive fault-tolerant QoS control algorithms for maximizing system lifetime of query-based wireless sensor networks," IEEE Trans. Dependable Secure Computing,vol. 8, no.2,2011

[5] J. H. Cho, I. R. Chen, and P. G. Feng, "Effect of intrusion detection on reliability of mission-oriented mobile group systems in mobile ad hoc networks," IEEE Trans. Reliab., vol. 59, no. 1, pp. 231–24

Optimizing Cluster Formation Based Routing Models for MANETs

Amandeep Kaur
Punjab Technical University
Jalandhar, India
adkresearch81@yahoo.com

***Abstract*—Over the last few years, there have been several improvements in optimising the complex models of routing for MANETs. MANET researchers are handling various challenges to improve network scalability, link connectivity, and broadcasting. To address these issues, the throughput of the network can be improved by efficient use of clustering. There are various improved formulations and enhancements to handle the above issues; we present a detailed analysis to discuss parameter requirements to optimise clustering problem for constructing hierarchical network. The primary objective of this paper is to present a analysis of different formulations of the optimized clustering problem in MANETs.**

Keywords—MANET, Cluster, Scalability, Routing, Adaptability, Optimization.

I. INTRODUCTION

Mobile Ad Hoc Networks (MANETs) are wireless networks consisting of entirely mobile nodes that communicate without using base infrastructure. Nodes in these networks act as routers, as well as communication end-points. Rapid changes of connectivity, network partitioning, higher error rates, collision interference, and bandwidth and power constraints together pose new challenges for this type of networks. In recent years, sensor networks have also received significant interest from the research community.

MANETs are unique in communications in that no base infrastructure is required for connectivity. They are autonomous group of mobile users that communicate over reasonably slow wireless links. Topologies are consequently dynamic in that mobile nodes can arrival and depart the network in an arbitrary fashion. The nodes must behave as transmitters, receivers, and routers (Layer 2 for nodes on the same subnet and Layer 3 for nodes on different subnets). In an ad hoc network, mobile nodes communicate with each other using multi-hop wireless links. Designing and implementing these MANETs can be quite demanding and complex due to the nature of wireless communications. Adverse effects, such as channel capacity variations, routing failures, and erratic power control play havoc with these designs and implementations. In an effort to optimize QoS parameters, customer expectations can be realized in the most practical and efficient manner possible. Actor nodes are devices which typically have more processing, memory and communication power and energy. Wireless sensor and actor networks can get information from the environment, interpret it and take actions based on the interpretation results. The challenges include designing ad hoc routing protocols that can perform well in the following two aspects: *Connectivity*: The protocol can swiftly identify feasible routes for connecting peer ad hoc network hosts and *Link Quality*: The protocol is able to select links that have high bandwidth to form the routes and, more importantly, to rebuild the routes when the channel quality of some links has already deteriorated. A review of these secure routing protocols for mobile ad hoc wireless networks indicates that most of the protocols presume the existence of a centralized or distributed trusted third party in the network. This assumption is also coupled with reconfiguration of nodes with encryption keys prior to joining the network.

Routing protocols for mobile ad hoc networks (MANETs) have been explored extensively in recent years. Routing in general consists of two entities, namely the routing protocol and the routing algorithm. Routing complexity includes various issues; few of them are:

- Decentralized control
- Unpredictable channel
- Data Loss
- Route Maintenance
- Node mobility
- Adaptability
- Effectiveness
- Scalability
- Adequate security

Clustering involves the creation of a hierarchical network where the network is divided into clusters, with certain nodes in each cluster being chosen to be cluster heads. The process of selecting which nodes would be best suited to be cluster heads and which regular nodes should be assigned (connected) to which cluster head is known as the clustering problem as shown in Figure 1.

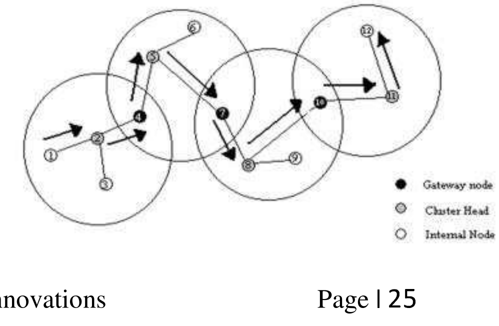


Fig 1. Cluster Formation

There are two levels of communications (Figure 2) in clustering method: Intra-cluster that do not use CS technique and inter cluster that use CS technique. The data size in inter-cluster transmissions is the same as the data in intra-cluster transmissions. Thus, reducing the number of transmissions can effectively reduce the energy consumption of mobile nodes. For intra-cluster transmissions, the nodes transmit their data to the CH following the shortest path routing (in terms of number of hops). For inter-cluster transmissions, a minimal cost (in terms of number of hops) backbone tree is created that connects all CHs and transmit the data projections along the backbone tree.

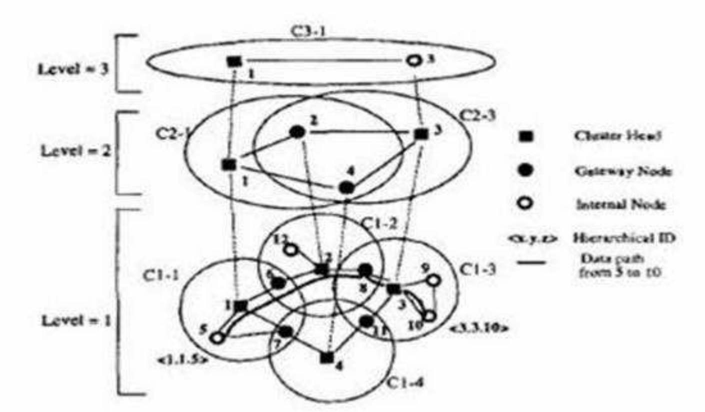


Fig 2. Level of clustering

The rest of this paper is organized as follows. In Section II, we review the related work. In Section III, the detail comparison is presented. Finally, we conclude this paper.

II. MANET ROUTING

Recently, there have been several attempts to design ad hoc routing protocols that meet the above challenge [6], [15], [16], [21], [26].

AODV (Ad hoc On-Demand Distance Vector Routing Protocol) is a reactive routing protocol designed by Charles E. Perkins and Elizabeth M. Royer [26]. This protocol uses four types of control messages in the aim to send data packets. The first type is HELLO messages. This type of messages, exchanged periodically to maintain a neighborhood base. RREQ, RREP, RRER, are used to establish a path to destination when any node wants to send a data. This number of control packet has a signified effect of the waste of resources. To overcome the problem of energy consumption in this protocol, authors designed a new solution that reduces the HELLO messages number exchanged and to include the factor of energy consumption that will be useful later for the routing messages. Firstly, they minimized the exchange number of Hello messages. Secondly, they replaced the regular periodic instant of sending hello message by another proportionally to the energy stored in the battery of the node. The node receiver of this hello message, do the inversely action to extract information proportionally to the node sender energy, and the same information enclosed in hello message.

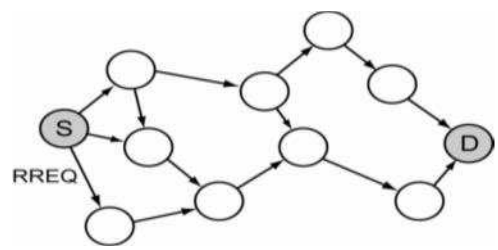


Figure 3(a). Route Request

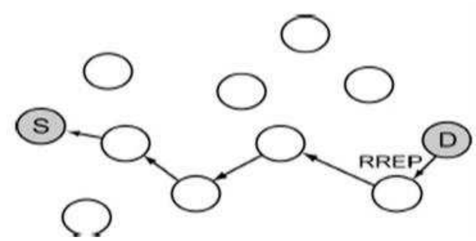


Fig 3(b).Route Reply

The primary objective of this paper is to present different formulations of the optimized clustering problem in MANETs. LEACH is one of the most famous hierarchical routing protocols, which can guarantee network scalability and prolong network lifetime than other ordinary routing protocols. The energy can be well balanced among nodes since each node takes turn to become the cluster head at different rounds. However, the cluster head nodes are randomly chosen and the cluster heads use direct transmission to send their data.

LEACH [24] is fully distributed and its data transmission delay is very small. However, the algorithm take the assumption that all the Cluster Head can directly communicate with the sink, so the assumption may not be practical, and then it cannot be much suitable for large-scale applications. At the same time, this approach of dividing clusters may bring additional costs and overlay issues.

Taking account of the overhead of dividing clusters, PEGASIS [22] makes an improvement on LEACH by constructing a node chain instead of cluster group. In all sensor nodes, only one chose node as a gateway to communicate with sink, the other nodes in the chain take turn to be gateway. After receiving message, each node takes an aggregation with its own sensing data. At last those data are transmitted to gateway node. Although PEGASIS performs better than LEACH by eliminating the overhead of dynamic cluster formation, because transmission is asynchronous, the time of transmission will be prolonged too much. Hierarchical-PEGASIS conducts a further improvement; it allows concurrent transmission when the nodes are not adjacent.

A (SEAD) [23] protocol was proposed to minimize energy consumption in both building a dissemination tree and disseminating data to mobile Scalable Energy-efficient Asynchronous Dissemination sinks. When the sink joined the tree, the Steiner tree was built recursively and SEAD found the minimal cost entry to the tree for the sink using unicast. Gandham [4] tried to use an ILP (Integer Linear Program) to determine the locations of multiple base stations. They aimed at minimizing the energy consumption per node and prolonging the network longevity.

Younis and Fahmy [21] propose a hybrid energy-efficient distributed clustering routing (HEED) protocol where the CHs are probabilistically selected based on their remaining energy and the sensors join clusters such that the communication cost is minimized. HEED assumes a multihop connection between the CHs and to the sinkIn EAP, each CH is probabilistically selected based on its ratio of the remaining energy to the average remaining energy of all the neighbor sensors within its cluster range. This is in contrast to HEED that only chooses CHs based on a sensors' own remaining energy. To further extend network lifetime, EAP introduces the idea of "intracluster coverage" that allows a partial set of sensors to be active within clusters while maintaining an expected coverage.

The Information Fusion-based Role Assignment (InFRA) [25] algorithm builds a cluster for each event including only those nodes that were able to detect it. Then, cluster-heads merge the data within the cluster and send the result toward the sink node. The InFRA algorithm aims at building the shortest path tree that maximizes information fusion. Thus, once clusters are formed, cluster-heads choose the shortest path to the sink node that also maximizes information fusion by using the aggregated coordinators distance. A disadvantage of the InFRA algorithm is that for each new event that arises in the network, the information about the event must be flooded throughout the network to inform other nodes about its occurrence and to update the aggregated coordinators-distance. This procedure increases the communication cost of the algorithm and, thus, limits its scalability.

III. CONCLUSION

In recent years, the routing protocol in MANETs has become one of the most important research area, and there have been existed a large number of research achievements. In this paper, we make a great deal of analysis and research, and classify the Hierarchical-based routing (Clustering). Nevertheless, there still exist a series of challenges for routing protocols in MANETs.

The following parameters describe some of those issues that must be considered for optimizing the cluster based approach:

• Effectiveness: how to effectively utilize bandwidth and energy for specific application; how to efficiently divide the whole networks into clusters and coordinate the workloads of all nodes.

• Adaptability: how to adapt the mobile networks and make sensor nodes self-organizing and self reconfigurable.

• Scalability: how to satisfy dense networks with a large number of nodes and try to prolong the lifetime.

• Security: how to make routing protocols secure in MANETs and assure that the transmitted messages are not eavesdropped and tampered.

As our study reveals, it is not possible that a routing algorithm is suitable for all scenarios and for all applications. Although many routing protocols have been proposed in MANETs, many issues still exist and there are still many challenges that need to be resolved in optimizing the existing Cluster Classification Models.

REFERENCES

[1] H. Yang, H. Luo, F. Ye, S. Lu, and L. Zhang, "Security in Mobile Ad Hoc Networks: Challenges and Solutions," IEEE Wireless Comm., vol. 11, no. 1, pp. 38-47, Feb. 2004.

[2] P. Sakarindr and N. Ansari, "Security Services in Group Communications Over Wireless Infrastructure, Mobile Ad Hoc, and Wireless Sensor Networks," IEEE Wireless Comm., vol. 14, no. 5, pp. 8-20, Oct. 2007.

[3] A.M. Hegland, E. Winjum, C. Rong, and P. Spilling, "A Survey of Key Management in Ad Hoc Networks," IEEE Comm. Surveys and Tutorials, vol. 8, no. 3, pp. 48-66, Third Quarter 2006.

[4] L. Zhou and Z.J. Haas, "Securing Ad Hoc Networks," IEEE Network Magazine, vol. 13, no. 6, pp. 24-30, Nov./Dec. 1999.

[5] L. Zhou, B. Cchneider, and R. Van Renesse, "COCA: A Secure Distributed Online Certification Authority," ACM Trans. Computer Systems, vol. 20, no. 4, pp. 329-368, Nov. 2002.

[6] H. Chan, V. Gligor, A. Perrig, and G. Muralidharan, "On the Distribution and Revocation of Cryptographic Keys in Sensor Networks," IEEE Trans. Dependable and Secure Computing, vol. 2, no. 3, pp. 233-247, July 2005.

[7] P. Yi, Z. Dai, Y. Zhong, and S. Zhang, "Resisting Flooding Attacks in Ad Hoc Networks," Proc. Int'l Conf. Information Technology: Coding and Computing, vol. 2, pp. 657-662, Apr. 2005.

[8] B. Kannhavong, H. Nakayama, A. Jamalipour, Y. Nemoto, and N. Kato, "A Survey of Routing Attacks in MANET," IEEE Wireless Comm. Magazine, vol. 14, no. 5, pp. 85-91, Oct. 2007.

[9] H. Nakayama, S. Kurosawa, A. Jamalipour, Y. Nemoto, and N. Kato, "A Dynamic Anomaly Detection Scheme for Aodv-Based Mobile Ad Hoc Networks," IEEE Trans. Vehicular Technology, vol. 58, no. 5, pp. 2471-2481, June 2009.

[10] J. Newsome, E. Shi, D. Song, and A. Perrig, "The Sybil Attack in Sensor Network: Analysis & Defenses," Proc. Third Int'l Symp. Information Processing in Sensor Networks, pp. 259-268, 2004.

[11] J.Stine and G.de Veciana, "A improved paradigm for QoS in wireless ad hoc networks using synchronous signaling and node state", IEEE J.Select.Areas Commun., vol.17, pp.1454-1465, 1999.

[12] J.Biswas, M.Barai, S.K.Nandy, "Energy Efficient hybrid QoS routing protocol for ad-hoc wireless networks, local computer networks", in:

29th Annual IEEEE International Conference on Nov.2004, pp.180-187.

[13] R. Dube, C.D. Rais, and S.K. Tripathi, “Improving NFS Performance over Wireless Links,” IEEE Trans. Computers, vol. 46, no. 3, pp. 290-298, Mar. 1997.

[14] R. Dube, K.-Y. Wang, C.D. Rais, and S.K. Tripathi, “Signal Stability Based Adaptive Routing (SSA) for Ad Hoc Mobile Networks,” IEEE Personal Comm., vol. 4, no. 1, pp. 36-45, Feb. 1997.

[15] J.J. Garcia-Luna-Aceves and S. Murthy, “A Path-Finding Algorithm for Loop-Free Routing,” IEEE/ACM Trans. Networking, vol. 5, no. 1, pp. 148-160, Feb. 1997.

[16] T. Goff, N. Abu-Ghazaleh, D. Phatak, and R. Kahvecioglu, “Preemptive Routing in Ad Hoc Networks,” Proc. MOBICOM ’01, pp. 43-52, July 2001.

[17] H.-Y. Hsieh and R. Sivakumar, “Routing: On Using the Ad Hoc Network Model in Cellular Packet Data Networks,” Proc. ACM MOBIHOC ’02, pp. 35-47, June 2002.

[18] Broch, D.A. Maltz, D.B. Johnson, Y.-C. Hu, and J. Jetcheva, “A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols,” Proc. MOBICOM ’98, pp. 85-97, July 1998.

[19] X.-H. Lin, Y.-K. Kwok, and V.K.N. Lau, “RICA: A Receiver-Initiated Approach for Channel-Adaptive On-Demand Routing in Ad Hoc Mobile Computing Networks,” Proc. 22nd IEEE Int’l Conf. Distributed Computing Systems (ICDCS ’02), pp. 84-91, July 2002.

[20] Y.-K. Kwok and V.K.-N. Lau, “A Novel Channel-Adaptive Uplink Access Control Protocol for Nomadic Computing,” IEEE Trans. Parallel and Distributed Systems, vol. 13, no. 11, pp. 1150-1165, Nov. 2002.

[21] O. Younis and S. Fahmy, “HEED: A hybrid, energy-efficient, distributed clustering approach for ad-hoc sensor networks,” IEEE Trans. Mobile Comput., vol. 3, no. 4, pp. 366–379, Oct.–Dec. 2004.

[22] M. Aravindhan ,” A Peculiar Cluster Based Routing in Wireless Sensor Network “ ,International Journal of Computer Science and Mobile Computing, Vol.2 Issue. 12 , December – 2013 , pg. 376 - 382

[23] Yih-Chun Hu, “SEAD: secure efficient distance vector routing for mobile wireless ad hoc networks”, Adhoc Networks Volume 1, Issue 1, July 2003, Pages 175–192.

[24] LI Yan,ZHANG Xi-huang,LI Yan-zhong , “Low Energy Adaptive Clustering Hierarchy protocol(LEACH)”, Proceeding of the Hawaii International Conference on System Sciences, Hawaii, January 2000.

[25] Mo Xiaoyan, "Study and Design on Cluster Routing Protocols of Wireless Sensor Networks", Dissertation, Hang Zhou, Zhe Jiang university,2006.

[26] Perkins C, Royer E. Ad hoc On-demand Distance Vector Routing (AODV). July 2003, RFC 3561.

Node Failure Investigation in Zigbee Sensor Network

Amritpal Kaur
Deptt. of ECE
SBSSTC, Ferozepur, Punjab, India
gurdevgill2123@gmail.com

Jaswinder Kaur
Deptt. of ECE
SBSSTC, Ferozepur, Punjab, India
jaswinder.ece@gmail.com

Gurjeevan Singh
Deptt. of ECE
SBSSTC(Polywing), Ferozepur, Punjab, India
gurjeevansandhu@gmail.com

Abstract: This research work is implemented in ZigBee using IEEE 802.15.4 protocol stack, it is most widely used technique in wireless sensor network for low rate wireless personal area network. Hybrid topologies are designed by using the possible combination of ZigBee network topologies and then analyze the affect of router and end devices failure. The performance of the network is evaluated by using parameters: throughput, delay, data dropped and data traffic receive and sent. The results quantify that the combination of star and tree topologies gives good response and also effective to operate the network in worst condition.

Keywords— Hybrid Network Topology, Zigbee Wireless Sensor Network, Router and End Device Failure.

I. INTRODUCTION

Wireless sensor networks (WSN) are regularly used for real-time applications, such as environment examination, therapeutic care, and automobile traffic control. The sensor nodes have depict some source limits, which are not withstanding accurate under these circumstances. WSNs have to afford an unwavering coverage of the area of curiosity and also to get together rigorous time control activities [1]. Zigbee is a wireless personal area network based on IEEE 802.15.4 wireless protocol. Zigbee network defined first in 2004 and released in 2006. The second stack of the Zigbee network was defined as Zigbee 2006. It provides short distance communication with low complexity, low data rate and low power consumption. It is a two way technology which pointed to Wireless Sensor Network (WSN) [2]. Furthermore, it has several advantages such as self organization, smaller size of protocol stacks, and larger addressing space. Most commonly Zigbee also used in the medical field for patient monitoring or health and added together with self-care and self-management technologies can enhance their health outcomes [3]. The main aim of this technology is remote control and sensor applications, which is appropriate to operate in ruthless radio environments and isolated locations. IEEE 802.15.4 based WSN defines the physical and MAC layers. It use CSMA / CA mechanism and solves the problem of channel access [4]. The MAC layer of IEEE 802.15.4 standards operates in two modes, they are beacon enabled and non-beacon enabled mode. In beacon enabled mode, beacon messages are transmitted periodically for network organization and management. Beacon enabled are synchronized and allows the mode to operate on slotted Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) mechanism. In non-beacon mode nodes are not synchronized, because periodic beacon transmissions are absent. Therefore, this mode supported to unslotted CSMA/CA mechanism [5]. Zigbee protocol stack consists of three layers. The upper layer is application layer (APL), which provide interface between the Zigbee system to its end users and also defines the device functionality. The basic function of this layer is to convert the input into digital data, and/or converts digital data into output. The middle layer known as the network layer (NWK) is responsible for network structure, routing, and security such as encryption, key management, and authentication. The lower layer is the physical layer and functionality of this layer is to transmit or receive data using certain radio channels with specified modulation and spreading techniques [6, 7].

In [8] the author only analyzes the performance of tree topology in case of node failure. This paper investigate the performance of hybrid network which is implemented by the possible combinations of ZigBee routing schemes and then analyze the affect of router and end device failure on the performance parameter of networks.

II. SYSTEM MODEL

There are four Zigbee releases, which are ZigBee 2004 (released in December 2004), ZigBee 2006 (released in December 2006), and finally ZigBee 2007 and ZigBee Pro (released in October 2007). On each new release there is an improvement according to the previous releases and new releases have backward compatibility with the old releases. The Zigbee standard is based, at the first two layers of the ISO/OSI stack, on the IEEE 802.15.4 standard. IEEE 802.15.4 is a standard for wireless Personal Area Networks (PANs), which provide low data rate at short communication range with low cost. It also uses a non-persistent Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) Medium Access Control (MAC) protocol. An optional acknowledgement (ACK) message is required to confirm the successful delivery of a data packet. In case of ACK messages, the access mechanism of the non-persistent CSMA/CA MAC protocol is modified to some extent [9, 10]. The ZigBee protocol stack consists of four main layers: the application (APL) layer, the network (NWK) layer, the medium access control (MAC) layer, and the physical (PHY) layer. While the NWK and APL layers of the ZigBee protocol are defined by the ZigBee specification, the PHY and MAC layers are defined by the IEEE 802.15.4 standard. The application layer provide effective interface between the system and the end user. A framework for communication and applications in the network are provided by the application layer. The functionalities of network layer include maintenance, providing and organizing routing over a multihop network, route discovery. The medium access control (MAC) enables the transmission of MAC frames through the use of the physical channel. Besides the data service, it offers a management interface and itself manages access to the physical channel and network beaconing. It also controls frame validation, guarantees time slots and handles node associations. Finally, it offers hook points for secure services. The physical layer (PHY) ultimately provides the data transmission service, as well as the interface to the physical layer management entity, which offers access to every layer management function and maintains a database of information on related personal area networks [11]. The security mechanism of Zigbee includes Data Encryption, Sequential Freshness, Frame Integrity Checking Function, and Entity Authentication Service. Data Encryption is implemented with a symmetric cipher to protect data from being read by the parties without cryptographic key. ZigBee adopts AES-128 key for security [12]. The Frame Integrity Checking function use Message Integrity Code (MIC) in the each data frame to protect data from hackers or by those parties, which do not have cryptography key. Entity Authentication Service provides a secure means for a device to synchronize information with another device simultaneously based on a shared key [13]. There are three different devices supported by ZigBee, which are ZigBee Routers (ZR), ZigBee Coordinator (ZC), and ZigBee End Devices (ZED). ZRs are responsible from performing the IEEE 802.15.4's tasks for routing the packets. ZCs coordinate and manage all devices in the network. ZEDs are simple devices and do not have any routing capabilities [9]. According to their functionality, these devices can be categorized into full function devices (FFDs) and reduced function devices (RFDs). FFDs are able to forward frames for other devices and start up the network as the coordinator of the PAN. A coordinator can periodically broadcast beacon frames by using slotted CSMA/CA, so that nearby RFDs can discover it and join the PAN. The default addressing method for ZigBee networks is Distributed Address Assignment Mechanism (DAAM). DAAM reserves a unique address for each possible location with a given setting of the topological parameters. In DAAM the ZC/ ZR can locally allocate addresses to its children with global setting of the topological parameter and the knowledge of its own depth value [14]. In ZigBee, a device can obtain a network address from the coordinator or a router and join a network successfully. In case of network formation, the coordinator determines the maximum number of children of a router (*Cm*), the maximum number of child routers of a router (*Rm*), and the depth of the network (*Lm*). Note that a child of a router can be a router or an end device, so *Cm* ¸ *Rm*. Cm, Rm, and Lm parameters are used to calculate nodes' network addresses. While these parameters facilitate address assignment, they also prohibit a node from joining a network [15]. Three main types of network topology are considered in IEEE 802.15.4, namely, the star, generic mesh and cluster tree topology. In the star topology, a FFD takes up the role of the PAN coordinator; the other nodes communicate with the PAN coordinator. In generic mesh topology, a FFD can communicate with other nodes within its radio range and outside of its radio coverage through an intermediate device by forming a multi-hop network. In the tree-topology nodes associated to a single PAN coordinator are arranged in parent–child relationships by establishing a tree [16].

III. SIMULATION SCENARIO

Zigbee has been define the three routing schemes i.e. star, tree and mesh but in this paper the performance is analyzed by making three hybrid topologies by using possible combinations of the different routing schemes. So the hybrid topologies used to analyze under different network configurations are star-tree (ST), star-mesh (SM) and mesh-tree (MT). The hybrid topologies are designed by using two PAN coordinators in office scale network and two different topologies are assigned to these coordinators. The affect of device failure has been analyzed for these new hybrid networks using different network parameters. The main intention of this work is to quantify the performance of hybrid topologies better than the network containing single coordinator.

The simulation of Zigbee network has been done using OPNET 14.5 and presents the preliminary simulation result to illustrate the attributes of hybrid network in case of node failure. The devised system distributes different Zigbee devices in an area (an office network scale) of (100m x 100m). The performance of these networks has been examined under different network configuration as shown in Table 1.

Table 1. Simulation Parameters

Network Scale	100 m*100m
Number Of Nodes	50
Network Type	Mixed
Mobility Model	Random Waypoint (Record Trajectory)
Speed of Mobile Nodes	2 m/s
Pause Time	150 s
Simulation Duration	300 s

There are three networks, in each network there are two ZC. In case of MT network, tree topology assign to one coordinator and mesh topology assigned to another one. In case of ST network, star topology assign to one coordinator and tree topology assigned to second. In case of SM network, mesh topology assign to one coordinator and star topology assigned to another coordinator. Using Table1. nine scenarios has been designed. First, second and third scenarios analyze the performance of hybrid topologies under ZR failure. Fourth, fifth and sixth scenarios analyze the performance of hybrid topologies under ZED and ZR failure at same instant of time. Seventh, eighth and ninth scenarios analyze the performance of hybrid topologies under ZED failure.

IV. RESULTS and DISCUSSIONS

Following results are used to enumerate the performance of network.

A. Throughput

Throughput is the ratio of the total amount of data that a receiver receives from a sender to a time it takes for receiver to get the last packet. Throughput is the data quantity transmitted correctly starting from the source to the destination within a specified time (seconds). A low delay in the network translates into higher throughput. Throughput is quantified with varied factors including packet collisions, obstructions between nodes and the type of used topology. The results shown in fig. 1,2 and 3 discovered that maximum throughput achieved by ST network and minimum by MT network in all cases of failure. But throughput of ST and MT network degrades when ZED and ZR fail at similar time, whereas SM network mostly affected by the failure of ZR. As the MT network contain more number of routers and less number of end devices than SM and ST networks.

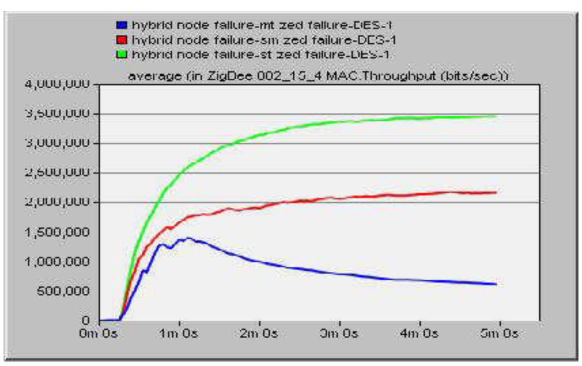


Fig 1. Throughput in MT, SM and ST Topologies in Case of ZED Failure

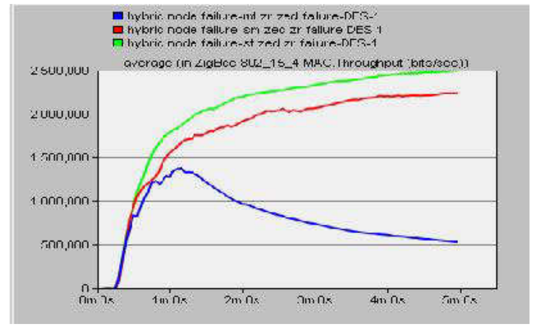


Fig 2. Throughput in MT, SM and ST Topologies in Case of ZED and ZR Failure

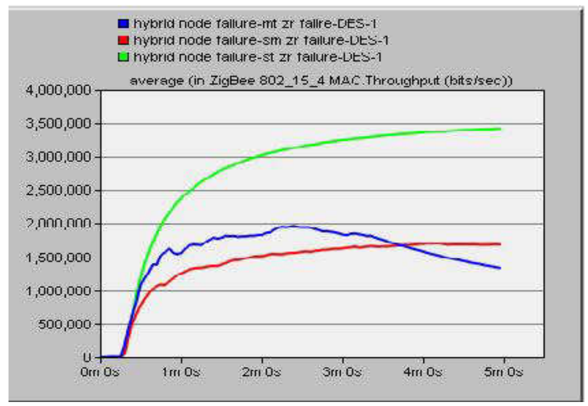


Fig 3. Throughput in MT, SM and ST Topologies in Case of ZR Failure

B. Data Dropped

This statistic records the total amount of data that was received from the upper layer and then dropped by all nodes in the network due to repeatedly failed retransmissions (i.e., exceeded the corresponding short retry or long retry threshold value). Maximum data dropped by MT network and minimum by SM. The ZED individual failure and ZED, ZR both failure affect the network in similar manner (referred to fig. 4 and 5), whereas ZR individual failure show less affect then other case as shown in fig 6.

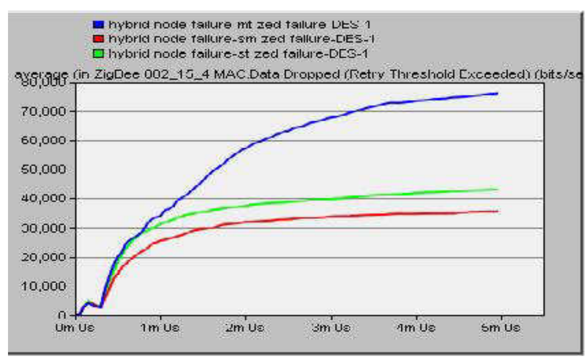


Fig 4. Data Dropped in MT, SM and ST Topologies in Case of ZED Failure

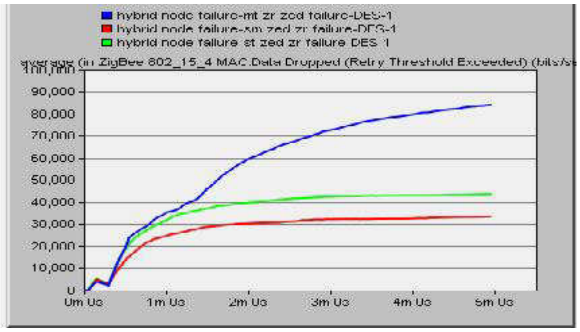


Figure5. Data dropped in MT, SM and ST Topologies in Case of ZED and ZR Failure

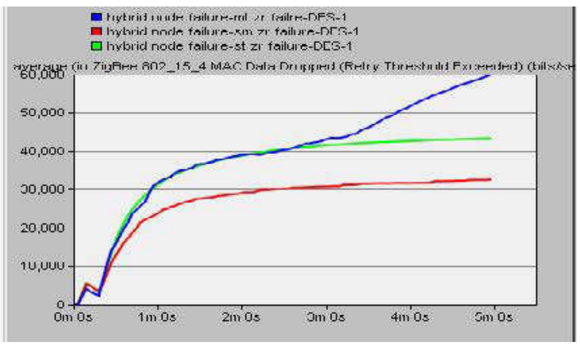


Fig 6. Data Dropped in MT, SM and ST Topologies in Case of ZR Failure

C. Delay

This statistic records the medium access delay experienced by the packets submitted for transmission on all interfaces in the network. This value is computed as the interval from the time the packet was inserted into the transmission queue until the time when the packet was sent to the physical layer for the first time. In case of ZED failure and ZED, ZR failure maximum delay shown by MT and SM network and minimum by ST network as referred to fig. 7 and 8.

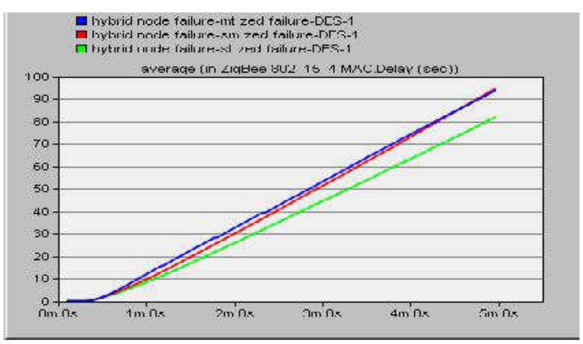


Fig 7. Delay in MT, SM and ST Topologies in Case of ZED Failure

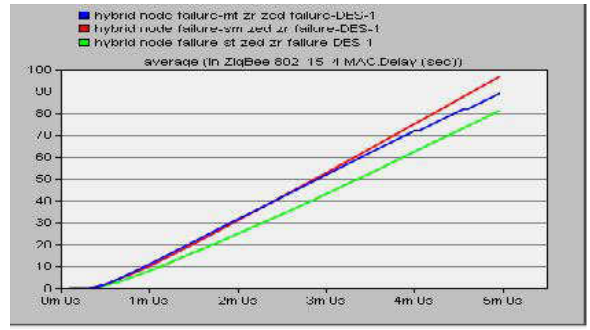


Fig 8. Delay in MT, SM and ST Topologies in Case of ZED and ZR Failure

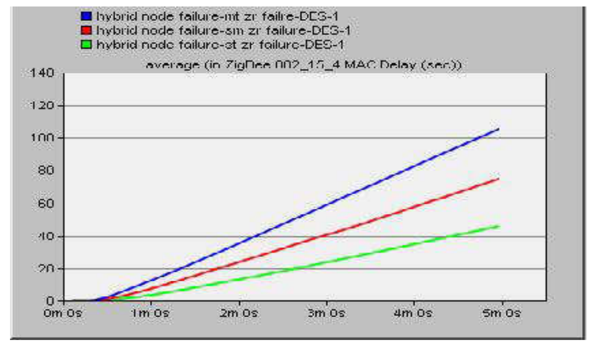


Fig 9. Delay in MT, SM and ST Topologies in Case of ZR Failure

Here the result shown in figure 9 shows that maximum delay achieve by MT network and minimum by ST network in case of ZR failure.

D. Data Traffic Received

These statistics record successfully received data traffic on this network interface from the physical layer. When these statistics are reported in units of bits/second, the physical and the MAC header sizes are included in the computation of the total amount of traffic received. These statistics record all the data received on the network interface regardless of the destination address. The given figure 10, 11 and 12 results conclude that maximum DTR observed in case of SM network and minimum in MT network.

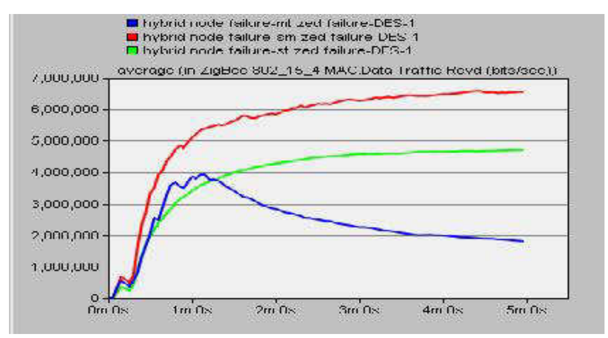


Fig 10. DTR in MT, SM and ST Topologies in Case of ZED Failure

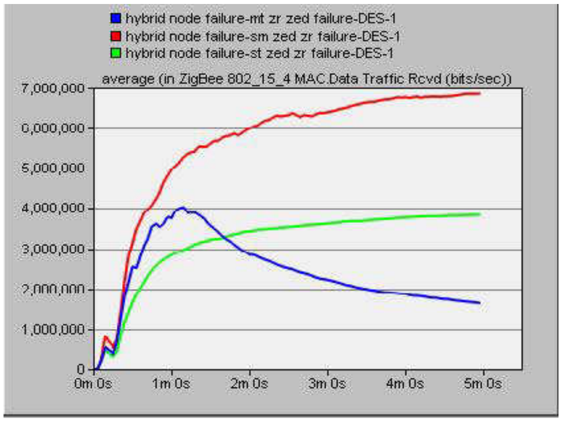


Fig 11. DTR in MT, SM and ST Topologies in Case of ZED and ZR Failure

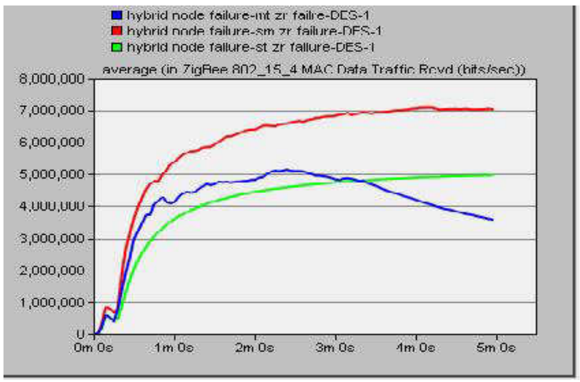


Fig 12. DTR in MT, SM and ST Topologies in Case of ZR Failure

E. Data Traffic Sent

These statistics record the amount of data transmitted by the network interface onto the physical layer. When these statistics are reported in units of bits/second, the physical and the MAC header sizes are included in the computation of the total amount of traffic sent.

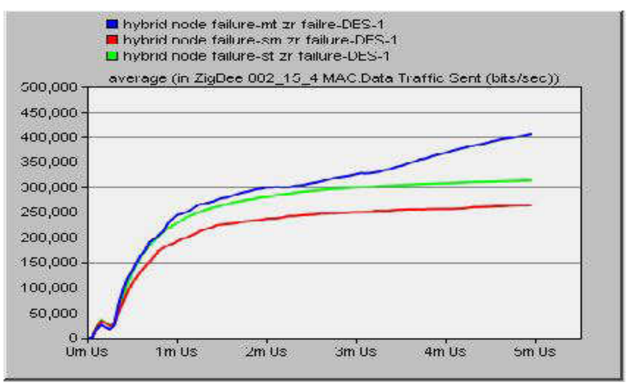


Fig 13. DTS in MT, SM and ST Topologies in Case of ZR Failure

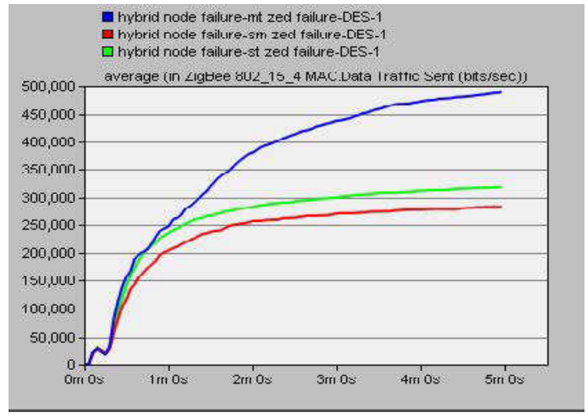


Fig 14. DTS in MT, SM and ST Topologies in Case of ZED Failure

The maximum DTS observed by MT network And minimum by SM network as referred to fig. 13, 14 and 15. According to these results MT send maximum data but receive minimum so it lost of data during communication, whereas SM sent minimum data and receive maximum that means it include duplicate packets during transmission and in case of ST average transmission and reception of data occurred which concludes that it loss minimum data packets and also the duplicate packet transmission is less.

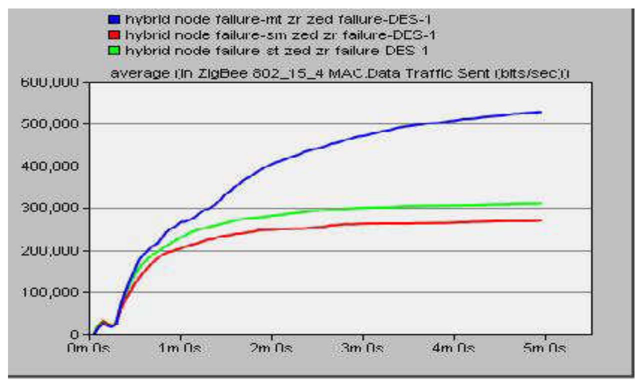


Fig 15. DTS in MT, SM and ST Topologies in Case of ZED and ZR Failure

V. CONCLUSION

The overall results conclude that maximum throughput achieved by ST network and minimum by MT network But throughput of ST and MT network degrades when ZED and ZR fail at same time instant, whereas SM network mostly affected by the failure of ZR. In case of DTS and DTR, the maximum DTS observed by MT network but receive minimum and minimum DTS observed by SM network and receive maximum. In case of ST average transmission and reception of data occurred which concludes that it loss minimum data packets. Due to failure of devices maximum delay achieve by MT network and minimum by ST network in case of ZR failure. According to these results ST hybrid network gives better performance than other topologies in case of node failure.

REFFERENCES

[I] Marco Gribaudo, Daniele Manini, Alessandro Nordio, “Transient Analysis of IEEE 802.15.4 Sensor Networks”,IEEE Transaction On Wireless Communications, VOL.10, NO. 4, APRIL 2011.

[II] Zavosh Abdollahzadeh Davani, Azizah Abdu Manaf, “A Survey on Key Management of ZigBee Network”, The International Conference on E-Technologies and Business on the Web (EBW2013).

[III] Rozeha A. Rashid , Hamdan Sayuti, Nurul Mu'azzah Abdul Latiff, Norsheila Fisal, Mohd Adib Sarijari, Abdul Hadi Fikri Abdul Hamid, Rozaini Abd Rahim, “Simple Scheduling Scheme for Smart Home and Ambient Assisted Living”, The Second International Conference on Informatics Engineering & Information Science (ICIEIS2013) – Malaysia.

[IV] Shi Longlonga, Qiu Chunlinga, Gao Penga, Jia Zhengsena, “The Research and Simulation of CSMA/CA Mechanism of Zigbee Protocol”, 2012 International Workshop on Information and Electronics Engineering (IWIEE), Procedia Engineering 29 (2012) 3466 – 3471 Engineering 00(2011) 000–000.

[V] Surender.R, P. Samundiswary, “Performance Analysis of Node Mobility in Beacon and Non-Beacon enabled IEEE 802.15.4 based Wireless Sensor Network”, IJCA (0975 – 8887), Volume 76– No.12, August 2013.

[VI] J. Pedro Amaro, Fernando J.T.E. Ferreira,RuiCortesão, Jorge Landeck, “Powering Wireless Sensor Networks Nodes for Complex Protocols on Harvested Energy”, CENTERIS 2012 – Conference on ENTERprise Information Systems HCIST 2012 - International Conference on Health and Social Care Information Systems and Technologies.

[VII] Yuan-Yao Shih, Wei-Ho Chung,Pi-Cheng Hsiu, Ai- Chun Pang, “A Mobility-Aware Node Deployment and Tree Construction Framework for ZigBee Wireless Networks”, IEEE transaction on vehicular technology, vol. 62, no. 6, july 2013.

[VIII] Mumtaj M.Ali Al-Mukthar, Teeb Hussain Hadi, “Modeling the Performance of Zigbee Cluster Tree Wireless Sensor Network in Presence of Failure”, Journal of Advanced Computer Science and Technology Research,Vol.3 No.3, September 2013, 116-126.

[IX] P. Medagliani, M. Martalò, G. Ferrari, “Clustered Zigbee networks with data fusion: Characterization and performance analysis” Ad Hoc Networks 9 (2011) 1083–1103.

[X] B.E. Bilgin, V.C. Gungor, “Performance evaluations of ZigBee in different smart grid environments”, ELSEVIER, Computer Networks 56(2012) 2196–2205.

[XI] http://en.wikipedia.org/wiki/ZigBee.

[XII] E. Yüksel, H. R. Nielson, and F. Nielson, "Zigbee-2007 security essentials," in Proc. 13th Nordic Workshop on Secure IT-systems, 2008, pp. 65-82.

[XIII] Zavosh Abdollahzadeh Davani, Azizah Abdul Manaf “A Survey on Key Management of ZigBee Network” The International Conference on E-Technologies and Business on the Web (EBW), (SDIWC).

[XIV] Li-Hsing Yen, Wei-Ting Tsai, “The room shortage problem of tree-based ZigBee/IEEE 802.15.4 wireless networks”, ELSEVIER, Computer Communications 33 (2010) 454–462.

[XV] Meng-Shiuan Pan, Yu-Chee Tseng “The Orphan Problem in ZigBee-based Wireless Sensor Networks” MSWiM, page 95-98. ACM, (2007).

[XVI] Francesca Cuomo, Emanuele Cipollone, Anna Abbagnale “Performance analysis of IEEE 802.15.4 wireless sensor networks: An insight into the topology formation process” Computer Networks 53 (2009) 3057–3075.

Performance Analysis of Static and Dynamic Routing in Computer Network Using Cisco Packet Tracer

Jaswinder Kumar
CT Institute of Technology & Research, Maqsudan,
Jkheer29@gmail.com

Samiksha
CT Institute of Technology &Research, Maqsudan,Jalandhar
samikshadhupar@gmail.com

Anurag Sharma
CT Institute of Technology & Research, Maqsudan, Jalandhar
er.anurags@gmail.com

Rahul Malhotra
CT Institute of Technology & Research, Maqsudan, Jalandhar

***Abstract*--This research paper analysis the comparison between static ,default and dynamic routing technique using Cisco packet tracer.This research paper main focus on various type of problem and benefit related static routing and dyamic routing using RIPv2 protocol.we are analysis various parameter like flexibility,routing update,control etc and find out which routing technique has provide us better results and which routing is better for small and large internetwork and other technical aspects.**

Keywords--CISCO, RIP, RIPv1, RIPv2, EIGRP and OSPF

I. INTRODUCTION

Internetworking stands for the interconnection of two or more computers for sharing of data and other information by devices such as routers and switches. A network is said to be established if two or more computers are sharing data or resources such as scanner, printer etc between them and internet protocol are principles and rules by which the data is to be send through routers. In internetworking, routing play very significant role in moving packets from LAN to WAN and vice versa. It also provides a best path for IP Packet to reach at the destination it occurs at network layer 3 of the OSI model. Router works at network layer of the OSI model and used to route IP packets to destination networks. This is possible as router used routing technique like static, default and dynamic routing. Static and default routing is configure by administrator manually. In dynamic routing, router used various routing protocol such as RIPv1, RIPv2, EIGRP and OSPF etc. Basically router performed routing to create a routing table and learn the neighbor route information.

II. TYPES OF ROUTING

There are mainly two methods for create a routing table:

A. Static Routing- In computer networking, the word static means manually. So in process administrator configure each router interface manually. So routing table is created, update and maintained by administrator manually, in this process router will not share our routing information with each other thus it reduced CPU/RAM overhead so as the result bandwidth is saved.

B. Dynamic Routing- In this routing, router learn all the routing update and other information with help of using routing protocol like RIPv1, RIPv2, EIGRP and OSPF etc. We will configure dynamic routing on each router with the help of RIPv2 protocol.

III. ROUTING INFORMATION PROTOCOL (RIP)

Routing information protocol is a distance vector protocol that works like a rumor. Routers using RIP advertise information about each subnet to their neighbors. Their neighbor in turn passes the information to the surrounding neighbors of their own and so on until all the routers are aware of information. Rip was first developed in 1969. It is available in three versions RIPv1, RIPv2, RIPng. Table1 shows the difference and table2 shows similarities between RIPv1 and RIPv2. By default RIPv1 has been used on routers because RIPv2 has limitation that it supports Classful IP. So it is not used in classless IP.

TABLE I. DIFFERENCE BETWEEN RIPv1 AND RIPv2

S.No.	Parameter	RIPv1	RIPv2
1.	Distance vector protocol	Yes	Yes
2.	Hop count	Up to 15	Up to 15
3.	Split horizon	Present	Present
4.	Poison reverse	Yes	Yes

TABLE II. SIMILARITIES BETWEEN RIPv1 and RIPv2

S. No.	RIPv1	RIPv2
1.	Classful protocol	Classes protocol
2.	It does not send subnet mask with route update	It send subnet mask with update
3.	It use broadcast IP 255.255.255.255 to send update	It used multicast IP 224.0.0.9 to send update
4.	It does not support CIDR and VLSM	It support CIDR and VLSM

RIP uses only hop count to determine the best way to a remote network; it sends the complete routing table out to all active interfaces every 30 seconds. It has send packet and routing information up to 15 hops. After 15 hops it does not provide any routing and other information so packet will be drop after 15 hops. The following are features of RIP:

1) *MULTICASTING*: RIPv2 supports multicasting against the broadcast updates by RIPv1 to share the routing information.

2) *TRIGGERED UPDATES*: when a route fails. It does not wait for the next periodic update. It immediately sends trigger update for the listed failed route. The updates are send to share its routing information with the neighbor, whenever a change occurs.

3) *CLASSLESS PROTOCOLS*: It supports Variable length subnet masking (VLSM) that enables to provide more than 1 subnet mask value. These protocols are known to class A, B and C network only. RIPv1 is a Classful addressing protocol.

4) *AUTHENTICATION*: It allows you to select the routers that you want to participate in RIPv2. The process of authentication is done by administrator.

RIP has three types of messages :

1) *HELLO MESSAGE*: These messages are send to aware neighbor routers about the presence of the respective router. It helps in creating neighbourships.

2) *UPDATE MESSAGE*: These messages are send by a router to update other and helps in exchanging information regarding the routing table.

3) *ACKNOWLEDGEMENT MESSAGE*: this message is used for acknowledgement to exchange information between two routers.RIP uses 4 timers to regulate its performance and keep track of the various routes connected.

There are four RIP timers are present in RIP Protocol:

1) *ROUTER UPDATE TIMER*: The default router updates its routing table information after 30 second. But we change update timer 10 second .so it take less convergence time for sending routing table update .so convergence time will be increase

2) *ROUTE INVALID TIMER*: The default router invalid timer value is 180 second .This periods indicate that after this period router declared a route as invalid. This happen only when router does not send and take any neighbor routing update information from route for that period

3) *HOLD DOWN TIMER*: The default value of this timer is also 180 seconds.

4) *ROUTE FLUSH TIMER:* The default router flush timer value is 240 seconds. After 180 second of invalid timer when router declared a route as invalid, router wait next 60 second. After 60 second it flushes out to the route from the routing table.

III. SETUP

Before setting the network and simulation following devices should be familiar as we have used them in our network.

A. Router- Router is an internetworking device which works at only layer 3 of OSI model. Router is a specialized computer which has some memory. It performs routing operation. It creates its own routing table in its memory and adds all routing information in the routing table. With the help of routing table it provide a best path to IP Packet.

B. Switch- Switch is also an internetworking device which work at layer 2 of OSI model and performed various switching operation.

C. Cables- Cable is wired communication medium between sender and receiver. We used copper straight cable to provide fast Ethernet between switches and PCs and used serial DCE to provide serial interface between two or more router.

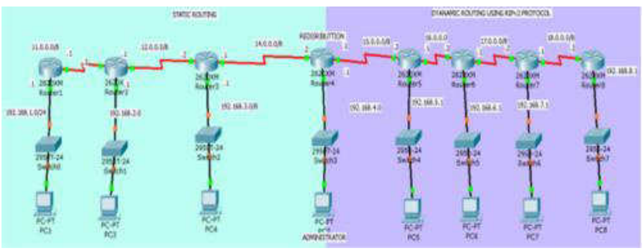


Fig.1. Simulation of static and dynamic routing in Cisco packet tracer

Figure 1 shows simulation of static, default and dynamic routing using RIPv2 protocol in Cisco packet tracer. We have developed a simulation networking model consist of Cisco router, switches and make a physical connection by connecting cable to serial and fast Ethernet by using simulation tool Cisco packet tracer Version 5.3.

D. Configuring CISCO Router- After implementation of physical model we are required to configuring of network with static routing means we are configure all router interface manually. Router will be configure in two ways-

i. GUI(Graphical User Interface)- With GUI we simply configure router by clicking the router then click configuration and then selected which type of configuration you want to configure as shown in fig.2.2

Fig.2. GUI to add network routes to static routing

ii. CLI (command line interface) mode – In this mode, we can configure static routing with the of various command

IV. SIMULATION CODE

Code used in Cisco packet tracer is given below :

```
hostname Router1
enable password 7 08224F4008
spanning-tree mode pvst
interface FastEthernet0/0
ip address 192.168.1.1 255.255.255.0
duplex auto
speed auto
!
interface FastEthernet0/1
no ip address
duplex auto
speed auto
shutdown
!
interface Serial0/0/0
ip address 11.0.0.1 255.0.0.0
clock rate 64000
!
interface Vlan1
no ip address
 shutdown
ip classless
ip route 192.168.2.0 255.255.255.0 11.0.0.2
ip route 0.0.0.0 0.0.0.0 11.0.0.2
line con 0
password 7 08224F4019
login
```

line vty 0 4

password 7 08224F4019

login

line vty 5 15

password 7 08224F4019

login

end

So these commands are implemented in all router interfaces. In this way we can implement all internetwork so that all router interface ping with each other but they cannot transmit the packet.

TABLE III. CHECK DIRECTLY CONNECTED ROUTER INTERFACE

Router1#sh ip int br					
Interface	IP-Address	OK?	Method	Status	Protocol
FastEthernet0/0	192.168.1.1	YES	manual	up	up
Serial0/0	11.0.0.1	YES	manual	up	up

We have implemented routing technique like static, default and dynamic routing.

TABLE IV. CONFIGURE AND DECRIPTION OF STATIC ROUTING IN CLI MODE

IP Route command Parameter	Description
ip route	Identifies the static route command
192.168.2.0	destination_network
255.255.255.0	Indicates subnet mask of network
11.0.0.2	ip address of next hop router in the path of destination

Table IV show that how to configure static route using CLI mode and gives the description of various command used in static routing

IV. SIMULATION

TABLE V. CHECK ROUTER INTERFACE STATUS

Router1#sh ip int br					
Interface	IP-Address	OK?	Method	Status	Protocol
FastEthernet0/0	192.168.1.1	YES	manual	up	up
Serial0/0	11.0.0.1	YES	manual	up	up

TABLE V up status of router interface show that routers are configures properly and they work properly.

```
Router1#conf t
Enter configuration commands, one per line. End with CNTL/Z.
Router1(config)#ip route 192.168.2.0 255.255.255.0 11.0.0.2
Router1(config)#ip route 192.168.3.0 255.255.255.0 11.0.0.2
Router1(config)#ip route 192.168.4.0 255.255.255.0 11.0.0.2
Router1(config)#exit
```

TABLE VI. CLI MODE TO ADD NETWORK ROUTE USING STATIC ROUTING ON ROUTER1

TABLE VII. CHECK STATIC ROUTE FOR ROUTER 1

```
Router1#sh ip route
Gateway of last resort is not set
C    11.0.0.0/8 is directly connected, Serial0/0
C 192.168.1.0/24 is directly connected, Fast Ethernet 0/0
S    192.168.2.0/24 [1/0] via 11.0.0.2
S    193.168.3.0/24 [1/0] via 11.0.0.2
S    193.168.4.0/24 [1/0] via 11.0.0.2
```

TABLE VI shows the setup of static routing and TABLE VII shows static route is present on router 1. Similarly configure all routers in this way.

V. DEFAULT ROUTING

At home we are using default routing ,router at home does not have routing it just pass to its default as shown in Fig.1of router R1 and R4.Default routing is done on end routers that is router which is connected to host.

TABLE VIII CONFIGURE AND DECRIPTION OF DEFAULT IN CLI MODE

IP Route command Parameter	Description
ip route	Identifies the default route command
0.0.0.0 0.0.0.0 11.0.0.2	0.0.0.0 0.0.0.0 ip-address or exit interface type and number.

Table VIII shows that how to configure default route using CLI mode and gives the description of various command used in default routing

TABLE IX. CONFIGURE DEFAULT ROUTING ON ROUTER 1

```
Router1#conf t
Enter configuration commands, one per line.  End with CNTL/Z.
Router1(config)#ip route 0.0.0.0 0.0.0.0 11.0.0.2
Router1(config)#exit
```

Table IX shows that configure of default routing on end router1 on the network . Similarly we configure end router.

TABLE X. CHECK DEFAULT ROUTE ON ROUTER1

```
Gateway of last resort is 11.0.0.2 to network 0.0.0.0
C    11.0.0.0/8 is directly connected, Serial0/0
C 192.168.1.0/24 is directly connected, Fast Ethernet0/0
S    192.168.2.0/24 [1/0] via 11.0.0.2
S    192.168.3.0/24 [1/0] via 11.0.0.2
S    192.168.4.0/24 [1/0] via 11.0.0.2
S    193.168.3.0/24 [1/0] via 11.0.0.2
S    193.168.4.0/24 [1/0] via 11.0.0.2
  S*   0.0.0.0/0 [1/0] via 11.0.0.2
```

"S*" indicate that default route has been implement in router1 and "C" shows directly connected routers. Similarly we have configured default route on all end router of internetwork.

TABLE XI. CONFIGURE RIPV2 COMMAND

IP Route command Parameter	Description
Router rip	Identifies the RIP route command
Version 2	Indicates RIPv2
15.0.0.0	Network address of directly connected interface of router

Table XI show that how to configure dynamic route (RIPv2) using CLI mode and gives the description of various command used in dynamic routing.

TABLE XII. CONFIGURE DYNAMIC ROUTE (RIPV2) ON ROUTER 5

```
Router5#conf t
Enter configuration commands, one per line.End with CNTL/Z.
Router5(config)#router rip
Router5(config-router)#ver 2
Router5(config-router)#network 15.0.0.0
Router5(config-router)#network 16.0.0.0
Router5(config-router)#network 192.168.5.1
```

Table XI show that configure dynamic route on Router 5 on network. Similarly we configure all outer5, router6, router7 and router8.

TABLE XII. CHECK RIP ROUTE ON ROUTER 5

```
Gateway of last resort is not set
C    15.0.0.0/8 is directly connected, Serial0/1
C    16.0.0.0/8 is directly connected, Serial0/0
R    17.0.0.0/8 [120/1] via 16.0.0.2, 00:00:11, Serial0/0
R    18.0.0.0/8 [120/2] via 16.0.0.2, 00:00:11, Serial0/0
C    192.168.5.0/24 is directly connected, FastEthernet0/0
R    192.168.6.0/24 [120/1] via 16.0.0.2, 00:00:11, Serial0/0
R    192.168.7.0/24 [120/2] via 16.0.0.2, 00:00:11, Serial0/0
R    192.168.8.0/24 [120/3] via 16.0.0.2, 00:00:11, Serial0/0
```

Table XII shows that RIPv2 protocol (dynamic routing) enable in Router5. "R" indicates that RIP route has been implementing in Router5

V. RESULTS

The results of simulation are shown in TABLE XIII. It shows that all router interfaces is up state. But if the network is failed then dynamic protocol warn with a notification to the network administrator that this interface of route is down state means it will not work at that time.But in case of static routing is not possible. So dynamic routing is more flexible than static routing.

TABLE XIII. IN CASE OF NETWORK FAILURE SERIAL0/1 INTERAFCE IS DOWN STATE

Router5#sh ip int br

Interface IP-Address OK? Method Status Protocol

FastEthernet0/0 192.168.5.1 YES manual up up

Serial0/0 16.0.0.1 YES manual up up

Serial0/1 15.0.0.2 YES manual up up

%LINK-5-CHANGED: Interface Serial0/1, changed state to down

%LINEPROTO-5-UPDOWN: Line protocol on Interface Serial0/1, changed state to down

Router5#sh ip int br

Interface IP-Address OK? Method Status Protocol

Serial0/1 15.0.0.2 YES manual up down

Table XIII also show that configure static route is more difficult than dynamic route so administrator overhead also increase. It also show that configure of static route on router should take more time than dynamic (RIP route). So static routing is only used for small network and dynamic routing used for large network.

TABLE XIV. SERIAL0/1 INTERFACE IS UP STATE

Router5#sh ip int br
Interface IP-Address OK? Method Status Protocol

FastEthernet0/0 192.168.5.1 YES manual up up

Serial0/0 16.0.0.1 YES manual up up

Serial0/1 15.0.0.2 YES manual down down

Router5#
%LINK-5-CHANGED: Interface Serial0/1, changed state to up

%LINEPROTO-5-UPDOWN: Line protocol on Interface Serial0/1, changed state to up

Router5#sh ip int br
Interface IP-Address OK? Method Status Protocol

FastEthernet0/0 192.168.5.1 YES manual up up

Serial0/0 16.0.0.1 YES manual up up

Serial0/1 15.0.0.2 YES manual up up
Router5#

VI. CONCLUSION

After the simulation and analysis of static, dynamic and default routing we have conclude that dynamic routing can be preferred over static routing where traffic control is complex. As static routing need administrator's presence every second and dynamic routing can manage the routes as well packet dropping occur due to network breakage by using protocols. Following table clearly defines difference between static and dynamic routing:

S. No.	Characteristics	Static	Dynamic
1.	Routing update	Manual	Automatic
2.	Network infrastructure	Small because less overhead	Large
3.	flexibility	No	Yes
4.	Administrator overhead (work)	high	low
5.	Administrator control	High	low

REFERENCES

[1] Ashish Saklani , Dr.S.C.Dimri " Impact of computer network routing based on imprecise routing tables and static v/s dynamic routing" The International Journal Of Engineering And Science (IJES) Volume2, Issue 7 , 2013 ISSN(e): 2319 – 1813 ISSN(p): 2319 – 1805.

[2] N.Nazumudeen, C.Mahendran "Performance Analysis of Dynamic Routing Protocols Using Packet Tracer" Department of CIVIL, CSE, ECE, EEE, MECHNICAL Engg. and S&H of Muthayammal College of Engineering, Rasipuram, Tamilnadu, India ,International Conference on Engineering Technology and Science-(ICETS'14) Volume 3, Special Issue 1, February 2014.

[3] Shravan K. Naivety and Santosh K. Balsu (2011) "Performance Comparison of EIGRP and ISIS/RIP Protocols", School of Computing/Telecom, International Journal of Scientific and Research Publications, Volume 2, Issue 2, February 2012 ISSN 2250-3153.

[4] Tuba, Milan: A Mathematical Model for Routing Comparison in Computer Networks, Bulletins for Applied Mathematics (BAM), LXXXVI-A 1565/98, Arad, July 1998, pp. 493- 503.

[5] Archana Dubey, Sushma Makloha, "Implementation, analysis and comparison of routing protocol (RIP & OSPF) using network simulator education version OPNET" Saoud Sarwar International Journal of Research in Engineering & Applied Sciences 633.

[6] Avinash Sridharan, Bhaskar Krishnamachari, Omprakash Gnawaliy " Routing Without Routes: The Backpressure Collection Protocol Scott Moeller", University of Southern California, Los Angeles, CA yStanford University, Palo Alto, CA.

[7] M i n g -Xin Yang; Sch. of Econ. & Manage " Research on the performence of dynamic routing algorithm", Hebei University of Science & Technology, Shijiazhuang, China.

[8]Vasudha Bahl "Performance Issues and Evaluation considerations of web traffic for RIP & OSPF Dynamic Routing Protocols for Hybrid Networks Using OPNETTM", Maharaja Agrasen Institute of Technology, New Delhi, India Volume 2, Issue 9 and September 2012

[9] Harilaos N. Psaraftis " Dynamic vehicle routing: Status and prospects".

[10] Waxman, B.M "Routing of multipoint connections" Computer & Communication. Research Center, Washington Univ., St. Louis, MO, USA Selected Areas in Communications, IEEE Journal on (Volume: 6, Issue: 9).

[11] Md Samil Farouqui ,Vishal Nigam, Gunjan Gandhi" Evolution of the Routing Protocols from RIP (1969) to OSPF (1998)" Volume 4, Issue 2, February 2014 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering Research Paper.

[12] Kiavash Mirzahossein, Michael Nguyen, Sarah Elmasry "Analysis of RIP, OSPF, and EIGRP Routing Protocols using OPNET". Simon Fraser University School of Engineering Science.

Shortest Path Computation Using ACO With DIJKSTRA's In Wireless Mesh Adhoc Networks

Bhawna Talwar
Dept of Computer Science
CT Institution of Higher Studies
Jalandhar, India
Bhawnatalwar2@gmail.com

***Abstract*—In multi-hop routing, the nodes are forwarding packets to each other which require some sort of routing protocol to take the routing decisions that is to select shortest best possible path, also in case of route failure. Therefore to solve this routing problem in Mobile Adhoc Networks based routing protocol, Ant colony optimization (ACO) technique inspired by behavior of real ants is used. In this paper AntNet routing protocol has been implemented on mesh topology and computed total Pheromone values from routing tables created and updated by ants for each node in the network. Then least total pheromone value is considered according to Dijkstra's algorithm to signify final shortest route. This paper proposed the idea which improves the route repeatability problem created by ants while path generation, by applying Dijkstra's concept in path selection phase which selects shortest path in less time. MANETs when uses ACO increase the throughput and packet delivery fraction by providing alternative paths which are not provided by applying only Dijkstra's. Therefore it is used in conjunction with ACO algorithm, best fit for dynamic topology. Hence alternative shortest paths are optimized and storage for least used paths is removed by this proposed combination.**

Keywords—MANETs, ACO, AntNet, AODV, CBR

I. INTRODUCTION

Mobile Ad Hoc networks (MANETs) are the networks in which mobile nodes operating in a distributed manner without any central infrastructure and is thus called self-organized network. The major issue related to MANETs is routing because of ad hoc network, nodes changes its domain and causes the changing of routes as well.

Ad hoc Routing Protocols were classified as follows:

- Proactive Protocols
- Reactive Protocols
- Hybrid Protocols

The Fig. 1 describes the three classifications of Ad hoc routing protocols and various protocols proposed under each category [2, 5 and 13].

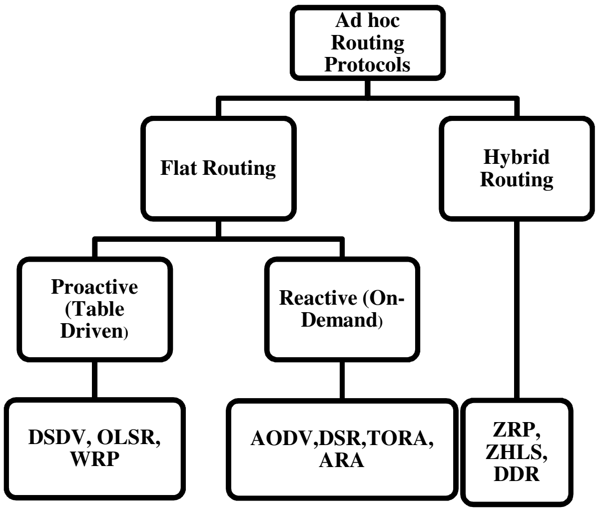


Fig. 1. Classification of Ad hoc Routing Protocols

II. ROUTING IN MANETS

The basic routing phases for MANETs are:

- Path Generation- In this phase path is generated according to distributed state information of the network and of the application, and traffic state information.
- Path Selection- On the basis of network and application state information, appropriate paths are selected.
- Data Forwarding- In this phase traffic is send along the selected route.
- Path Maintenance- The maintenance of the selected route via control messages is done.

The link failures occurred in Adhoc on-demand distance vector (AODV) routing protocol causes the problem of long route discovery and does not provide best alternative route. So to obtain shortest alternative routes, ACO technique is used.

III. ANT COLONY OPTIMIZATION

Swarm Intelligence (SI) [1] is the group of organism able to accomplish some type of task in decentralized and self-organized system. Commonly used swarm intelligence technique is Ant Colony optimization. Ant Colony Optimization (ACO) is based on foraging behavior of ant species in real life. MANET routing issues [12] have been resolved so far by using existing ACO based routing protocols and still enhancement in these protocols is in its progress. [6, 7 and 12].

A. AntNet

Di Caro and Marco Dorigo proposed AntNet [3] which is based on two mobile agents Forward Ant (FA) and Backward Ants (BA) traverse from source to destination d, to collect information regarding c-ongestion, delay and path in network. Source generates FA which uses routing table T_k which stores probability value P_{nd} for destination-neighbor pair, to find path to destination and record the route it has taken. When they reached at the destination FA dies and creates the BA which goes back to the source by moving along the same path followed by the FA but in opposite direction. During this traversal by BA, [6] it modifies routing table by increasing the routing probability of FA and decrease in probability of all other neighbor nodes, but this increase and decrease should be done at the total of all probabilities will remain 1. The update of the routing table is done by using the quantity, *r'*, which derived according to equation 1:

$$r' = \begin{cases} \frac{T}{C\mu}, C \geq 1, if \ \frac{T}{C\mu} < 1 \\ \quad 1, \ otherwise \end{cases} \quad (1)$$

Where *T* is the trip time from the current node to the destination, μ is average of *T,* and *C* is a scaling factor, usually set to 2.

IV. COMPUTATION OF PHEROMONE VALUES

The following steps to compute and update pheromone values by ants in ACO algorithm are discussed below:

Step1: Set parameters; initialize pheromone values

Step2: Do while (destination node not reached):

Computed Pheromone value at every node,

Traverse next node (if not destination)

Step3: Update values in Routing Table (using equation 1 mention in above section)

Step4: Print Routing Table

Step 5: End

V. DIJKSTRA'S ALGORITHM

Dijkstra's Algorithm was created in 1959 by Dutch computer scientist Edsger Dijkstra. Its basic concept is to find out shortest path from node to any other target node and then use path length to find shortest path iteratively [8, 9].

Let in Graph G with vertex V, edge E, and weight W length of the path is sum of weights on vertices. Therefore shortest path between initial vertex V_o and final vertex V_f is the length of path between V_o and V_f with lowest total cost by summation of weights [4, 10 and 11].

VI. SIMULATION ENVIRONMENT

Network Simulator NS2 is used to simulate ant based routing on MANET based AODV routing protocol. The radio propagation model is used for physical layer. To simulate the environment for the implementation of Antnet following considerations are taken in account:

- To simulate ant based routing protocol, number of nodes have been kept equal to number of ant agents.
- To define network mobility of nodes in flat grid area of 800×800 with bandwidth of 10Mbits/sec and delay of 10ms is simulated for all above mentioned protocols.
- Constant Bit Rate (CBR) traffic source is taken with packet size taken is of 1000bytes send at the interval of .015 sec.
- Simulation finishes at 10.0 sec.
- The useful data is extracted from trace file using awk scripts.

Fig. 2 describes the simulation of Antnet of 12 nodes using mesh topology in which source node is 0 and destination node is 11. The network scenario described above shows that little packet dropping occurs when nodes move according to set mobility model.

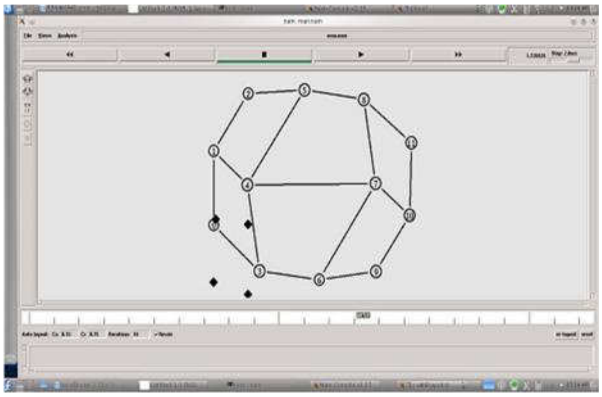


Fig. 2. Simulation of Antnet of 12 nodes

VII. RESULTS

A. Path Calculation

The routing table generated as text output file contains routing table at every node from node0 to node11 but here description of routing table at node11 in TABLE I. is considered while keeping in mind the above scenario in the Fig. 2.

TABLE I. Routing Table at Node11

Source Node	Next Node	Pheromone Value
0	3	0.500000
	1	0.500000
1	4	0.333333
	2	0.333333
	0	0.333333
2	5	0.500000
	1	0.500000
3	6	0.333333
	0	0.333333
	4	0.333333
4	7	0.250000
	1	0.250000
	5	0.250000
	3	0.250000
5	8	0.333333
	2	0.333333
	4	0.333333
6	9	0.333333
	3	0.333333
	7	0.333333
7	10	0.250000
	4	0.250000
	8	0.250000
	6	0.250000
8	11	0.333333
	5	0.333333
	7	0.333333
9	6	0.500000
	10	0.500000
10	7	0.333333
	11	0.333333
	9	0.333333
11	8	0.500000
	10	0.500000

The Source is node0 and Destination is node11, the various paths with total pheromone value have been constructed from above routing TABLE I. Path calculated from node0 to node 11 are described in Table II. From Table I, if node1 is opted as next node to node0 then we have calculated 21 paths according to more pheromone value to next neighbor node, and if same pheromone value found then consider all the nodes like from node0 to node3 is also consider in our case which further generated 17 total paths to destination node11. Also, excluded the loop forming paths like 0-1-2-5-8-7-6-3-4, 0-1-2-5-8-7-10-9-6-3-4 as both paths contain last node4 which further attached to node1, node3, node5, node7 which forms loop as they have already taken before node4.

Table II. Path Calculated with Total Pheromone Value

No. of Nodes	Calculated Path	Total Pheromone Value
If node1 is opted as next node to node0 then we have calculated 21 paths, excluded loop forming paths.		
6	0-1-2-5-8-11	1.99
6	0-1-4-7-10-11	1.66
6	0-1-4-7-8-11	1.66
8	0-1-2-5-8-7-10-11	2.57
8	0-1-2-5-4-7-10-11	2.49
8	0-1-2-5-4-7-8-11	2.49
8	0-1-4-3-6-9-10-11	2.57
8	0-1-4-3-6-7-10-11	2.32
8	0-1-4-3-6-7-8-11	2.32
8	0-1-4-5-8-7-10-11	2.32
8	0-1-4-7-6-9-10-11	2.49
10	0-1-2-5-8-7-6-9-10-11	3.40
10	0-1-2-5-4-7-6-9-10-11	3.32
10	0-1-2-5-4-3-6-9-10-11	3.40

10	0-1-2-5-4-3-6-7-10-11	3.15
10	0-1-2-5-4-3-6-7-8-11	3.15
10	0-1-4-3-6-9-10-7-8-11	3.15
10	0-1-4-5-8-7-6-9-10-11	3.15
12	0-1-2-5-8-7-4-3-6-9-10-11	3.98
12	0-1-2-5-4-3-6-9-10-7-8-11	3.98
If <u>node3</u> is opted as next node to <u>node0</u> then we have calculated 17 paths, excluded loop forming paths.		
6	0-3-6-9-10-11	1.99
6	0-3-6-7-10-11	1.74
6	0-3-6-7-8-11	1.74
6	0-3-4-7-10-11	1.66
6	0-3-4-7-8-11	1.66
6	0-3-4-5-8-11	1.74
8	0-3-6-9-10-7-8-11	2.57
8	0-3-6-7-4-5-8-11	2.32
8	0-3-4-7-6-9-10-11	2.49
8	0-3-4-1-2-5-8-11	2.57
8	0-3-4-5-8-7-10-11	2.32
10	0-3-6-9-10-7-4-5-8-11	3.15
10	0-3-6-7-4-1-2-5-8-11	3.15
10	0-3-4-1-2-5-8-7-10-11	3.15
10	0-3-4-5-8-7-6-9-10-11	3.15
12	0-3-6-9-10-7-4-1-2-5-8-11	3.98
12	0-3-4-1-2-5-8-7-6-9-10-11	3.98

B. Path Selection

According to Dijkstra's concept the best path is selected for minimum 6, 8, 10, and maximum12 nodes with lowest total pheromone value (low cost) from above TABLE II. are described in TABLE III. The first best optimal shortest path selected by ants will be 0-1-4-7-10-11 with the smallest total pheromone value 1.66. The four available paths for node0 to node11 containing six nodes with total pheromone value 1.66 are there, so if link failure occurs at node1 in first path then we can select the alternative shortest path. Hence resolve the problem of finding alternative path in minimal time without reinitiating route discovery process in AODV. Similarly for the number of nodes eight we have 6 paths available and for number of nodes ten we have 8 paths available and for number of nodes twelve we have 4 paths.

TABLE III. Path Selected with Total Pheromone Value

No. of Nodes	Selected Path	Total Pheromone Value
6	0-1-4-7-10-11	1.66
6	0-1-4-7-8-11	1.66
6	0-3-4-7-10-11	1.66
6	0-3-4-7-8-11	1.66
8	0-1-4-3-6-7-10-11	2.32
8	0-1-4-3-6-7-8-11	2.32
8	0-1-4-5-8-7-10-11	2.32
8	0-1-4-3-6-7-10-11	2.32
8	0-3-6-7-4-5-8-11	2.32
8	0-3-4-5-8-7-10-11	2.32
10	0-1-2-5-4-3-6-7-10-11	3.15
10	0-1-2-5-4-3-6-7-8-11	3.15
10	0-1-4-3-6-9-10-7-8-11	3.15
10	0-1-4-5-8-7-6-9-10-11	3.15
10	0-3-6-9-10-7-4-5-8-11	3.15
10	0-3-6-7-4-1-2-5-8-11	3.15
10	0-3-4-1-2-5-8-7-10-11	3.15
10	0-3-4-5-8-7-6-9-10-11	3.15
12	0-1-2-5-8-7-4-3-6-9-10-11	3.98
12	0-1-2-5-4-3-6-9-10-7-8-11	3.98
12	0-3-6-9-10-7-4-1-2-5-8-11	3.98
12	0-3-4-1-2-5-8-7-6-9-10-11	3.98

C. Analysis of Shortest Path Optimization

Shortest and alternative path are created by ACO & Pheromone Values is updated. Dijkstra's reduces the number of path in analyzing the shortest path. Hence from the Fig. 3 it has been analyzed that number of paths created by ACO in mesh topology are more which are reduced by Dijkstra's concept of removing least used paths. Thus by this analysis throughput and packet delivery fraction are increased shown in Fig. 4 and Fig. 5, also releases the storage space of least used paths and reduces the chances of stack overflow by ants.

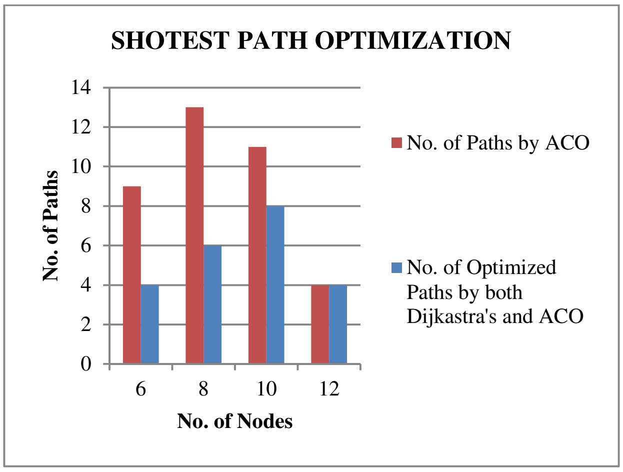


Fig. 3. Analysis Graph of Shortest Path Optimization

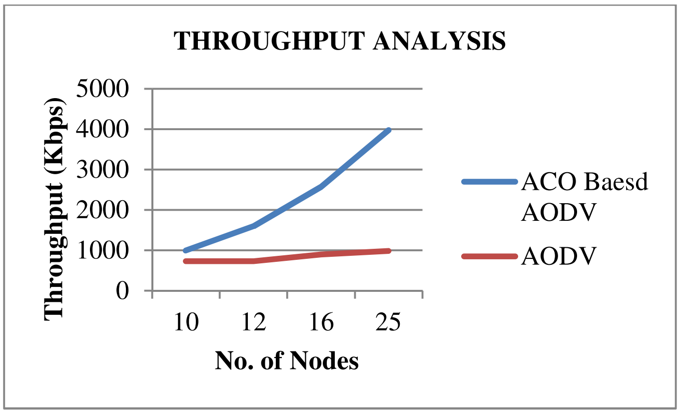


Fig. 4. Throughput Analysis

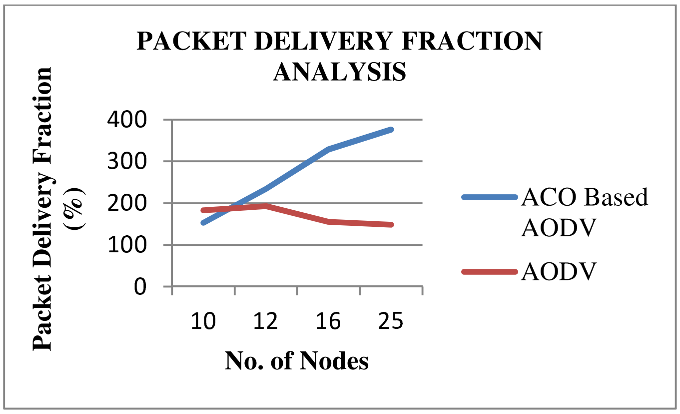


Fig. 5. Packet Delivery Fraction Analysis

CONCLUSION

Considering mesh topology which can withstand with high traffic and has feature to provide multiple paths on failure of node in network, thus causes reduction in route-discovery process and providing high connectivity for real time data and multimedia communication. Antnet implementation is done using mesh topology which provides many paths for every node to its destination and to select best shortest alternative path out of provided paths, ant agents of ant based routing technique are applied. By using Dijkstra's in conjunction with ACO the alternative shortest paths are optimized and storage for least used paths is released. This research work has handled link failures thus increased the throughput and reduced the dropping of packets as it shortest alternative path are provided by ants using pheromone value. Also optimized the shortest alternative path provided on applying ACO on MANET routing protocol.

REFERENCES

[1] Abraham, A., Guo, H., and Liu, H., "Swarm Intelligence: Foundations, Perspectives and Applications", Springer-Verlag Berlin Heidelberg, vol.26, pp.3-25, 2006.

[2] Boukerche, A., Turgut, B., Aydin, N., Ahmad, Z.M., Bölöni, L., and Turgut, D. "Routing protocols in ad hoc networks: A survey", Elsevier Computer Networks (55), pp 3032–3080,2011.

[3] Caro, D.G. and Dorgio, M., "AntNet: Distributed Stigmergetic Control for Communications Networks", Journal of Artificial Intelligence Research9, pp-317-365,1998.

[4] E. W. Dijkstra, "A note on two problems in connexion with graph[J]. Numerische Mathematic", Vol.1, pp. 269-271, 1959.

[5] Gupta, K.A., Sadawarti, H., and Verma, K.A., "Review of various Routing Protocols for MANETs", International Journal of Information and Electrical Engineering, ISSN: 1109-2742, Vol. 1 No. 3, pp. – 251-259, 2011.

[6] Gupta, K.A., Verma, K.A., and Sadawarti, H., "Analysis of various Swarm-based and Ant-based Algorithms", Proc. Of International Conference on Advances in Computing and Artificial Intelligence (ACAI 2011), an ACM Chapter Event, Chitkara University, Punjab, pp. 39-43, 2011.

[7] Gupta, K.A., Sadawarti, H., and Verma, K.A., "MANET Routing Protocols Based on Ant Colony Optimization", International Journel of Modelling and Optimization, vol.2, No.1, 2012.

[8] James B. Orlin, Kamesh Madduri, K. Subramani, and M.Williamson, "A faster algorithm for the single source shortest path problem with few distinct positive lengths" Journal of Discrete Algorithms 8, Publisher Elsevier, pp- 189-198, June 2010.

[9] Liu Xiao-Yan, Chen Yan-Li, "Application of Dijkstra Algorithm in Logistics Distribution Lines", Proceedings of the Third International Symposium on Computer Science and Computational Technology (ISCSCT '10) Jiaozuo, P. R. China, 14-15, ISBN 978-952-5726-10-7 , pp. 048-050, August 2010.

[10] L.Bing, C. Siew-Hwee, L. Shee-Ling, L.Sing-Meng, L.Soo-Chee, P.Foong- Ping, and T.Hwee-Har, "Integrating Case-based Reasoning, Knowledge-Based Approach and Dijkstra's Algorithm for Route Finding", Proceeding of the conference on Artificial Intelligence Applications, pp. 149-155, 1994.

[11] M.Barbehenn, "A Note on the Complexity of Dijkstra's Algorithm for Graphs with Weighted Vertices", IEEE Transactions on Computers, Vol. 47, No.2, 1988.

[12] Mahmood, Y.A., Hamid, H., and Lobiyal, K.D., "Improved Power Control MAC Protocol for wireless Ad Hoc Networks", WSEAS TRANSACTIONS on COMMUNICATIONS, ISSN: 1109-2742, Issue 1, Vol.10, 2011.

[13] Royer, M.E. and Toh, K.C., "A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks", Personal Communications, IEEE, Issue 2, Vol.6, pp 46-55,1999.

Analysis of Prominent Routing Protocols Based on Black Hole Attack for VoIP

Lovepreet Singh
Shaheed Bhagat Singh State
Technical Campus, Ferozepur
lovepreet.2524@gmail.com

Navdeep Kaur
Dept. of ECE
Shaheed Bhagat Singh State
Technical Campus, Ferozepur
navdeepkaurjhaj213@gmail.com

Gurjeevan Singh
Dept. of ECE
Shaheed Bhagat Singh State
Technical Campus, Ferozepur
gurjeevansandhu@gmail.com

Abstract--This research paper presents the fair performance analysis of MANETs routing protocols for voip application under black hole attack. In this paper we have analyze the performance of three routing protocols (AODV, DSR and TORA) by varying the number of black hole nodes. The research work has been based on simulation study in OPNET modeler. The performance comparison of routing protocols has been done in terms of Jitter, Throughput and Retransmission attempts. Finally the experimental results indicate that AODV is better than DSR and TORA.

Keywords—MANETs,AODV,DSR, TORA, OPNET, VOIP.

I. INTRODUCTION

As wired media has become much expensive. So there is emerging need of wireless network which is cheap mean of wireless communication like Bluetooth, Wi-Fi. MANETs, Wireless mesh networks (WMNs), WWAN are the types of wireless network which provide communication over wide areas [1,2]. In MANET network nodes are communicate to each other without any central control authority or any fixed infrastructure, so it is an open network [3]. The nodes of MANETs have wireless radio range, MANETs allows its wireless nodes to form such a topology in which source node can communicate with other node that does not lie in its direct range [4]. The nodes are transmitting or receiving data using smart antennas inbuilt in mobile node [2]. The intermediate nodes lie between source and destination node must route or forward the packets to destination node [5]. Due to security provided by MANET it can be used in military battlefields, classrooms and rescue sites etc. [5,6]. In these applications it is necessary that all nodes communicate and cooperate with in a given group of nodes. The multicast routing protocols in MANETs are divided into two following categories: Tree based and mesh based. In a multicast routing tree, only single path exists between source and destination node, while there are multiple paths between source and destination in mesh multicast protocol [7]. In [1], the simulation of two protocols i.e. AODV and DSR addressed in OPNET modules. Simulation results show that DSR is not suitable for wireless transmission. In [13], two routing protocols (AODV and DSR) were compared using VOIP and videoconferencing applications in OPNET modeler. The results show that AODV is better than DSR for both applications. In this Paper we have evaluated the performance of three routing protocols (AODV, DSR and TORA) by using VOIP application under black hole attack.

II. ROUTING PROTOCOLS

As the links between the nodes are wireless so the nodes are independent to move in the network or leave the network. Due to this mobility of nodes routing in MANETs varies. Therefore dynamic change in the network makes routing in MANET critical and difficult to implement [8]. In MANETs routing Protocols are used to generate route between numerous nodes. Each routing protocol uses its own routing technique to determine a routing path between source and destination node. The characteristics like performance can vary depending on the density of nodes in a specific area their speed and direction [9]. In this Paper we have evaluated the performance of AODV, DSR and TORA routing protocols.

A. Ad Hoc On-Demand Distance-Vector Routing Protocol (AODV)

AODV supports multicast routing [5]. AODV is a source initiated on-demand routing protocol. In AODV, if a source node wants to send data to destination node then firstly it will check the route, if route exists then it simply forwards the packets to destination node, if route does not exists, a route discovery process begins [6]. AODV takes over the three route discovery process: Route Request (RREQ): When source node wants to initiate communication with destination node then it spread the RREQ message through intermediate nodes in network. Route Response (RREP): After receiving RREQ message, destination node reply as a RREP to source node and clarify the route in between source node and destination node, so communication will start in between them. Route Error (RERR): When the link breakage happens the node must invalidate the existing route in the routing table entry. The node must list the affected destinations and determine which neighbor can be affected with this breakage. The node sends the RERR message to corresponding neighbors.

B. DSR Routing Protocol (Dynamic Source Routing)

The Dynamic Source Routing protocol (DSR) [4] is based on source routing, which means that the originator of each packet determines an ordered list of nodes through which the packet must pass while travelling to the destination [2]. The key advantage of a source routing design is that intermediate nodes do not need to maintain up-to-date routing information in order to route the packets that they forward, since the packet's source has alr eady made all of the routing decisions [5]. The header of each packet must include the complete information that packet will follow from source to destination [10]. The source node is responsible for listing the route which includes the address of intermediate nodes in packet header. With the help of this table, the next "node" can be identified by its address to which it to transmit the packet on its way to destination node. The working of DSR includes two important stages: Route Discovery and Route Maintenance. In Route Discovery process the source node transmit RREQ packets which will be received by the nodes which are within its radio range. RREQ helps in identifies the destination node for which the route discovery is requested. If the route is identified, then the source node receives a Route Reply (RREP) packet which includes the sequence of intermediate node address on its way to reach destination node. The route maintenance term is used to detecting the error or failure in the route. Then source node repairs the route or it may start the route discovery process for an alternate route to the destination node [2].

C. TORA Routing Protocol (Temporally Ordered Routing Algorithm)

Temporally Ordered Routing Algorithm (TORA) is a source initiated on-demand routing protocol which is based on the link reversal algorithm. Whenever the topological changes occur in the network, each node which has routing information about the neighbor node sends the control message to the other node which is within its radio range. Basic functionality of TORA can be divided into three main processes: Creating Routes, Maintaining Routes, and Erasing Route. TORA has multiple routes from source to destination node. Routes can be created by assigning the directions to the links for whole network or sometimes to the part of the network as desired. Maintaining Routes refers to the topological changes which are always expected in the mobility environment where MANETs are worked on. By this, it means that when there is a topological change, all the routes to the destination have to be re-established and redefined within a finite time. Erasing route is one process to erase the other routes from the network and is done by clear (CLR) messages [11].

III. SECURITY ATTACK ON MANET

MANET often suffer from many security attacks because of its features like open medium, changing its topology dynamically, lack of central monitoring management, cooperative algorithms and no clear defense mechanism [12]. A black hole attack can occur at single-node or at several nodes in collusion. In the Black hole attack a malicious node advertises the route with less hope count for a given destination. Once route is established via malicious node then it will drop all the data packets [6,13,14].

IV. SIMULATION ENVIRONMENT

In carrying out the performance of AODV, DSR and TORA in the presence of black hole attack for VOIP application we have performed the different following experiments. In first scenario 200 nodes are taken in which 50 nodes are Black hole nodes. In second scenario 200 nodes are taken in which 100 nodes are Black hole nodes. In third scenario 200 nodes are taken in which 150 nodes are Black hole nodes. The protocol used for investigation is AODV, DSR and TORA. In each scenario the placement of nodes is random over an area of 1000*1000. The simulation was run for 60s with a seed value of 128. OPNET Simulator 14.5 [15] was used to analyze the performance of AODV, DSR and TORA protocol. We used OPNET modeler, as OPNET modeler provides a comprehensive development environment supporting the modeling of communication network and distributed system and it also support the VOIP application. OPNET modeler provides better environment for simulation, data collection and data analysis. The Simulation parameters used in our scenario are shown in Table 1.

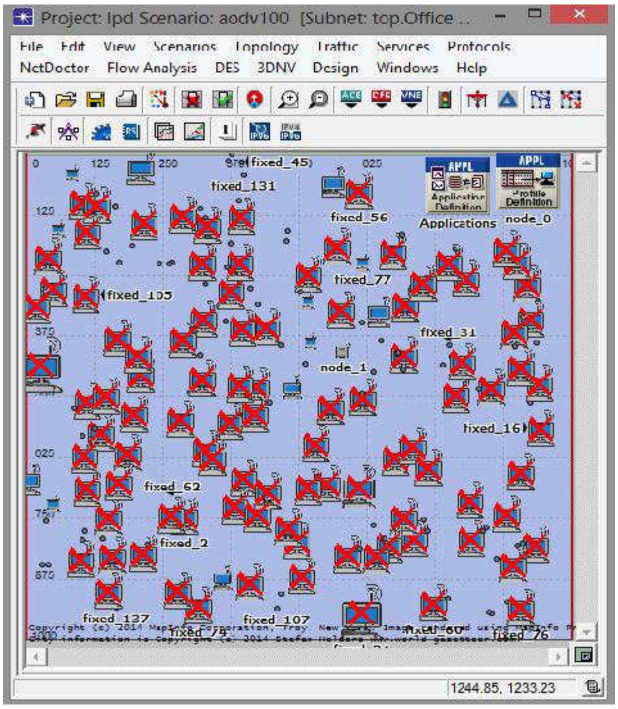


Fig. 1. Snapshot of research work.

TABLE I. Simulation Parameters

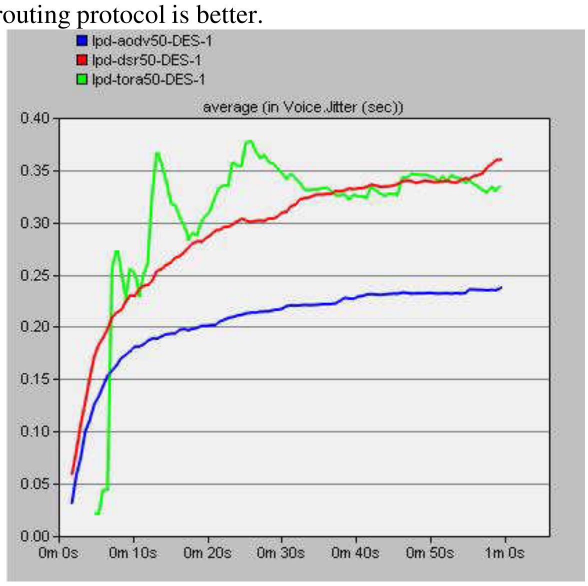
Network Parameters	Values
Total Number of Nodes	200
Number of Black Hole Nodes	50,100,150
Simulation Time	60 sec (1 min)
Simulation Area	1000 m X 1000 m
Routing Protocols	AODV, DSR,TORA
Data Rate	18mbps
Application Name	VOIP

V. RESULTS AND DISCUSSIONS

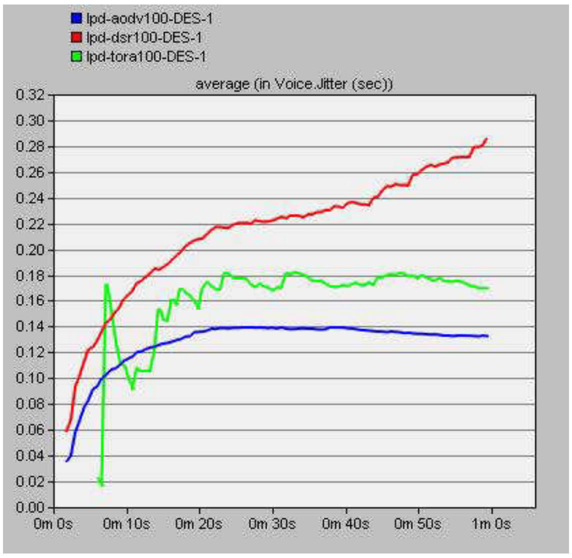
Simulation is conducted to evaluate the performance of MANET using OPNET modeler 14.5. All the result is clarified in the presence of AODV, DSR and TORA routing protocols. The result is evaluated by varying the number of black hole nodes, for application i.e. VOIP. We analysed the performance in the terms of wireless LAN Jitter, Throughput and Retransmission Attempt using different range of scenario.

A. Jitter

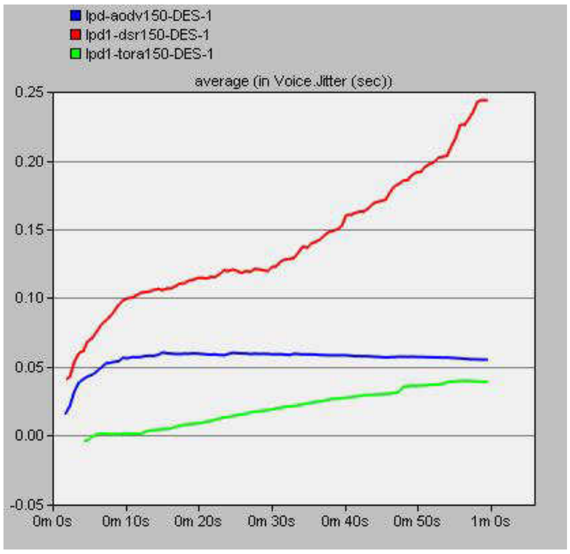
If two consequetive packets leave the source node with time stamps t1 & t2 and are played back at the destination node at time t3 & t4, then: jitter = (t4 - t3) - (t2 - t1). This shows the latency in packets caused by any interference or network congestion. MANETs is a variable network because topology may changes rapidly. Fig. 2 shows the comparison of routing protocols on the basis of jitter using different number of black hole nodes. With the increase in number of black holes nodes of the network i.e. from 50 to 100 nodes, jitter also becomes less. The performance of AODV routing protocol is better.



(a)



(b)

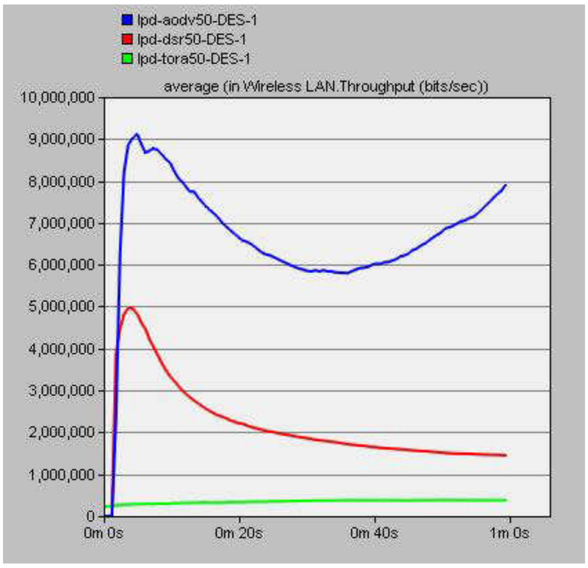


(c)

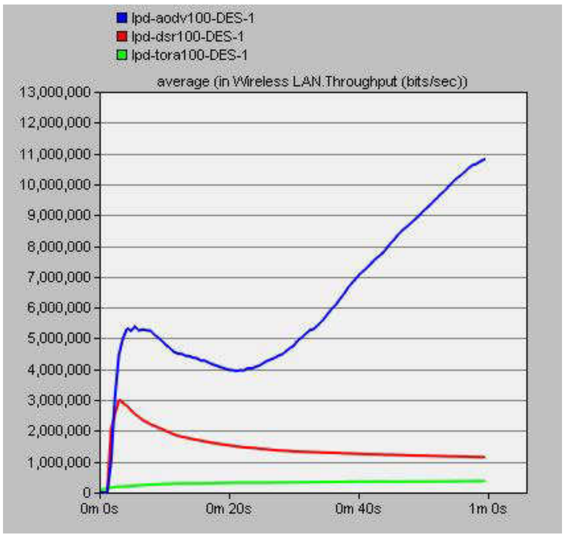
Fig. 2. Jitter for different black hole nodes of (a) 50 nodes, (b) 100 nodes, (c) 150 black hole nodes

B. Throughput

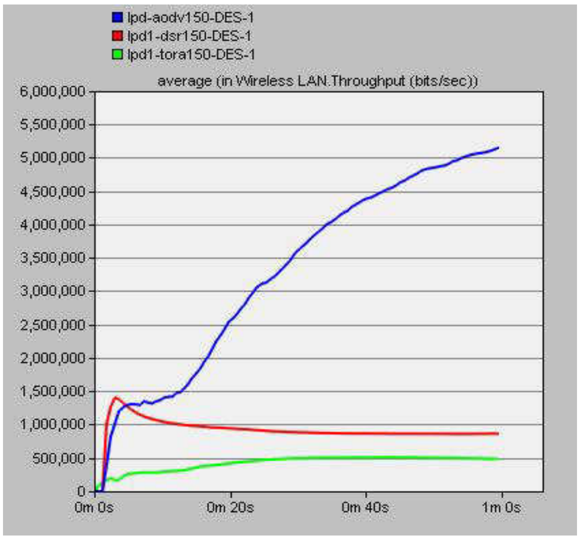
Represents the total number of bits (in bits/sec) forwarded from wireless LAN layers to higher layers in all WLAN nodes of the network. Fig. 3 shows the average throughput of routing protocols using different number of black hole nodes. With the increase in number of black hole nodes of the network the average throughput decreases. Fig. 3 (a) shows that throughput of AODV increases sharply in the beginning of simulation time. From the Fig. 3 it is clear that throughput of AODV is better than throughput of DSR and TORA.



(a)



(b)

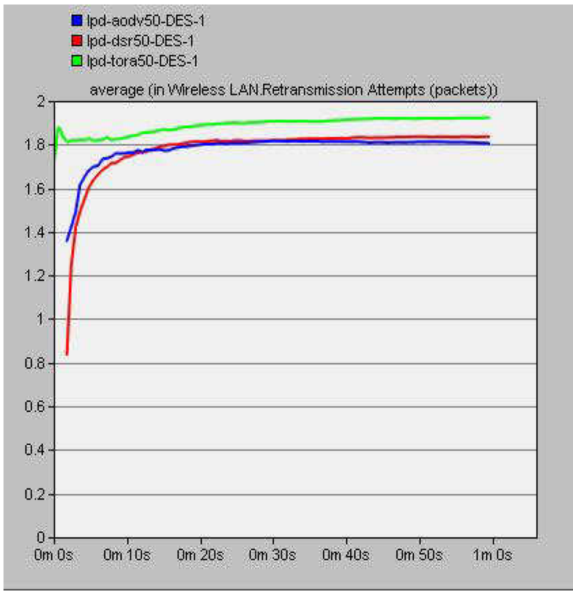


(c)

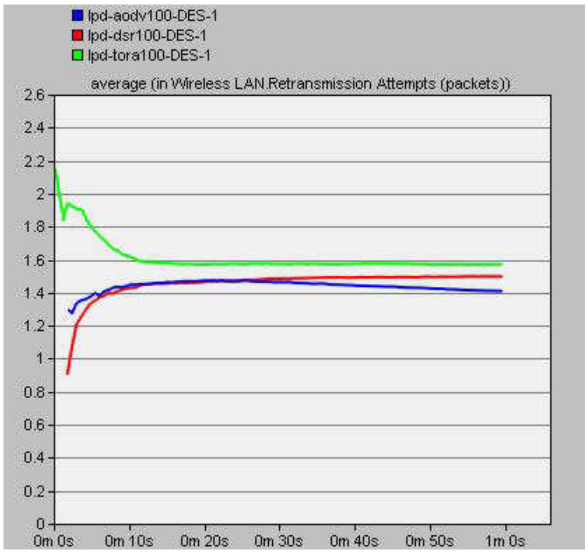
Fig. 3 Throughput for different black hole nodes of (a) 50 nodes, (b) 100 nodes, (c) 150 black hole nodes

C. Retransmission Attempts

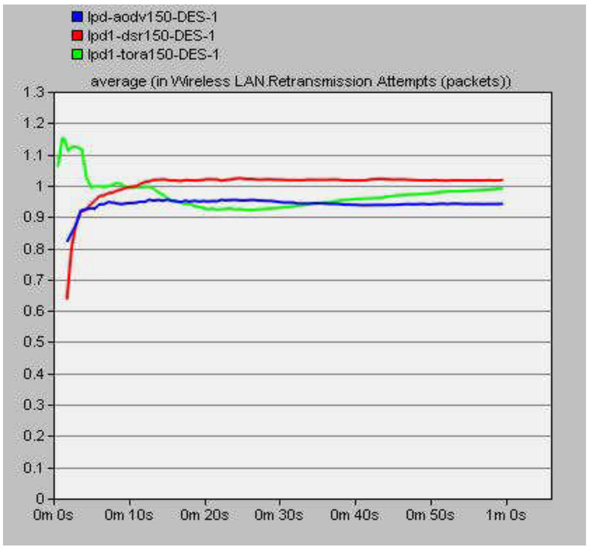
Total number of retransmission attempts by all WLAN MACs in the network until either packet is successfully transmitted or it is discarded as a result of reaching short or long retry limit. For 802.11e-capable MACs, the retransmission attempt counts recorded under this statistic also include retry count increments due to internal collisions. The result shows that in starting the retransmission attempt is less but with passage of time the transmission attempt increases. The result also shows that as we increase the no. of black hole nodes the retransmission attempt decreases. The result also shows that retransmission attempt of AODV is less as compare to DSR and TORA.



(a)



(b)



(c)

Fig. 4 Retransmission Attempts for different black hole nodes of (a) 50 nodes, (b) 100 nodes, (c) 150 black hole nodes

VI. CONCLUSION

In this research analysis the performance of AODV, DSR and TORA by varying no. of Black hole nodes is carried out for VOIP applications in terms of Jitter, Throughput and Retransmission Attempt. In this experiment the placement of nodes is random. Here each node is move with the speed of 5 m/s in an area of within the network range 10, 00,000 sq. m. The study of simulation shows that as we increase the no. of nodes the AODV performs better than DSR and TORA in terms of Jitter, Throughput and Retransmission Attempt. The result shows that the throughput of AODV is better than DSR and TORA. The result also shows that as we increase the no. of nodes the jitter and retransmission attempt of AODV is less than DSR and TORA. The result also shows that AODV protocol can detect black hole nodes faster than DSR and TORA. In the future, the work can be extended to analyze the performance of parameters by using different applications in MANETs.

REFERENCES

[1] Guo, Lei and Peng, Yuhuai and Wang, Xingwei and Jiang, Dingde and Yu, Yinpeng, "Performance evaluation for on-demand routing protocols based on OPNET modules in wireless mesh networks," Computers \& Electrical Engineering, vol. 37, pp. 106-114, 2011.

[2] Vats, Kuldeep and Sachdeva, Monika and Saluja, Krishan, "Simulation and performance Analysis of OLSR, GRP, DSR Routing Protocol using OPNET," IJETED volume 2, Issue 2, March 2012.

[3] Mohebi, Amin and Kamal, Ehsan and Scott, Simon, "Simulation and Analysis of AODV and DSR Routing Protocol under Black Hole Attack," International Journal of Modern Education \& Computer Science, vol. 5, 2013.

[4] Gupta, S Balaji and Navneeth, T and Sundar, S and Vidhyapathi, CM, "Performance Evaluation of MANET Routing Protocols under Varying Node Mobility," International Journal of Engineering \& Technology (0975-4024), vol. 5, 2013.

[5] Aujla, Gagangeet Singh and Kang, Sandeep Singh, "Comprehensive Evaluation of AODV, DSR, GRP, OLSR and TORA Routing Protocols with Varying number of nodes and traffic applications over MANETs," Department of CSE, Chandigarh Engineering College, India, April 2013.

[6] Ramaswamy, Sanjay and Fu, Huirong and Sreekantaradhya, Manohar and Dixon, John and Nygard, Kendall E, "Prevention of Cooperative Black Hole Attack in Wireless Ad Hoc Networks," International Conference on Wireless Networks, 2003.

[7] Nguyen, Hoang Lan and Nguyen, Uyen Trang, "A study of different types of attacks on multicast in mobile ad hoc networks," Ad Hoc Networks, vol. 6, pp. 32-46, 2008.

[8] Saxena, Nidhi and Kumar, Sanjeev and Saxena, Vipul, "Performance Analysis of AODV Routing Protocol under the Different Attacks Through The Use Of OPNET Simulator," International Journal of Innovative Research and Development, vol. 2, pp. 244-248, 2013.

[9] Tamilarasan, Santhamurthy, "A Quantitative Study and Comparison of AODV, OLSR and TORA Routing Protocols in MANET," International Journal of Computer Science Issues(IJCSI), vol. 9, pp. 364-369, January 2012.

[10] Mali, Baldev Ram and Barwar, NC, "Performance of MANET Routing Protocols considering Impact of Node Density under Different Traffic Patterns" International Journal of Computer Applications (0975--8887) on Wireless Communication and Mobile Networks, vol. 1, pp. 26-30, January 2012.

[11] Tamilarasan, S and Sivaram, Dr R, "An Analysis and Comparison of Multi-Hop Ad-Hoc wireless Routing Protocols for Mobile Node," International Journal of Science and Applied Information Technology, vol. 1, pp. 1-5, march-April 2012.

[12] Saini, Akanksha and Kumar, Harish, "Effect Of Black Hole Attack On AODV Routing Protocol In MANET," IJCST, vol. 1, pp. 57-60, December 2010.

[13] Khurana, Sumit and Grover, Amit, "Simulation and Analysis the Effect of Varying no. of Nodes on AODV and DSR for different Applications," IJCA, vol. 77, 2013.

[14] Tseng, Fan-Hsun and Chou, Li-Der and Chao, Han- Chieh, "A survey of black hole attacks in wireless mobile ad hoc networks," Human-centric Computing and Information Sciences, vol. 1, pp. 1-16, 2011.

[15] OPNET official website, http://www.opnet.com.

Performance Evaluation of Star ZRP POVRP Based on VANET's

Harjit Singh
Dept. of CSE
CTIEMT, Shahpur, Jalandhar, Punjab,
INDIA, harjit1985@gmail.com

Manju Bala
Dept. of CSE
CTIEMT, Shahpur, Jalandhar
manju.ctgroup@gmail.com

Manoj Kumar
CTIEMT, Shahpur, Jalandhar, Punjab,
INDIA,
drmanojkumarindia@gmail.com

***Abstract*--As more individuals manage the cost of cars, the interest for in-vehicle diversion is expanding. With the late improvements in figuring and remote correspondence innovations, arranges that can structure the premise for such requisitions could be imagined. The street security standard is moving from a latent one (air packs, ESP, and so on.) to a dynamic one, where sensors, radars, Polaroid's, route frameworks, and chip, usually introduced in vehicles, are coordinated with remote correspondence frameworks to help provisions, for example, stopping support, path keeping, versatile voyage control, and numerous others. Vehicular Ad hoc Networks (Vanets) can productively caution and update drivers through immediate remote vehicle to vehicle interchanges, in the long run diminishing response time and data accessibility restriction. Numerous administrations and requisitions in vehicular specially appointed systems (Vanets) oblige safeguarding and secure information interchanges. To enhance driving wellbeing and solace, the movement related status data will be telecasted customarily and imparted around drivers. Without the security and protection ensure, ambushers could track their intrigued vehicles by gathering and examining their activity messages. Consequently, unnamed message validation is a key necessity of Vanets. In this paper the re-enactment effects demonstrates fundamentally proposed to enhance the administrations in vehicular specially appointed system by dissecting parameters like throughput, End-to-End Delay, jitter etc in gatherings like STAR, ZRP, POVRP & IPOVRP by using the Qualnet 5.0.1 test framework.**

Keywords--MANET, VANET, GPS and LORA.

I. INTRODUCTION

The expanding interest of remote correspondence and the needs of new remote gadgets have a tendency to research on self sorting out, self recuperating systems without the impedance of concentrated or pre established framework/power. The systems with the nonappearance of any unified or pre established base are called Ad hoc systems. Specially appointed Networks are gathering of directing toward oneself versatile nodes. VANET helps vehicle drivers to convey and to arrange around themselves to maintain a strategic distance from any discriminating circumstance through Vehicle to Vehicle correspondence e.g. street side mischance's, congested roads, velocity control, free entry of crisis vehicles and unseen deterrents and so on. Also security provisions VANET additionally give solace requisitions to the street clients. Without security, a Vehicular Ad Hoc Network (VANET) system is completely open to different strikes, for instance, causing of false advised messages and additionally camouflage of bona fide forewarn messages, therefore bringing on mischance's. This makes security a part of true concern in building such organizes. VANET [1] are of prime hugeness, as they are inclined to be around the first business procurement of exceptionally delegated framework advancement. Vehicles are the bigger a piece of eversingle one of center points, which are prepared for molding self dealing with frameworks with no previous taking in of each other, whose security level is low and they are the most vulnerable some bit of the framework which could be snared adequately. The cut-off of VANET innovation is high with a broad mixed bag of orders being passed on in backing of clients, business establishments, for instance, toll courts, diversion associations and moreover law necessity powers. Then again, without securing these frameworks, mischief to life and property could be done at a more amazing degree, so security schemas adeptly send and acknowledge road information, for instance, events, progressing development information, or surface condition. In this appreciation, the reasonable controlling framework should be illustrated with including such components specified long ago. Such a directing framework should surety a steady and robust segment over Vanets. Existing controlling meetings, which are usually expected for MANET, don't make usage of the characteristics of VANETs. As to turf, the maker should consider that cars in differing ways move at unique speed, go down in various bearing. In such a component framework, independent correspondence associations may not be a durable and the controlling ways that movement continually are frail.

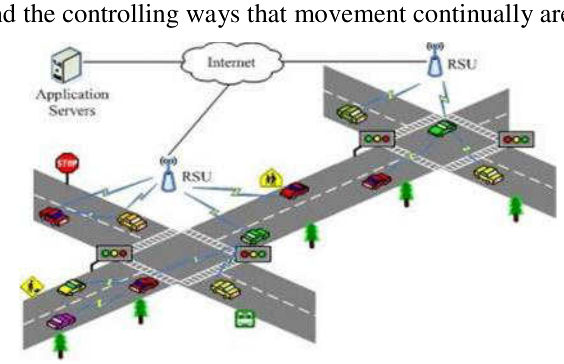


Fig. 1. Structure of Vehicular Ad-hoc Network [5]

II. PROTOCOLS ANALYZED IN THIS PAPER

A. Source Tree Adaptive Routing Protocol (Star)-

The STAR convention is proposed for usage by nodes (static and flexible) in a specially appointed system or a web. A switch in STAR confers to its neighbors the parameters of its source guiding tree, which includes every one association that the switches needs to attain every known end (and area range) in the impromptu system or web. To direct transmission information exchange limit and essentialness, a switches gives movements to its source guiding tree exactly when the switch gets new objectives, the probability of orbiting, or the probability of node disillusionments or framework parts. The STAR[3] steering convention could be run either by using a perfect regulating philosophy (ORA) or by using a base overhead directing strategy (LORA). The central system tries to fulfill an essential to give perfect routes between nodes according to a portrayed metric and won't ponder the volume of steering messages - to be totally straightforward its execution qualities will depend on upon the frequence of guiding tables updates. On the other hand, the LORA approach according to its name strives to keep the overhead of running messages to a base while surrendering the optimality essential of ORA. As LORA slopes towards an all the more direct correspondence with on-enthusiasm steering assemblies and in like manner is furthermore captivating from an algorithmic viewpoint this diagram focuses on STAR-LORA. A serious a discussion of the ORA mode could be found.

B. Zone Routing Protocol (Zrp)

ZRP joins both proactive and touchy methodology of directing. It takes advantage from proactive and touchy methodologies and evacuates the impediment. ZRP[2] limits the reach of proactive steering systems to neighboring hubs provincially, however ZRP utilizes responsive directing to pursuit the coveted hubs by questioning the particular system hubs internationally as opposed to sending the question to all the hubs in system. ZRP utilizes "Intra-zone" and "Inter-zone" steering to give adaptable course revelation and course upkeep in the numerous specially appointed situations. Inter-zone steering performs course finding through responsive directing convention comprehensively while intra-zone steering focused around proactive steering keeping in mind the end goal to keep up avant-garde-course data by regional standards inside its own particular directing extent. The general normal for ZRP is that it diminishes the system overhead that is created by proactive steering and it likewise handles the system postpone that is brought on by touchy directing conventions and perform course disclosure all the more effectively

C. Poistion Based Routing Protocols (Povrp)

The element and profoundly portable nature of VANET, where nodes carry on extremely fast and progressions its area every now and again requests such directing technique that can manage the earth of such system. These solicitations tend the examiners to use positions of nodes as a piece of appeal to give compelling correspondence from start to end. Such system in which geological positions of nodes are used to perform data coordinating from start to end of the line is called position based directing[4]. Position based routing expect that each node have looking into its physical/ geological position by GPS or by some other position choosing organizations. In it each nodes furthermore has the taking in of start, end and other neighboring nodes. As appeared differently in relation to topology based directing, position based steering uses the additional information of every one participating hub to fitting in VANET, that additional information is amassed through GPS. Position based directing gives hop-by-hop correspondence to vehicular frameworks. A position based directing gathering includes various true parts, for instance, "beaconing", "range organization and servers" and "recovery and sending methods"

D. Implementation

The convention execution is examined using the Qualnet test framework structure 5.0.1. The re-enactment parameters used for reproducing the circumstances of vehicular specially appointed system is showed in the table 1.

1). Simulation Parameters

TABLE.I. Shows parametric values

Parameter	Value
Coordinate	1500m X 1500m
Mobility Model	Random Waypoint Model
Simulator	QUALNET 5.0.1
ROUTING PROTOCOL	STAR, ZRP AND POVRP
Network	IPv4
MAC	IEEE 802.11e
NO OF NODES	20 to 100

2). Simulation Scenario

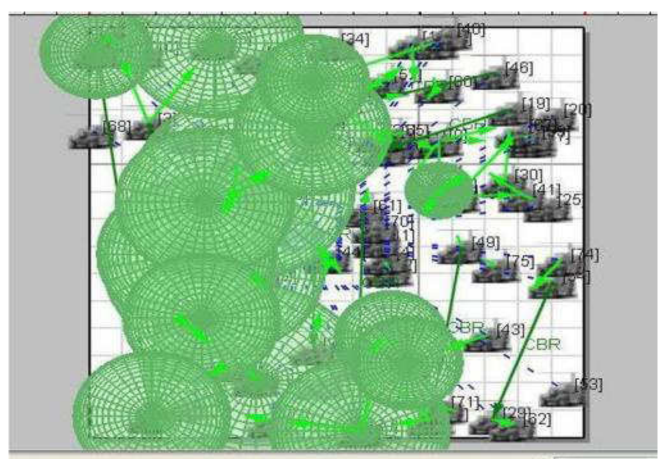


Fig.1. A Scenario illustrating the analysis of STAR, ZRP AND POVRP protocols in Vehicular Ad-hoc Network

III. SIMULATION RESULTS

We assess the execution of different routing protocols, ZRP, STAR and POVRP in VANETS where vehicles are sparsely dispersed.

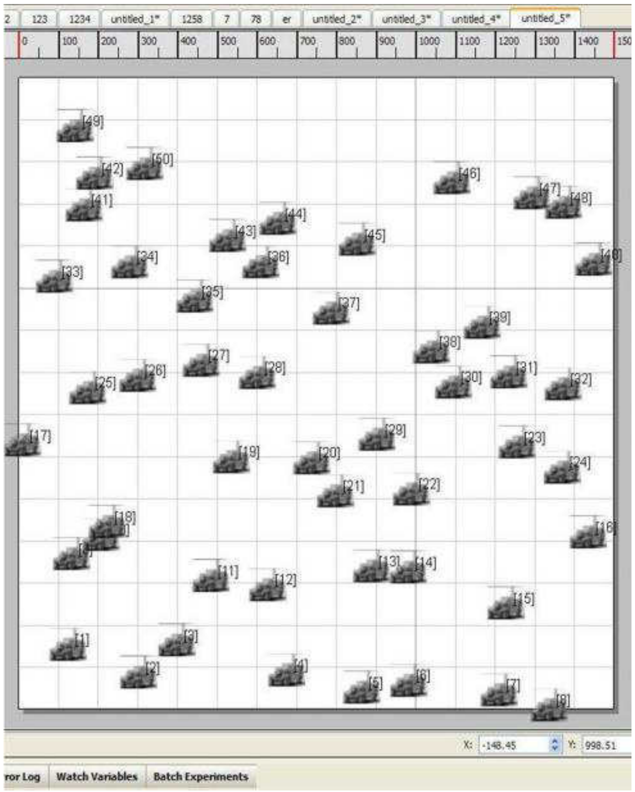


Fig. 3. Sparsely distributed Scenario

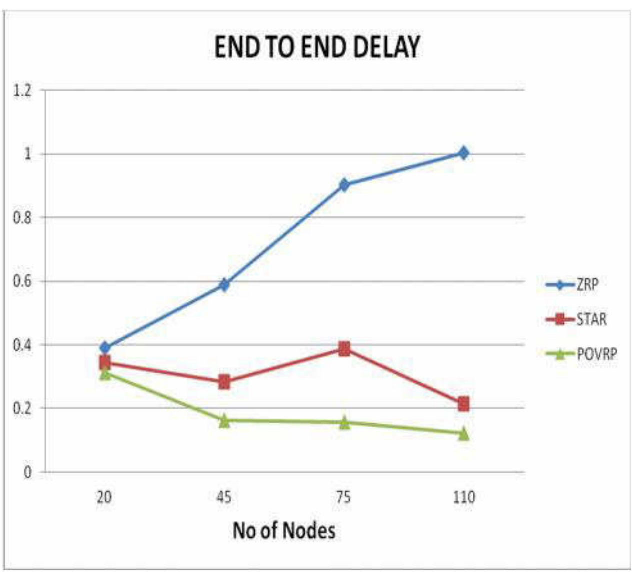


Fig 4. Analysis of End To End Delay

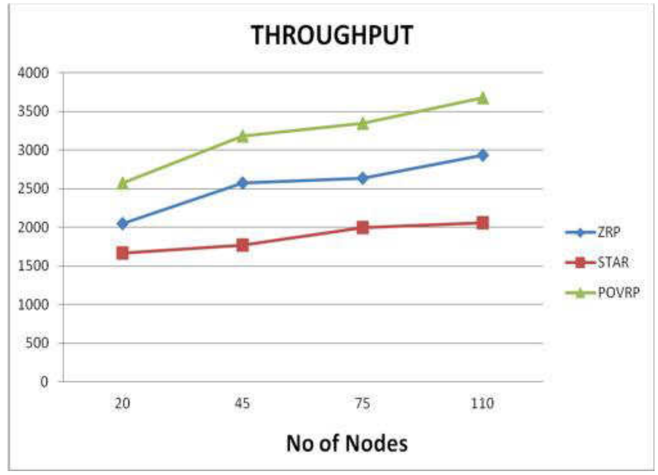


Fig 5. Analysis of Throughput

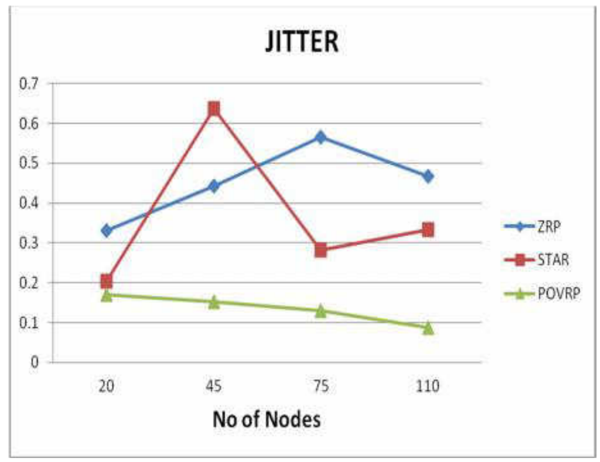


Fig 6: Analysis of Jitter

IV. CONCLUSION

In this paper we have analysed, that POVRP routing protocol is better as compared to STAR and ZRP routing protocols in VANETs. POVRP routing protocol is the position based routing protocol i.e. position based routing expects that each node had looking into its physical/ geological position of GPS or by some other position choosing organizations and STAR is a source tree adaptive protocol is a Proactive based routing protocols and it may failing VANET due to consumption of more bandwidth and large table information. ZRP routing protocol is hybrid protocol i.e. combination of proactive and reactive routing protocol. ZRP protocol works on inter-zone and intra-zone concept that increase the routing overhead and hence its performance gets degraded. Due to this reason the POVRP is outperformed in all the three metrics i.e. End to end Delay, Throughput and Jitter.

REFERENCES

[1] Bijan Paul, Md. Ibrahim, Md. Abu Naser Bikas " VANET Routing Protocols: Pros and Cons" International Journal of Computer Applications, (0975 – 8887) Vol. 20, No.3, April 2011

[2] Z. J. Haas, “The Zone Routing Protocol”, Sixth WINLAB Workshop on Third Generation Wireless information Networks, New Brunswick, NJ, Nov. 1997.

[3] Seet, B.-C., “A-STAR: A Mobile Ad Hoc Routing Strategy for Metropolis Vehicular Communications.” 1980.

[4] Bilal Mustafa, " Issues of Routing in VANET" Master thesis Computer Science june 2010, thesis no-MCS-2010-20 .

A Survey of Issues and Challenges in Development of Futuristic Multihop Wireless Networks

Sonam Bhatia
Dept. of CSE
GNDU, RC Jalandhar
dbhatia297@gmail.com

Jyoteesh Malhotra
Dept. of CSE
GNDU, RC Jalandhar
jyoteesh@gmail.com

Abstract—Wireless multi-hop networks, in various forms and under various names, are being increasingly used in military and civilian applications. A multi-hop wireless network consists of a number of nodes that form a dynamic topology with ad hoc wireless links among neighboring nodes. These nodes are typically limited in resources such as bandwidth and battery. In multihop wireless networks, packets of a flow originating from a source node are relayed by intermediate nodes and travel towards their destination along a multihop wireless path. In this paper, brief overview of multihop wireless network. This paper mainly focuses on challenges of Multihop Wireless Networks and also provides the suitable solution to various problems.

Keywords—Multihop Wireless Network, Routing, Energy, Transport layer.

I. INTRODUCTION

The term wireless networking refers to technology that enables two or more computers to communicate using standard network protocols, but without network cabling. It's cheap and doesn't require wires to connect to your computer, instead wireless networks radio waves to send and receive network signals .As a result computer can connect to wireless network at any location in home and office. That’s why wireless networking has become so popular.

A multihop wireless network is a network of nodes connected by wireless communication links. With the help of digital packet radios, the links are usually implemented, Due to limited range of radios, many pairs of node in multihop wireless networks may not be able to communicate directly, and hence they may need other intermediate nodes to forward packets for them. [1] In such a network, packets are transmitted through multiple hops to reach their destinations. The multihop path for each flow is determined by the routing protocol. The neighbor node with the shortest distance to the destination is selected as the relaying node for the next hop for each node in most of the existing multihop routing protocols. Such selection results in a longer per-hop progress and less hop counts and, thus, smaller end-to-end delay for packet delivery [2]. Wireless multi-hop networks, in various forms, e.g. wireless sensor networks, vehicular networks, mesh networks , and under various names e.g. ad-hoc networks, hybrid networks, delay tolerant networks, are being increasingly used in military and civilian applications. Characteristic features of a wireless multi-hop network defined as First, Wireless devices are self-organized or assisted by some infrastructure to form a network. The ad-hoc networks are based on multi-hop networks. Second, Communication is mostly via wireless multi-hop paths. Third Packets are forwarded from the source to the destination. [3]

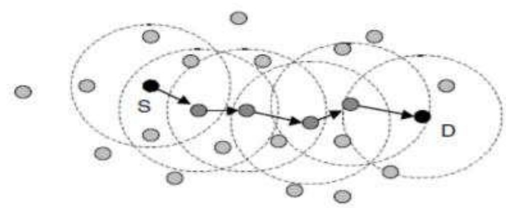


Fig. 1. Multihop packet delivery

The rest of this paper is organized as follows: In Section II , we present the constraints and challenges in multihop wireless networks In Section III, we present work related to these challenges and In Section IV, we discuss issues in tabular form .Open issues and future scope are provided in Section V. Conclusion is presented in Section VI. Open issues and future scope related to wireless multihop networks have been discussed before concluding in section 6.

II. CONSTRAINTS AND CHALLENGES OF MULTIHOP WIRELESS NETWORKS

A. Routing Challenges in Wireless Multihop Networks:

One of the important challenges is how to route packets efficiently in a wireless multi-hop network. The availability of many intermediate nodes between a source and a destination results in having many optional paths/routes to follow. The challenge is to pick an optimal path that satisfies the needed performance requirements - this is the responsibility of the routing protocol. With multi-hop wireless ad hoc networks, messages may be transmitted via multiple radio hops, and thus a routing protocol is essential for the success of such networks. In the last several years after substantial research efforts, routing for multi-hop wireless networks becomes a broadly investigated problem.

B. Routing in multi-hop wireless networks presents a great challenge mainly due to the following facts:

First, wireless links are unreliable because of channel fading .Second; achievable channel rates may be different at different links since link quality depends on distance and path loss between two neighbors. Third, as the wireless medium is broadcast in nature, the transmission on one link may interfere with the transmissions on the neighboring links [4]

C. Energy Consumption in Wireless Multihop Networks:

Energy consumption in multihop wireless networks is a crucial issue that needs to be addressed at all the layers of a communication system, from the hardware up to the application. [5] A key issue in developing control protocols for multihop wireless networks is to reduce the energy or power consumption. This is obviously an important issue for battery powered networks since the power consumption often limits the lifetime of the network [11]

To keep wireless devices networked, they need to consume energy. On the other hand, vast deployment of multi-hop wireless networks in the near future will cause an explosion in the number of devices around the world. Consequently, energy consumption by these devices will increase .Each node needs to be aware of its scarce energy budget while it operates in a multi-hop wireless network. Relay traffic in the network should be balanced between nodes to prevent quick failure of those nodes which forward many packets on behalf of other nodes. Failure of nodes due to battery exhaustion not only isolates them from other nodes but also harms multi-hop connectivity of the network.

In many applications of wireless multi-hop networks nodes may run on batteries. High energy consumption reduces operational lifetime of battery-powered devices, and hence the autonomy of the system relying on them. [6]

D. Transport Layer Constraints in Wireless Multihop Networks:

In this section, a thorough understanding of the major problems that arise with Transmission Control Protocol (TCP), when applied over multihop wireless networks, is discussed

E. Routing failures:

Packets that are buffered at intermediate nodes along the route will be dropped when a route failure occurs. This large amount of packet drops may cause a series of time-outs at the TCP sender. As a result, the retransmission time-out (RTO) value will be doubled for each subsequent time-out. In addition, TCP does not have any indication of the route re-establishment duration, because the route re-establishment event after route failures depends on the underlying routing protocol.

F. Random wireless losses:

Due to channels fading and interference, Channel errors of wireless usually occur randomly and in bursts, which cause it to be error prone in nature. High bit error rate (BER) introduces two challenges to the reliable TCP protocol. First, Packets loss due to corruption violates the assumption of TCP's congestion control mechanism in that congestion is the main reason for most losses. In this case, TCP will mistakenly take this as an indication of network congestion and reduce the sending rate unnecessarily whenever a packet loss is detected,. Second, In the presence of severe wireless channel contention, where the BER is high and the link-layer local recovery mechanism is unable to recover the lost packets, TCP will face a large amount of consecutive packet losses. The current TCP's "error detection and recovery" mechanism is inefficient to recover from these losses. Either only one lost packet is recovered per round-trip time (RTT), or a time-out is needed to recover from substantial losses.

G. Hidden and exposed terminal:

In wireless networking, the hidden node problem or hidden terminal problem occurs when a node is visible from a wireless access point (AP), but not from other nodes communicating with that AP. This leads to difficulties in media access control

III. RELATED WORK

Considering the challenges and constraints highlighted in previous section of mutihop wireless networks Literature has been surveyed based on Routing, Energy, and Transport layer. The work done in order to resolve these issues by various authors has been complied in this section.

Jia-Chun Kuo, Member, IEEE, and Wanjiun Liao, Member, IEEE [2] in 2007 presented the multihop path for each flow is determined by the routing protocol and also tell about wireless multihop networks and develop an analytical framework for the hop count distribution in a multihop wireless network with an arbitrary node density.

Kai Zeng, Member, IEEE, Wenjing Lou, Senior Member, IEEE, and Hongqiang Zhai, Member, IEEE [4] in 2008 presented in a wireless multi-hop network, one of the important challenges is how to route packets efficiently. And the impact of candidate selection and maximum end-to-end throughput of OR.

Longbi Lin, Xiaojun Lin, and Ness B. Shroff [11].in 2008 explained the problem of minimizing the total power consumption in a multi-hop wireless network subject to a given offered load. And presented total power consumption of multi-hop wireless networks can be substantially reduced by jointly optimizing power control, link scheduling, and routing.

Sherin Abdel Hamid, Hossam, Glen Takahara in 2011 [9] presented solution to routing problems like distance vector, link state routing, also covered core components like routing discovery routing selection. The basic distance vector routing works in theory but has a serious drawback in practice.

Behrouz Maham, Member, IEEE, Are Hjørungnes, Senior Member, IEEE, and Ravi Narasimhan, Member, IEEE [5] in 2011 presented an energy issue in multihop networks. Also formulated the problems of finding the minimum energy cooperative route for a wireless network under Rayleigh fading, and proposed a space-time coded cooperative multihop routing for the purpose of energy savings, constrained on a required outage probability at the destination.

Ammar Mohammed Al-Jubari1 Mohamed Othman1, Borhanuddin Mohd Ali2 and Nor Asilah Wati Abdul Hamid [7] in 2011 explained major problems that arise with TCP, when applied over multihop wireless networks, is discussed and provide solution to problem. The approaches cross-layer approaches and layered approaches applying TCP algorithms to the problem of multi-hop conditions and characteristics.

Maggie Cheng, Quanmin Ye, and Lin Cai [8] in 2013 also presented the cross layer and layer approach address how to minimize end-to-end delay jointly through optimizing routing and link layer scheduling. The two cross-layer schemes, a loosely coupled cross-layer scheme and a tightly coupled cross-layer scheme also covered. Through simulation, show that the proposed routing and scheduling schemes can outperform their counterparts in each layer, and the integrated cross-layer schemes are superior to the combination of the existing routing and scheduling schemes.

Kai Zeng, Wenjing Lou, Ming Li [1] in 2011 explained multihop wireless networks and also present energy efficiency ,capacity.

Javad Vazifehdan [6] in 2011 prsented energy consumption and energy awareness in multihop wireless networks is a crucial issue that needs to be addressed at all the layers of communication system, from the hardware up to the application. The work presented in this dissertation provides a platform for energy-efficient and energy-aware communication in personal types of wireless multi-hop networks.

Guoqiang Mao [3] in 2011 presented features of wireless multi-hop network and overview recent development and discuss research challenges and opportunities in the area, with a focus on the network connectivity.

Josh Broch David A. Maltz David B. Johnson Yih-Chun Hu Jorjeta Jetchev[10] explained Many different protocols have been proposed to solve the multihop routing problem in ad hoc networks, each based on different assumptions and intuitions term like distance vector and link state and find some drawbacks in algorithm also.

IV. ISSUES

Various authors have discussed about the issues in the survey as given above. Table given below summaries these issues.

Table I. Issues of Multihop Wireless Networks

Paper Name	Issues covered	Remarks
Capacity of Opportunistic Routing in Multi-Rate and Multi-Hop Wireless Networks[2008]	Routing in multihop wireless networks	OR has great potential to improve the end-to-end throughput under different settings, and proposed multi-rate OR schemes achieve higher throughput bound than any single-rate geographic OR

Low-Complexity and Distributed Energy Minimization in Multi-hop Wireless Networks [2008]	Issue cover how to reduce the energy or power consumption	The performance bound of solution, achieved in a distributed method, is provably tighter than a centralized solution.
Energy-Efficient Space-Time Coded Cooperation in Outage-Restricted Multihop Wireless Networks [2011]	Energy consumption	Space time coded Cooperative multihop routing for the purpose of energy savings, constrained on a required outage probability at the destination. The calculated distributed and centralized power allocations are independent of instantaneous channel variation It is shown that energy savings of up to 72% is achievable in line networks with 3 relays
Energy-aware Wireless Multi-hop Networks by Javad [2011]	covers energy consumption and energy awareness in multihop wireless networks is a crucial issue	study of energy-aware and Energy-efficient design and analysis of wireless multi-hop networks.
TCP performance in multi-hop wireless ad-hoc networks: challenges and solution [2011]	Routing failure, hidden and exposed terminal problem	Existing TCP enhancements are not enough in isolation to provide a complete reliable TCP solution. Therefore, designing a suitable solution to improve TCP performance in multi-hop wireless networks is still a task for future research.

IV. OPEN ISSUES AND FUTURE SCOPE

Various authors have discussed about the challenges of multihop wireless networks in literature. As given in pervious section. However there is ample scope of improvements in literature in the Routing, Energy consumption, Transport layer issues that has been presented here in this section.

- In a wireless multi-hop network, one of the important challenges is how to route packets efficiently.
- More Energy consumption reduces the lifetime of battery-powered devices.
- A fundamental issue in multi-hop wireless networks is that performance degrades sharply as the number of hops traversed increases.
- Reliable data transfer is one of the most difficult tasks to be accomplished in multihop wireless networks.

Many applications of multihop wireless networks will be designed which have expected to have good routing techniques, and consume less energy so that battery lifetime can be maximized. Many researchers are working to remove the issues in multihop wireless networks but still many issues exist which need to be solved in future.

V. CONCLUSION

Wireless multi-hop networks have attracted significant research interest As the Internet users' requirements of flexibility and mobility increase, multi-hop wireless networks have become the best solution to satisfy these requirements. These networks are complex distributed systems that consist of wireless mobile or static nodes and have some constraints. In this survey paper we have covered issues like how to route packet efficiently, Energy consumption, major problems that arise with Transport layer, when applied over multihop wireless networks, is discussed .It is hoped that survey done here in this paper will prove to be helpful to researchers working in area of wireless multihop networks.

REFERENCES

[1] Kai Zeng, Wenjing Lou, Ming Li, "Multihop Wireless Networks: Opportunistic Routing", 2011.

[2] Jia-Chun Kuo, Member, IEEE, and Wanjiun Liao Member, IEEE "Hop Count Distribution Of Multihop Paths in Wireless With Arbitrary Node Density: Modeling and Its Applications", 2007.

[3] Guoqiang Mao "Wireless Multi-hop Networks: Current Research and Future Challenges".

[4] Kai Zeng, Wenjing Lou, and Hongqiang Zhai, "Capacity of Opportunistic Routing in Multi-Rate and Multi-Hop Wireless Networks", IEEE Transactions Wireless Communication, VOL. 7, NO. 12, December 2008.

[5] Behrouz Maham, Are Hjørungnes, and Ravi Narasimhan "Energy-Efficient Space-Time Coded Cooperation in Outage-Restricted Multihop Wireless Networks" IEEE Transactions On Communication, VOL. 59, NO. 11, November 2011.

[6] Javad Vazifehdan, "Energy-aware Wireless Multi-hop Networks", July 2011.

[7] Ammar Mohammed Al-Jubari1*, Mohamed Othman1, Borhanuddin Mohd Ali2 and Nor Asilah Wati Abdul Hamid "TCP performance in multi-hop wireless ad hoc networks: challenges and solution", Journal on Wireless Communications and Networking 2011.

[8] Maggie Cheng, Quanmin Ye, and Lin Cai, "Cross-Layer Schemes for Reducing Delay in Multihop Wireless Networks", IEEE Transactions Wireless Communication VOL. 12, NO. 2, February 2013.

[9] Sherin Abdel Hamid, Hossam, Glen Takahara, "Routing for wireless multihop networks-unifying and distribution features", December 2011.

[10] Josh Broch David A. Maltz David B. Johnson Yih-Chun Hu Jorjeta Jetchev, "A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols"

[11] Longbi Lin, Xiaojun Lin, and Ness B. Shroffs, "Low-Complexity and Distributed Energy Minimization in Multi-hop Wireless Networks", Center for Wireless Systems and Applications (CWSA) School of Electrical and Computer Engineering, Purdue University in USA, 2008.

Future Enhancements with Sixth Sense Technology

Gaurav Kumar
Dept. Computer Science Engineering
CTIEMT, Shahpur
Jalandhar, India
kumargourav@engineer.com

Ravinder Kaur
Electronics & Communication Engineering
CTIEMT, Shahpur
Jalandhar, India
er.rkaur@gmail.com

Abstract—**The work done here familiarizes with current sixth sense technology, which offers freedom of intermingling with the digital world using only hand gestures. The sixth sense technology is the one which empower user to interact to the digital world in most efficient way. Devices using Sixth Sense technology are very much different from the computers. This paper provides the basic idea of sixth sense technology, its advantages and future enhancements.**

Keywords— ***Introduction, Elements, Working, Future Enhancements.***

I. INTRODUCTION

When humans come across something, someone or some place, the five natural senses are utilized [1] which include eye, ear, nose, tongue, mind see Fig.1 and body to perceive information about it; that information assist to make decisions and chose the appropriate actions to take. Perhaps the most worthwhile information that can assist to make the right decision is actually not naturally perceivable with the five senses, namely the data, information and knowledge that mankind has collected about everything and which is gradually increasing and is all available online.

Even though the miniaturization of computing devices permits us to carry computers in our pockets, custody us persistently and keep connected to the digital world but still there is no link between our digital devices and our interactions with the physical world. Information is restricted traditionally on paper or digitally on a screen. Sixth Sense bridges this gap, take along intangible digital information out into the tangible world, and letting us to intermingle with this information thru natural hand gestures. Sixth Sense liberates information from its precincts and impeccably integrates it with reality, and thus making the entire world your computer.

Sixth Sense technology is a wearable gesture centered device that augments the physical world with digital data and facilitates people to use natural hand gestures and interact with that digital information. Currently digital devices such as computers, mobile phones, tablets, etc. are accessed to use internet and get information that is required. With Sixth Sense technology device used will be no larger than cell phone and probably be as small as a button on the shirt to bring internet to the user in order to interact in real world. This technology will allow us to interact with our world like never before. Information on anything required, from anywhere will be available within a few seconds. The most interesting part of the device is its capability to scan objects or even people and project out information regarding what is being required [2].

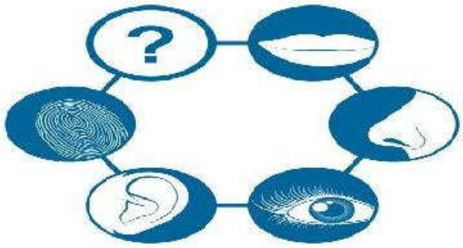


Fig.1 Human six senses

II. SIXTH SENSE

Sixth Sense is well-defined as Extra Sensory Perception or ESP. It encompasses the reception of information not gained through any of the five senses. Not even it is taken from any experiences from the past or known. Sixth Sense aims to more flawlessly integrate online information and tech into everyday life. By making existing information essential for decision-making beyond what we have access to with our five senses, it efficiently gives users a sixth sense [3].

III. BASIS OF CONSTRUCTION AND WORKING

The Sixth Sense technology comprises of a pocket projector, a mirror and a camera enclosed in like a pendant wearable device. The projector, camera and sensors are connected to a coding device (such as laptop, or other digital device) in the user's pocket. The projector projects visual information on supporting surfaces like wall, physical objects around user which can be used as interfaces, at the same time the camera recognizes and tracks user hand gestures plus physical objects using computer-vision based techniques. The software program compile the video streamed data recorded by the camera and tracks the locations of the color markers(Red, Blue, Green) situated at the tips of the user fingers. The movements and arrangements of these markers are inferred into gestures that act as interaction directions for the projected application interfaces.

The role of different hardware device used in sixth sense technology is given below:

i. Portable Projector

The portable projector is use for various type of visual information to be displayed which is captured from the camera provided.

ii. Color Markers

The color markers red, blue and green are used for gesturing the objects, data or any other work. The color marker is used as accordingly to programming in the hardware in which hardware components can easy to gesture and work agreeing to the user need.

iii. High Mega Pixel Camera

The camera in the sixth sense technology play an important role for gesturing and tracking the user finger tips covered with color markers. As the moment of these markers decide what work is to be done. The camera captures the data and transmits it to the other hardware device for further software processing.

iv. Advanced Cell Phone

The cell phone (or any other digital device used) manages the transmitting data, receiving data, gesturing, video recording and other gestured data. This digital device completely plays the role of a CPU which manages the requirement of user.

v. Sensor

Sensor is hardware device designed and programmed accordingly to response when it sense the gesture or the point of objects.

vi. Reflected Mirror

The mirror is implanted under the projector shown in Fig.2 to reflect the images agreeing to the user needs.

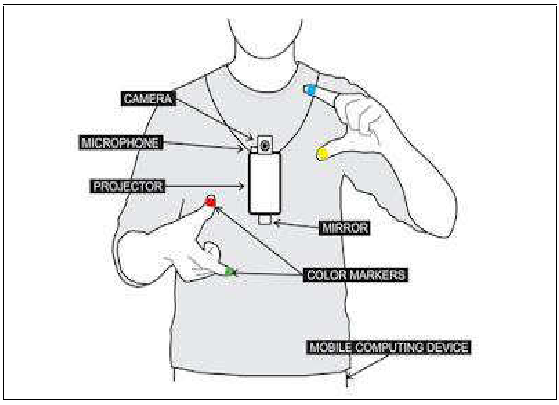


Fig.2 Elements of sixth sense technology

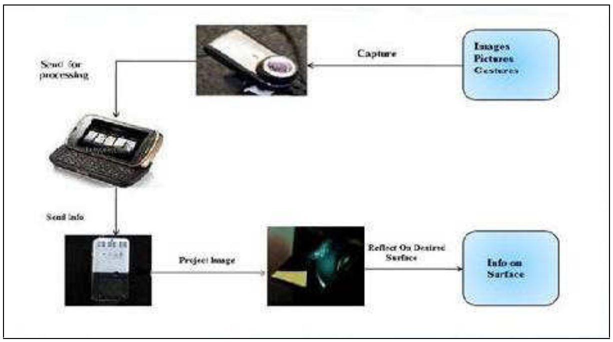


Fig.3 Working of sixth sense technology

IV. ADVANTAGES AND FUTURE APPLICATIONS

The Sixth Sense prototype is used to implement several applications that have shown the effectiveness, viability and flexibility of the system.

- With the help of this technology in the coming future easily voice and video calls can be made without need of mobile phone, right away click on your hand and it will automatically do the needful shown in Fig.4.

Fig.4 Voice/ Video Calling application

- As this device has small size hence it is portable and can be taken anywhere in the world .
- Games can be played without any need of remote or other device just watch and play. Moreover it can easily draw the pictures or other data efficiently.
- Now passengers need not to worry about delay in airplanes or trains, as if airlines flight or train is cancelled, data will be automatically send to the passenger that today flight is cancelled due to weather condition shown in Fig.5.

Fig.5 Air ticket displaying delay

- This technology is also available open source. The device help to read, moreover can decoded the barcode and show the results .If user want to see the movie, just click and watch the movie and enjoy .
- The map application permit the user to navigate a map that will be displayed on a close by surface by using only hand gestures, allowing the user to zoom in, zoom out or pan using instinctive hand movements [6].
- Another application is to check the time; by moving the limbs and acting as to drawing a circle on the user's wrist will displays an analog watch.[6] The computer recognizes the gesture, keeping the track of the red marker cap and sends command to projector to display the image of a watch onto the user's wrist.

V. RELATED TECHNOLOGIES

There are some upcoming technologies that are associated with sixth sense technology.

i. Augmented Reality

Augmented Reality is a term for a live direct or indirect view of a physical real world environment whose components are linked by virtual computer generated imagery [4].

ii. Gesture Recognition

In computer science it is a language technology which interprets human gestures by using mathematical algorithms.

iii. Computer Vision

In this technology, the required information is interpreted / extracted from an image by machines.

iv. Radio Frequency Identification

It is an electronic tagging technology [4] that allows the tracking, detection of tags and consequently the objects that they are connected to.

VI. FUTURE ENHANCEMENTS

As this technology is taking pace future enhancements can be foreseen few are discussed here:

- *Hand finger print device* recognize the user finger and revels the entire detail related to the person to the user. This will help blind users as whenever blind user meet another person device will help you to recognize the person i.e have this person earlier met or he is new one.
- To get free from of color markers.
- To integrate camera and projector inside mobile Computing device.
- Whenever the pendant- style wearable device is placed on table, it should allow user to use the table as multi touch user interface.
- Applying this technology in various interests like versatile gaming, education systems etc.
- Enhance to 3D gesture tracking.
- To make sixth sense work as fifth sense for disabled person. [5]

VII. CONCLUSION

In the work done, new and advanced technology Sixth Sense technology has been discussed, which help user to do work according to their fingertips moment. The Sixth Sense Device makes work much easy, efficient and convenient than before. This technology can be used as a replacement of the fifth senses for handicapped peoples The device take full care of the ease and working conditions according to need of the user. This will enable individuals to make their own application depending upon needs and imagination.

REFERENCES

[1] Pranav Mistry, “Sixth Sense Technology”, http://www.pranavmistry.com/projects/sixthsense/

[2] Raxa K. Bathani, “Sixth Sense Technology OR WUW (Wear Ur World)”, Research Expo International Multidisciplinary Research Journal, Volume - II , Issue - II June – 2012. ISSN : 2250 -1630

[3] Monika Arora, “Basic Principles of Sixth Sense Technology”, VSRD-IJCSIT, Vol. 2 (8), 2012, 687-693

[4] Rohit Gupta,“ Sixth Sense Technology”, seminar report http://www.authorstream.com/Presentation/Karankinng-368341-sixth-sense-technology-report/

[5] K.R Arjun “Sixth Sense Technology”, http://dspace.cusat.ac.in/jspui/bitstream/123456789/2207/1/SIXTH%20SENSE%20TECHNOLOGY.pdf

[6] P. Mistry, P. Maes. SixthSense – A Wearable Gestural Interface. In the Proceedings of SIGGRAPH Asia 2009, Sketch. Yokohama, Japan. 2009

Power Consumption Analysis of Static and Mobile Wireless Sensor Networks with Converge Cast Routing

Jatinder Pal Singh
Dept. of ECE, CT Institutions Jalandhar
Punjab, India
jhsing11@gmail.com

Love Kumar
Dept. of ECE, DAVIET Jalandhar
Punjab, India
er.lovekumar@gmail.com

Abstract- **A Wireless Sensor Network (WSN) is an ad-hoc network. WSN is a network that contained battery-powered nodes which route the data from sensor node to sink node. Therefore, power management is a major issue for both static and mobile wireless sensor networks. In this network sensor nodes and sink can be static or both are in mobile mode. It is depends on the application of the wireless sensor network. In this paper power measurement is reported for both static and mobile wireless sensor networks. Experiment is performed for 802.15.4 physical layer energy consumption with modulation type, number of nodes and distance between them.**

Keyword- Power Management, Converge-cast Routing, Static Nodes, Mobile Nodes and Sensor Network.

I. INTRODUCTION

Wireless sensor networks consist of large numbers of sensors that act cooperatively to provide "usable chunks of predigested information rather than a confusing wash of number". A WSN provides refined information, i.e. it processes the raw data collected by individual sensors before presenting it to the user.

In this network sensor nodes and sink can be static or both are mobile. This amounts to providing a service or a collection of services based on sensor capabilities and on the underlying communications infrastructure. Due to their vast array of applications, WSNs have been viewed from different perspectives and standpoints a Wireless Sensor Networks (WSN) may be described as a network of small, autonomous, battery-powered nodes and routed to a sink, which typically lacks energy constraints [1]. If some sources do not coordinate their routes to sinks, it is possible that one or more nodes may be exhausted due to overuse and the network may be partitioned. When this occurs, the energy in sources that are disconnected from sinks that is wasted of energy and there is no mean to route the data. Consider the network shown in Fig.1.It consists of ten nodes out of which three nodes are sources and one node is a sink (Z). An edge between two nodes indicates that those nodes can communicate with other, assumed that communication is bidirectional. Within the network, it is obvious that certain nodes consume more power to maintaining connectivity than others.

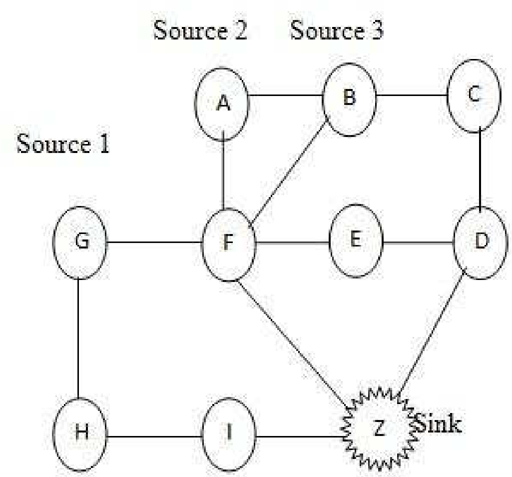


Fig 1.Wireless Sensor Network

Most of the current research in wireless sensor networks (WSN, for short) is constraint driven and focuses on optimizing the use of limited resources (for example, power) at each sensor. While such constraints are important, there is a need for more general performance metrics describing the effectiveness of WSNs.

Each *source* determines the optimal path routing to a *sink* .For example, the optimal routes to the sink might be GFZ from source G, AFZ from source A and BFZ from source B. However all the above routes require the use of node F. To reducing the energy expenditure of node F, it is require source G to use the non-optimal route GHIZ.

The paper is organized as follow: section giver general introduction to wireless sensor network while section II describe simulation scenario and result for variation in number of node. In last the conclusion has given in section III.

II. SIMULATION SETUP & RESULT

A number of routing schemes have been proposed that attempt to maximize the efficiency of WSNs. Although many schemes are derived from a combination of others, they can be loosely grouped into the categories of minimum hop routing, minimum energy routing, loading balancing routing and potential based routing. In this paper power measurement has been done for Converge-cast routing.

Since the energy required for transmission and the transmission distance, a longer sequence of small hops may require less energy than a short sequence of long hops [2]. In the converge-cast routing, all the sensor nodes send the data to the single Sink node.

In case of optimal path routing an intermediate sensor node that used by the number of sources it will expire quickly [4] while non-optimal path routing, to route the data toward sink node, the number of intermediate node will increased. To solve this problem and to control the power consumption, converge-cast routing can be used to maintain the energy balance [6][5] in this scenario.

The sink node coordinates all the sensor nodes in star, tree or hybrid topology. In this paper we investigate the power consumption on each sensor node. In the figure.2 a typical wireless sensor network is shown with converge-cast routing with OmNet++ for 10 nodes.

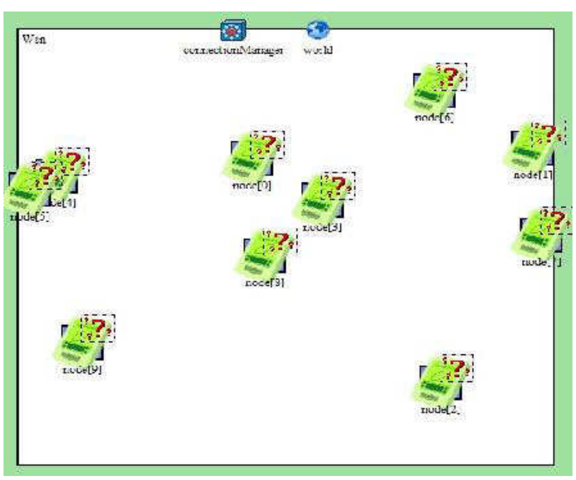


Figure 2.Wireless Sensor Network

Similar Scenario has been used for 20, 30, 40, 50, 60, 70, 80 90, 100 sensor nodes. Powers consumptions for all the scenario has been reported in next section.

III. POWER CONSUMPTION FOR STATIC NODES

The power consumption on each static sensor node is shown in figure.3 and table no.1 with simulation time of 10 seconds. In this as the number of sensor nodes are increased in the network, they consume more power per unit time span.

TABLE No.1 Power consumption for different Static Sensor node

Sensor node	Power consumption (mWs)	Simulation Time(s)
1	32.14	10s
2	32.33	10s
3	32.56	10s
4	32.14	10s
5	32.14	10s
6	32.4	10s
7	32.56	10s
8	32.14	10s
9	32.33	10s
10	32.56	10s

The power consumption [1][8] of the sensor node, it's depends upon that how many time that sensor node route the data.

The power consumption [9] is also dependent on routing [7] type that is optimal path and non optimal path routing [3]. It has been analyzed from figure 3 that there is less power variation for 10 sensor nodes.

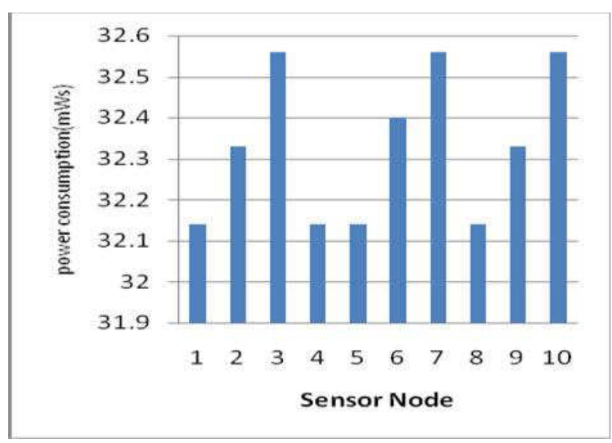


Fig 3. Power consumption

Similar as comparison of power for 10,20,30,40,50,60,70,80,90 &100 sensor nodes given in figure 4.

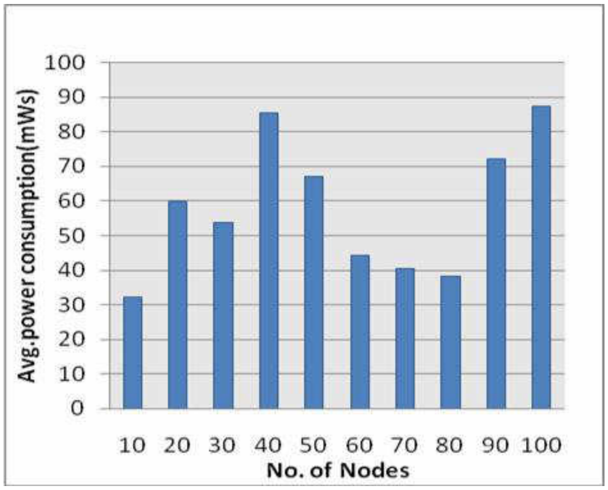


Fig 4. Power consumption

It has been depicted from figure 4. There is large variation in the power consumption as we increase the number of node. The numerical values of consume power is tabulated in table 2.

Table 2: Power consumption for different static sensor node

No. of node	Avg. power consumption(mWs)	Simulation Time(s)
10	32.14	10s
20	59.88	10s
30	53.76	10s
40	85.34	10s
50	67.01	10s
60	44.15	10s
70	40.44	10s
80	38.4	10s
90	72.24	10s
100	87.45	10s

It has been analyzed at 100 nodes in the simulating scenario maximum power is consumption and a lot of variation has been observed for different sensor node.

IV. MOBILE SENSOR AND SINK NODES

If a sensor node or sink node moves fast enough to deliver data with tolerable delay, Wireless Sensor Networks are restricted against mobility range and power consumption in the network [10]. In this mobility relay approach, the mobile sink gets the data from mobile sensor nodes and transports data with physical movements. This approach trades data delivery latency for the reduction of energy consumption of nodes. There are two types of scenario in the wireless sensor networks.

In the first scenario approach only sink node can be move and remaining sensor nodes becomes static. When the sink node comes in the range of the sensor node then the sensor nodes sends data to the sink nodes. If the sink node is not in the range in sensor nodes then they are not sends the data and becomes in the sleep mode.

In the second scenario approach both sink node and sensor node can be moved. It can be used where the continuous monitoring is required and data sends to the sink node in continue manner.

V. POWER CONSUMPTION FOR MOBILE NODES

The power consumption on each mobile node is shown in figure.5 and table no.3 with simulation time of 10 seconds. The simulation 10 nodes move with different mobility speed and they up dates their status with every 1 second in the network. Lot of power variations has been observed in this simulation with respect to the speed.

Table 3: Power consumption for different mobile sensor node.

Simulation Time(s)	No. of Nodes	Mobility Speed(mps)	Update Interval	Avg. power consumption (mWs)
1	10	5	1s	507.95
2	10	10	1s	507.7
3	10	15	1s	507.59
4	10	20	1s	558.7
5	10	25	1s	514.63
6	10	30	1s	560.4
7	10	35	1s	512.62
8	10	40	1s	558.8
9	10	45	1s	561.63
10	10	50	1s	539.31

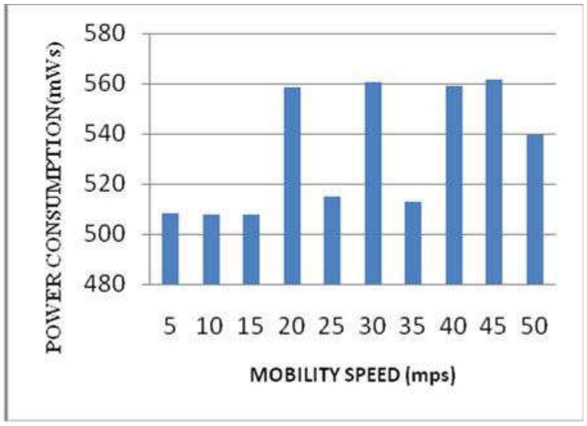


Fig 5. Power consumption in mobile nodes

V. Conclusion

Simulation has been performed to find the optimal value of the IEEE standard 802.15.4 physical layer energy consumption and it has been reported that how it is affected by the physical layer, number of nodes and distance between them. It has been analyzed that mobile nodes power consumption is more than the power consumption in the static nodes. Also as the mobility speed increases, according to that overall power consumption increased. So static nodes are better than mobile nodes and they are more suitable for static node applications. Further the scenario can be optimized for power for different sensor node and distance between them.

REFERENCES

[1] E. Shih et. al., "physical layer driven protocol and algorithm design for energy efficient WSNs , proc. ACM mobicom01,pp. 271-286, 2001

[2] C.E. Perkins and E. M. Royer, "Ad-hoc On –Demand Distance Vector Routing ," in Proceedings of the Second IEEE Workshop on Mobile Computer Systems and Applications, Orleans , USA , pp. 90-100,1999.

[3] D. B. Johnson, et al., "DSR: The Dynamic Source Routing Protocol for Multi-hop Wireless Ad hoc Networks, "in Ad hoc Networking, C.E. Perkins, Ed., ed: Addison –Wesley, pp. 139-179, 2000.

[4] A. Woo , et al., "Taming the Underlying Challenges of Reliable Multi hop routing in sensor networks, "in Los Angeles, California , USA, Conference on Embedded Networked Sensor System, pp. 14-27, 2003.

[5] W. R. Heinzelman, et al., "Energy-Efficient Communication Protocol for Wireless Micro sensor Networks, "in 33rd international Conference on system Sciences, Hawaii, USA, pp. 548-552, 2000.

[6] H. Dai and R. Han, "A Node –Centre Load Balancing Algorithm for wireless sensor Networks," in IEEE GLOBECOM – Wireless Communication, San Francisco, USA, pp.548-552, 2003.

[7] S. Singh, et el., "Power-Aware Routing in mobile Ad Hoc Networks, " in Mobile Computing and Networking, Dallas, Texas, United States, 1998,pp. 181-190, 1998.

[8] G. J. Pottie and W.J. Kaiser, "Wireless Integrated Network Sensor," Communications of the ACM, vol. 43, pp. 51-58, 2000.

[9] B. Zhang and H. T. Mouftan, "Energy-aware on – demand routing protocols for wireless ad hoc networks," Wireless Networks, vol. 12, pp. 481-494, 2006.

[10] M. Grossglauser and D. Tse , "Mobility increases the capacity of ad-hoc wireless networks," IEEE/ACM Trans. Netw., vol. 10 , no.4,pp. 477-486, Aug 2002.

Comparative Analysis of Different Internetworking Protocols

Romil Kumar
Dept. of ECE
CTIEMT, Shahpur
Jalandhar, India
kumar_romil@yahoo.com

Charanjit Singh
Dept. of ECE
CTIEMT, Shahpur
Jalandhar, India
cssohal1984@gmail.com

Sangram Singh
Dept. of CA
CTIMIT,Maqsudan
Jalandhar, India
sangrampasricha@gmail.com

***Abstract*—In modern internet era, a routing protocol determines how the routers communicate with each other by sending update packets about the routes between the source node and the destination node. A routing protocol plays a significant role in selecting an optimal path for the packets, when a number of routes exist between the source node and the destination node while sending minimum routing update packets but still maintaining the convergence of the network. This paper has explored three eminent protocols namely Routing Information Protocol (RIP), Enhanced Interior Gateway Routing Protocol (EIGRP) and Open Shortest Path First (OSPF) protocols by taking into consideration multiple factors as Convergence time, Jitter, End-to-End delay, Throughput and Packet Loss etc.**

Keywords—Internetworking, Routing,RIP,EIGRP, OSPF

I. INTRODUCTION

For Internet and similar network communications, TCP/IP provides end to end connectivity, including specifications on data formatting, addressing, transmission, routing and reception at the destination. *Fig. 1* shows TCP/IP protocol suite. As shown in *fig. 1*, it basically contains four generalization layers which includes link layer, internet layer, transport layer and application layer. There is separate role of each layer such as information related to local network communication technologies is provided by the link layer while Internet Layer provides connections between different local networks and establishes internetwork communication. Transport layer handles peer to peer communication while Application layer contains protocols specific for data communication services on process to process level [1].

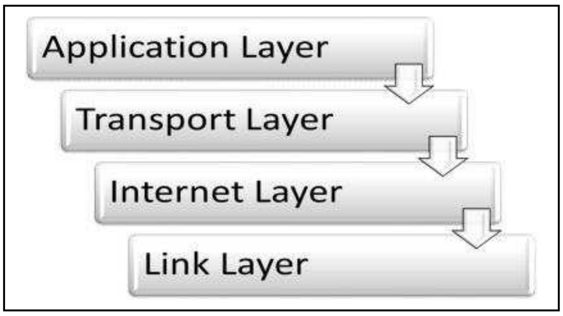


Fig. 1. TCP/IP protocol suite

Routing is the process of directing data packets through designated paths in a network. Shortest-path routing is a typical example where, as the name suggests, length is the main aspect. Each route is assigned a length in the network and those routes whose lengths are equal to the distance—the shortest paths—are selected for data packet transfer from source to the destination [2]. The protocols are rules that allows for communication between the different network devices. There are two main categories of protocols known as routed protocols and routing protocols. Routed protocols are known as those network protocols that provide sufficient information to utilize base as an addressing scheme for directing packets from one host to another such as protocol Internet (IP) is an example of it [3]. Routing protocols are the means of information transmission between routing nodes in the network; thereby play a dynamic role in present network communications. They play an important role in determining the optimal path according to the algorithm used for each node, for communication with one another [4]. *Fig. 2* shows three important routing protocols used in networks.

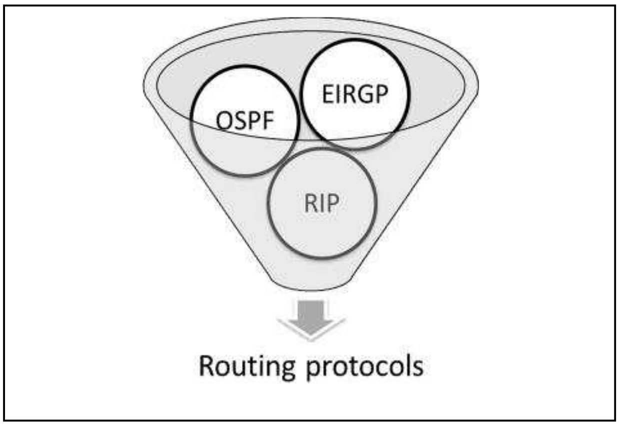


Fig. 2. Internetworking routing protocols

The paper explores three eminent routing protocols. Section II starts with a distance vector routing information protocol followed by section III which discusses about OSPF protocol. Section IV provides idea about EIGRP protocol while section V compares the three protocols by taking into account various recital parameters. Finally section VI terminates with brief conclusion.

II. ROUTING INFORMATION PROTOCOL

Routing Information Protocol (RIP) is a distance-vector routing protocol. Each router retains a table in distance vector routing algorithm that conserves the best known distance and the optimal route to be followed to reach at each destination [4]. It is the first routing protocol implemented on TCP/IP. It is an interior gateway protocol (IGP) used to exchange information and for routing traffic in the global Internet. As it is an Interior gateway protocol, which means that routing is performed within an autonomous system. An autonomous system is a collection of networks under a common administration. RIP evolved as an Internet routing protocol, and other protocol suites use modified versions of RIP [5-6]. Hop count is used to determine the best path in the RIP between locations. Basically a hop count is the number of routers through which the packet/datagram reaches the destination network. In RIP, the maximum number of hop counts is 15. If still the packet is unable to reach the destination then at 16th hop, the packet is discarded by the RIP to prevent from undesirable loops [6]. Therefore, the size of the network is limited in RIP. In a RIP network, each router broadcasts its entire Routing updates to the neighboring routers every 30 seconds. When a router receives a routing update that includes the changes, then it updates its own routing table and then sends the updated table to its neighbors. Matric value plays an important role in the RIP as the RIP routers maintain the best path in their routing tables(the path having the lowest matric value will be the best path)[1][7].

The database of the RIP router stores the information such as IP Address of the computer, Matric Value, Administrative Distance and Interface port for receiving the packet.There are two versions of this protocol. RIPv1 is a classful protocol, that broadcasts update every 30 seconds. Its hold-down period is 180 seconds. RIPv1 has a Hop count metric of maximum 15 hops. The second version is RIPv2 which uses multicasting. Also, it supports triggered updates i.e. when a change occurs, a RIPv2 router will immediately sends its routing table to its connected neighbors. RIPv2 is a classless protocol which supports variable-length subnet masking (VLSM) and authentication. The paths that have the smallest hop counts are selected by the RIP even when the path is slowest in the network. RIP performs well in small networks. However, it may be inefficient in larger networks [8].

III. OSPF ROUTING PROTOCOL

The drawbacks of RIP are overcome by the OSPF which stands for "OPEN SHORTEST PATH FIRST". OSPF is a Link state advertisement (LSA) type routing protocol designed by the Internet Engineering Task Force (IETF). OSPF is also a classless interior gateway protocol. It is a routing protocol that supports Variable Length Subnet Mask (VLSM). OSPF works under a large autonomous system [7]. The backbone area in OSPF is known as the default area, and is referred to as area0. A contiguous backbone area must be defined within each autonomous system. Route summarization is used by OSPF as the areas routing are carried out by the backbone router. OSPF also supports load balancing up to 16 equal paths. A router with OSPF process is defined by the Router ID[9]. A router Id is a 32 bit number that identifies the router uniquely. While configuring OSPF, it undergoes following processes:

A. Neighbours

Link state advertisements are sent by router to find its neighbors. Hello protocol is used to elect the neighbors. The Hello protocol sends the Hello Packets at every 10 seconds to each interface[6].After the hello packets sent, the area ID should be same to become the neighbors. After the neighbors established and an adjacency is formed, then the OSPF routers exchange the hello packets. If the hello packet is not received from the neighbor within 40seconds then the neighbor is considered as dead. This is known as the dead interval time[9].

B. Adjacency

After the hello packets are exchanged between the routers, an adjacency is created by proceeding from hello packets exchange process to database exchange process[7]. In order to minimise information exchanged on a segment such as the Ethernet, the router elects DR (Designated Router), a BDR(Backup designated router) as well as other OSPF routers called DROTHER. DR(Designated Router) is the router having the highest router Id, and the router with the second highest router id is the BDR(Backup designated router). All other routers are considered as DROTHERs [10]. The drawback of this process is that in a running network when a new router is intalled with the highest router id, the DR will not change. Therefore, loopbacks with the highest router ids are used for its elimination.OSPF stores only the best path in its routing table. This is the main drawback of OSPF as if the best path gets disturbed then it is time consuming to find the alternate best path as the OSPF protocol first calculates the matric of the alternate route.

IV. EIGRP ROUTING PROTOCOL

To overcome the drawbacks of the OSPF routing protocol, EIGRP is used. EIGRP is a cisco proprietary protocol that supports distance vector protocols and link state advertisements. It is an advanced distance vector routing protocol that uses the theory of autonomous systems to group routers performing the similar tasks [4]. Same autonomous system number should be used by the EIGRP router to share the routes. EIGRP uses 3 tables for the routing decisions. These are known as the routing table , the neighbor table and the topology table. EIGRP maintains the neighbor table. The neighbor table carry information regarding the directly connected routers. The topology table is also maintained by EIRGP that carry the information regarding the route enteries for each destination. EIGRP selects the best path from the topology table and enters that path in its routing table[5][10]. The second best path remains in the topology table. EIGRP works on the principle of bandwidth and delay as the metrics to find the best route. EIGRP only sends the triggered updates when some changes happened in the network. EIGRP update its routing table and exchange its updated routing table to the other router in the same autonomous system. Hence, saving the bandwidth and system time. EIGRP multicasts the updates to 224.0.0.10. EIGRP sends the hello packets at every 5 seconds on LAN and in all other cases the time becomes 60 seconds. If a router that sends the hello packets doesn't receive any reply in 15 seconds on LAN and 180 seconds in all other cases, then the router removes that entry from the routing table [9].

C. *EIGRP operation*

EIGRP does not depend on TCP/IP to exchange routing information. EIGRP is protocol independent uses its own proprietary transport layer protocol known as RTP(Reliable Transport Protocol).

B. *EIGRP Packet Types*

1) Hello – Used for detection of neighbor routers. Hello packets are sent at interval of every 5sec . Hello packets are sent multicast to IP address 224.0.0.10.

2) Acknowledgment – Packets are used to acknowledge for the query, update and reply and are reliable.

3) Update - Used on discovery of a new neighbor or a topology change by a router.

4) Query- Used when EIGRP routers lost the information and does not have backup paths, then the router send the qurery packets to one or all neighbors.

5) Reply – Reliable Packets that are used in response to the query [5].

V. Performance Comparison Between RIP, OSPF And EIGRP

There are various parameters such as hop count, convergence time, protocol type, algorithm used etc. that distinguish these protocols from one another. All these protocols have different characteristics and features. Table 1 summarizes different features for these protocols and provides a brief comparison between them.

VI. Conclusion

Different internetworking protocols have been discussed in this paper such as RIP, OSPF and EIGRP. It has been observed from the discussions that RIP protocol is sufficient only for small networks while for large networks, OSPF and EIGRP are better choice. OSPF is an open standard protocol that also provides security and multipath facilities. EIGRP is EIGRP is robust, efficient and fast hybrid protocol that is having added advantage of least convergence time as compared to other two whereas convergence time is more for RIP. Thus, making EIGRP suitable for real time applications.

Table I. Performance Comparison Between RIP, SPF And EIGRP Protocol

Sr.No	Features	RIPv1	RIPv2	OSPF	EIGRP
1	Cisco Proprietary	No	No	No	Yes
2	VLSM Support	No	Yes	Yes	Yes
3	Auto Summarization	Yes	Yes	No	Yes
4	Path Matric	Hops count	Hops count	Bandwidth	Bandwidth, Reliability, Delay and Load
5	Hop count Limit	15	15	No Limit	255
6	Convergence	Slow	Slow	Fast	Very Fast
7	Algorithm used	Bell-man Ford	Bell-man Ford	Dijkstra's algorithm	Diffusing Update Algorithm
8	Sending Updates	Broadcast	Multicast	Multicast	Multicast
9	Classfull or Classless	Classfull	Classless	Classless	Classless
10	Protocol Type	Distance vector	Distance vector	Link State	Hybrid
11	Table	Routing Table	Routing Table	Routing, Neighbor and Topology Tables	Routing, Neighbor and Topology Tables
12	Discontigeous Networks	No	Yes	Yes	Yes

REFERENCES

[1] Y. Joo,V. Ribeirot, A. Feldmannt, A.C. Gilbert, and W.W.G. Yang, "TCP/IP traffic dynamics and network performance: A lesson in workload modeling, flow control,and trace-driven simulations," Computer Communication Review,ACMSIGCOM, April 2001.

[2] F. Le, and J.L. Sobrinho, “Interconnecting Routing Instances,” IEEE/ACM Transactions on Networking, vol. 22, Issue:2, pp. 540-553, April, 2014.

[3] M.H. Perez, and A.M. Santiago, “Process of Migration of Routing Protocols in a LAN: RIP to OSPF, ” IEEE International Conference on Electrical and Electronics Engineering, Veracruz, pp. 1-5, Sept. 2006.

[4] I. Fitigau, and G. Toderean, “Network performance evaluation for RIP, OSPF and EIGRP routing protocols,” IEEE International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Petesti, pp. 1-4, June 2013.

[5] M.N. Islam, and M.A.U. Ashiqu, “Simulation Based EIGRP over OSPF Performance Analysis,” Master Thesis in Electrical Engineering Emphasis on Telecommunications, Blekinge Institute of Technology, Thesis no: 4983, May 14, 2010.

[6] R.M. Pethe, S.R. Burnase, “Technical Era Language Of The Networking – EIGRP,” International Journal of Engineering Science and Technology (IJEST) NCICT Special Issue, pp. 1-5, Feb 2011.

[7] P. Rakheja, P. Kaur, A. Gupta, and A. Sharma, “Performance Analysis of RIP, OSPF, IGRP and EIGRP Routing Protocols in a Network,” International Journal of Computer Applications, vol. 48, no. 18, June, 2012.

[8] Y.NavaneethKrishnan, C.N. Bhagwat and A. Utpat, “Performance Analysis of OSPF and EIGRP Routing Protocols for Greener Internetworking,” IJCA Proceedings on International Conference in Distributed Computing & Internet Technology, ICDCIT:1-5, Jan, 2013.

[9] Cisco press, Interconnecting Cisco Internetworking Devices, Part 2, ICND2, Study Guide, vol 1, Version 1.0, 2007.

[10] T. Lammle, Cisco Certified Network Associate, Study Guide, 6th Edition, Wiley India Pvt. Ltd., 2007.

[11] Cisco press, Interconnecting Cisco Internetworking Devices, Part 2, ICND2, Study Guide, vol 2, version 1.0, 2007.

Localization Techniques in Wireless Sensor Networks

Karanjit Kaur
Dept. of Electronics & Communication,
CT Institute of Technology & Research,
Maqsudan, Jalandhar
ghotrakaranjit@gmail.com

Amanjot Singh
Dept. of Electronics & Communication,
CT Institute of Technology & Research,
Maqsudan, Jalandhar

Harjit Pal Singh
Dept. of Electronics & Communication,
CT Institute of Engineering, Management & Technology,
Shahpur,Jalandhar

Monica Sharma
Dept. of I.T,
CT Institute of Technology,
Shahpur, Jalandhar

***Abstract*— To determine the location of sensor nodes is the one of the key issue in wireless sensor networks. Location information is crucial for both network organization and for sensor data integrity. Sensor nodes are required to know their locations with high degree of precision in many wireless sensor applications. Localization method helps in assisting sensor nodes to find out their location in sensor network . This paper reviews the localization techniques used to locate the nodes which are in the communication range of each other to validate the performance of localization algorithms.**

***Keywords – Localization algorithms, Wireless Sensor network, Anchor*.**

I. INTRODUCTION

Wireless sensor network (WSN) is an extensive research area offering wide range of applications in the numerous fields. Sensor network supports a large number of sensing nodes that are distributed over a geographical area so as to instrument and observe the events in that particular environment[3]. The collected data are not usable without knowing the location of an event that is the location of sensor. The positions are estimated and computed by using less expensive and accurate mechanism. Using GPS in large scale is not cost effective and not feasible[2] in dense foliage and indoors because it requires line of sight between receiver and satellites. Due to poor signal reception with low accuracy, it becomes inefficient. Therefore, several algorithms proposed to solve the localization problem. This literature categorizes the localization techniques with new viewpoints.

A. *Localization Techniques*

One of the more important task is to measure the accurate location information that greatly improve the performance of tasks such as routing, energy conservation, data aggregation and maintaining network security[4][6]. Localization in wireless sensor networks is performed by using the following steps:

- Distance estimation- It involves measurement techniques[2][3] to evaluate the relative distance between nodes.
- Position computation- It comprises algorithms[4] in which the coordinates of the unknown node are to be calculated with respect to the location of neighbouring nodes or anchor nodes.For location estimation depending on the reference nodes' location ,localization algorithms require techniques implementation. These techniques are known as multi-lateration (ML) techniques[10]. Some simple ML techniques are[4][8]:

1) *Atomic Multilateration* : The position of node is to be determined by a mechanism ,if a node receives three reference points or beacons. Fig.1 shows three beacon nodes and one unknown node.

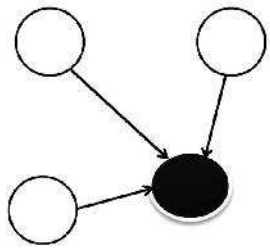


Fig.1 Atomic multilateration[10]

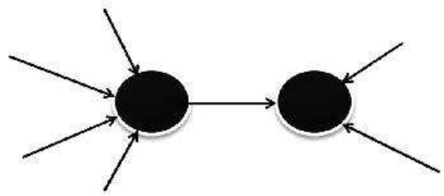


Fig. 2 Iterative multilateration[10]

2) *Iterative Multilateration* : Sometimes, a node estimate its location and sends out a beacon ,it enables the other nodes to receive atleast a three beacons. It occurs only when a node not in the direct range of three beacons in the network. Fig.2 shows two unknown nodes receive beacon nodes.

3) *Collaborative Multilateration*: The two or more nodes are to be collaborated with each other, when they not able to receive atleast three beacons each. There are six participating nodes, out of them four are beacons, whose positions are known and two are unknown nodes.

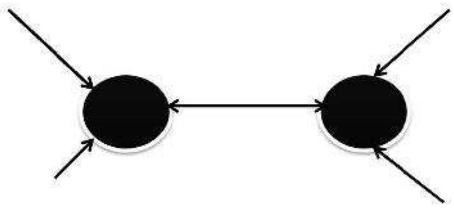


Fig. 3 Collaborative multi-lateration

Fig.3 shows the two unknown nodes receive beacons nodes. Various localization algorithms can be classified as follows[2][3][4]:

- Localization algorithms[3][4][8]- the information concerning distances and positions is manipulated by using algorithms in order to allow all nodes within network to estimate their position.

Relative localization algorithms[2][9] evaluates relative position of the nodes or the coordinate system is chosen for the group of nodes . The anchor nodes are not required and in location aided routing applications, relative positions are just sufficient than calculating the absolute positions. Absolute positions are manipulated by implementing absolute localization algorithms making use of anchor nodes which transmit their location information to unknown nodes. The geographical locations of anchor nodes are known by the use of GPS or through manual installation prior to the localization process .The number of anchor nodes used greatly determine the accuracy of algorithm

II. RANGE BASED VERSUS RANGE FREE

Once distance information to at least three anchors is available through any of these mechanisms, a simple trilateration or multilateration algorithm yields the location of the sensor nodes.

In range–based algorithms[2][4] the distance information[2][6] is obtained from:

A. Timing information, or the signal propagation time or time-of-flight (TOF) of the communication signal is used to measure distance between the receiver and the reference point.

B. Time difference of arrival (TDOA) is used to calculate the distance between two nodes.

C. Received signal strength information (RSSI) infers the distance between the receiver and the reference point from the fact that attenuation of the radio signal increases as the distance between the receiver and transmitter increases. Using these measurements in methods like triangulation or trilateration[4][6] a node location is uniquely specified when atleast the coordinates of 3 reference points or anchor nodes are available for a node. Range-free methods do not require additional hardware, but they generally only work well when networks are dense. Sparse networks by nature contain less connectivity information and are thus more difficult to localize accurately. These algorithms require that each node knows which nodes interact with each other i.e. in the communication range of each other, their location estimates and ideal radio range of sensors.

Range free techniques are most cost-effective[4] because they do not require sensors to be equipped with any special hardware but use less information than range based.

B. Localization Algorithms

A localization algorithm estimates the location of nodes in the network area based on the input data.

i. Non-learning based algorithms

The non-learning based localization algorithms classified into anchor or reference based and anchor free classes.

1) Non-learning anchor based localization algorithms

The nodes which are aware of their positions because of adding GPS or manual installation are used in some algorithms to estimate the location of other nodes which don't know their positions. By using anchors, better accuracy can be achieved. The accuracy of location depends on the number of anchors and their distribution in the network. These anchor based algorithms can be implemented in the fixed, mobile or hybrid networks. Fixed networks consist of static nodes and they are used to localize unknown nodes. Hybrid networks contain of static sensor nodes and mobile beacon acts as a static one (broadcasting its accurate position) and represents many virtual static beacons. The goal of using these schemes is to localize static nodes. Mobility makes WSNs more flexible and enables more possible applications. The first issue in mobile wireless sensor networks (MWSN)s is the latency. And line of sight (LOS) requirement is an another challenge available in localization techniques. A sensor moves from a position with good LOS to a position with bad LOS is possible. So, LOS availability for mobile nodes is required. Localization algorithms can be implemented in centralized, locally centralized and distributed manner.

- Centralized approach: The information about the entire network is transferred to a central unit for analysis, and then computed positions are sent back into the network. It is more accessible for small scale area networks. It is easy to implement and more accurate than other implementations because the existence of global information.

- Locally centralized approach: The two or three central units are available in this technique. It is proposed as a solution to scalability problem in centralized approaches.

- Distributed approach: The relevant computation is done on the sensor nodes themselves in these algorithms. It is more flexible for large scale networks. In the non-learning category of algorithms , Euclidean properties are used to localize the sensor nodes. The most important are Triangulation, Trilateration, Multilateration and Proximity based. For fixed networks, convex position algorithm [16] is implemented which is a centralized localization algorithm. By a single centralized node this algorithm is executed. So, it is not possible for ad hoc applications. CBLALS method [20] is a locally centralized localization algorithm for fixed networks. It uses TDOA measurement technique along with ultrasound and RF signals to locate indoor sensors.

The first step in DV-hop algorithm [18] is that all the anchor nodes flood their location to the network through message. The message is propagated hop by hop and counts the hop count from anchor to node heard. Each node with counting table maintains the minimum number of hops that is away anchors. The average hop distance is calculated by anchor nodes and transfer back to network as correction factor. When a non-anchor node gets correction factor from nearest anchor, uses it to estimate its distance to anchors. Then, unknown node applies trilateration to estimate its location.

The proposed algorithm given in [19] is called distributed grid-based transmitting power (DGL). The communication range of anchor nodes can be changed by increasing their transmitting power. A rectangular coordinate system is established by each node and divides it into square grids.

2) Non-learning anchor free localization algorithms

This category has not any anchor node as compare to anchor based. In these schemes, the algorithm is used to find the relative positions of the nodes instead of nodes' positions in the coordinate system through a reference nodes (anchors).These can be used in fixed, hybrid and mobile networks . The localization techniques are to be implemented based on Centralized, Locally Centralized and Distributed approaches. The methods which are given in [22-25] are anchor free and non-learning localization algorithms. Basically, the data analysis and information visualization is determined by using the multi-dimensional (MDS-MAP)so as to represent the data in geometrical form. The shortest distance between all pairs of nodes is computed and then a distance matrix is obtained and applies MDS to construct relative location of nodes. The absolute nodes' location can be estimated by transforming relative locations only if there was sufficient anchors available.MDS has high computation cost and requires global information. IMDS approach [23] is based upon MDS-Map method which is locally centralized algorithm for large scale networks.

III. CONCLUSION

This paper, a new classification for localization techniques is proposed. Although localization algorithms were classified based on different key features like anchor existence, movement in network, etc. In order to understand the operation of varies localization methods this classification is usable and it also helps to implement a new localization algorithm.

REFERENCES

[1]. Kiran Yedavalli and Bhaskar Krishnamachari, "Sequence based localization in wireless sensor networks", IEEE Transactions On Mobile Computing, , ISSN 1536-1233,Vol. 7, No. 1, pp. 81-94, Jan 2008.

[2]. Xia Zhenjie and Cheni Changjia, " A Localization Scheme with Mobile Beacon for Wireless Sensor Networks", 6th International Conference on ITS Telecommunications Proceedings, pp1017-1020,2006.

[3]. Yi Shang, Ying Jhang, Markus Fromherz, "Localization from connectivity in wireless sensor networks", IEEE Transactions on Parallel and Distributed Systems,Vol. 15, No. 11,pp. 961-974, November 2004.

[4]. Masoomeh Rudafshani and Suprakash Datta, "Localization in wireless sensor networks", ACM, IPSN'07,April 25- 27, 2007.

[5]. Tao Chen, Zheng Yang, Yunhao Liu, Deke Guo, Xueshan Luo, "Localization Oriented Network Adjustment In Wireless Adhoc And Sensor Networks", IEEE transactions on Parallel and Distributed Systems, ISSN. 1045-9219,Jan 2013.

[6]. Rui Huang, Gergely V. Zaruba, "Monte Carlo Localization Of Wireless Sensor Networks With A Single Mobile Beacon", Springer, LLC 2008.

[7]. Neal Patwari, Joshua N. Ash, Alfred O. Hero III,Neiyer S. Correal, "Cooperative localization in wireless sensor networks", IEEE signal Processing Magazine, July 2005.

[8]. Zhen Hu, Dongbing Gu, Zhengxun Song, Hongzuo Li, "Localization in wireless sensor networks using a mobile anchor node", Proceedings of 2008 IEEE /ASME, International Conference on Advanced Intelligent Mechatronics, July 2-5 ,2008.

[9]. M. Castillo-Effen, W.A. Moreno, M. A. Labrador, K. P. Valavanis , "Adapting sequential Monte-Carlo Estimation to Cooperative Localization in wireless sensor networks", IEEE International conference on Mobile Adhoc and Sensor Systems, pp. 656-661, Oct 2006.

[10]. Kazem Sohraby, Daniel Minoli, Taieb Znati , "Wireless Sensor networks", Technologu,Protocols and applications,Wiley publications, 2007.

[11]. Stoleru, R. and T. He, J.A. Stankovic, and D. Luebke,. "A high-accuracy, low-cost localization system for wireless sensor networks", Proceedings of the 3rd international conference ,pp 13–26, 2005.

[12]. Tang, T. and Q. Guo, and B. Peng, 2008. "Sorted TDOA optimization based localization algorithm for wireless sensor network" Jisuanji Gongcheng yu Yingyong(Computer Engineering and Applications), (44) 25.

[13]. Stefano, G.D. and A. Petricola, "A distributed AOA based localization algorithm for wireless sensor networks" Journal of Computers, (3): 1. 2008.

[14]. Zhuang, W. and G. Song, J. Tan, and A. Song, "Localization for hybrid sensor networks in unknown environments using received signal strength indicator" Information and Automation, 2008. ICIA 2008.International Conference on: 567–572, 2008.

[15]. Bulusu, N. and J. Heidemann, and D. Estrin. "GPS-less low-cost outdoor localization for very small devices" IEEE personal communications (7), pp. 28–34, 2000.

[16]. Doherty, L. and L. El Ghaoui, "Convex position estimation in wireless sensor networks" INFOCOM, 2002.

[17]. Savarese, C. and J.M. Rabaey, and J. Beutel, 2002. "Location in distributed ad-hoc wireless sensor networks: Acoustics, Speech, and Signal Processing, 2001". Proceedings (ICASSP'01) 2001 IEEE International Conference on: 2037–2040.

[18]. Srinivasan, A. and J. Wu, 2007. "A survey on secure localization in wireless sensor networks", Encyclopedia of Wireless and Mobile communications.

[19]. Sun, C. and J.Xing, Y. Ren, Y. liu, J.Sha, J.Sun, 2012."Distributed Grid-based Localization Algorithm for Mobile Wireless Sensor Networks",Advances in Electronic Engineering, Communication and Management (1) :315–321.

[20]. Shu, J. and R. Zhang, L. Liu, Z. Wu, and Z. Zhou, 2009."Cluster-based Three-dimensional Localization Algorithm for Large Scale Wireless Sensor Networks": Journal of Computers, (4),Pp.585.

[21]. Xing, J. and D. Wang, Y. Liu, 2011. "Distributed Range-free Localization Algorithm for 3D Wireless Sensor Networks under Irregular Radio Propagation Model": Applied Informatics and Communication, (224), pp. 299–306.

[22]. Pan, J.J. and J.T. Kwok, Q. Yang, and Y. Chen, 2006. "Multidimensional vector regression for accurate and low-cost location estimation in pervasive computing", IEEE Transactions on Knowledge and Data Engineering (18), pp. 1181–1193.

[23]. Shang, Y. and W. Ruml, 2004. "Improved MDS-based localization": INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies (4) ,pp. 2640–2651.

[24]. Hari, N.P. and H. Balakrishnan, E. Demaine, and S. Teller, 2003. "Anchor-Free Distributed Localization in Sensor Networks", Proceedings of the 1st international conference on Embedded networked sensor systems, pp.340–341.

[25]. Youssef, A. and A. Agrawala, and M. Younis, 2005. "Accurate anchor-free node localization in wireless sensor networks", Performance, Computing, and Communications Conference, 2005. IPCCC 2005. 24th IEEE International, pp.465–470.

Scruitinize the Movement of Sensor Nodes in Zigbee Network Topologies

Sandeep Kaur
AIET
Faridkot, Punjab
sandeepsandhu790@gmail.com

Harinderpal Singh
AIET
Faridkot, Punjab
harinder.ycoe@gmail.com

Gurjeevan Singh
DIC ECE, Poly-Wing, SBSSTC
Ferozepur, Punjab
gurjeevansandhu@gmail.com

Abstract—**This research work is implemented on 802.15.4/ZigBee sensor network. Zigbee used for low rate low power and low cost wireless communication. This paper based upon the mobility management in ZigBee sensor network. The simulation is done using OPNET modeler 14.5. The performance of ZigBee routing scheme is examine using parameters packet delivery ratio, data traffic received, data traffic send, and network load.The results show that tree topology have better performance in data traffic received, data traffic send and network load than other topologies. The packet delivery ratio is effective in case of mesh topology. The results show that tree topology is best in mobility management scenario.**

***Keywords*—Mobility, ZigBee, IEEE802.15.4, OPNET, PDR.**

I. INTRODUCTION

Wireless sensor networks (WSNs) is described as the collection of sensors organised in a given area, that can sense and control the physical or environmental conditions such as temperature , pressure , volume etc. and pass their data to main location through the network [2].ZigBee is based on IEEE 802.15.4 wireless protocol [1]. First time ZigBee network is defined by Alliance and released in 2006. The second stack of the ZigBee network was defined as ZigBee 2006 [8].ZigBee operates in the industrial, scientific and medical (ISM) radio band, at 2.4 GHz. ZigBee also utilizes 868MHz band in Europe and 915 MHz band in USA, Australia [9, 10].ZigBee is intended to be used in those applications that require only a low data rate, long battery life, secure networking, smaller size of protocol stacks and larger addressing space [1]. IEEE 802.15.4 devices typically operate in a limited personal operating space. The IEEE 802.15.4 is concerned with PHY and MAC layers only. The upper layers are defined by the ZigBee Alliance. The combination of IEEE 802.15.4 and ZigBee specifications allows defining a WSN protocol stack able to support a variety of WSN applications [3].In this paper we examine the Performance of ZigBee network topologies using random waypoint mobility.

II. ZIGBEE NETWORK

ZigBee builds upon the physical and media access control layers, which are define in IEEE standard 802.15.4 for low-rate WPANs. The application and network layers are defined by ZigBee specification.

A. ZigBee Devices

Two types of devices are specified in the IEEE802.15.4 framework according to their capabilities and available resources i.e. Full-Function Devices (FFD) and Reduced-Function Devices (RFD) [3].

1) *ZigBee coordinator (ZC):* The coordinator is a fully functional device and act as the parent of other nodes. It is root of the network or may act as a bridge to other networks. To initiate a network it is essential to place at least one ZigBee coordinator in the network. This will work as trust center, which secure keys and store information [1]. ZigBee coordinator is responsible for starting the network and choosing the key network parameters [5].The coordinator allows the association of other nodes to the network [3].
2) *ZigBee Router (ZR)*: ZR act both as RFD and FFD. In RFD, it act as child of ZigBee Coordinator (parent) and perform the function of intermediate router passing data from ZC to ZED. ZigBee Router as FFD is a parent to its child node (ZED) [1]).Routers that acts as FFD are also known as Coordinators, since they coordinate all devices that transfer data to coordinator via the FFDs themselves [3]. ZigBee routers are devices capable of routing data .In the case of many-to-one multi-hop ZigBee-based WSN(ordinary) sensor nodes are likely to assume the role of network routers[5].
3) *ZigBee End Device (ZED):* ZigBee end-devices have no routing capability – these devices rely on their parents, the coordinator or routers, to transmit/route their packets [5].It is a reduced functional device, work as a child node and receives data from its parent nodes. So, it requires lower memory usage, which makes it cheaper than ZC or ZR [1].

B. ZigBee Topologies

“Topology” refers to how the data is transmitted through the hardware components. They describe the physical and logical arrangement of the network nodes[5]. There are three network topologies i.e. Star topology, Mesh topology, Tree topology in the ZigBee network, we explain all these topology discussedbelow.

1) *Star Topology*: The star topology consists of a coordinator and several end devices. In this

topology, the end device communicates only with the coordinator. Any packet exchange between end devices must pass along the coordinator. The main advantages of star topology are its simplicity, energy savingand predictable. The drawbacks are limited scalability [7].

2) *Mesh topology*: A mesh topology provides greater flexibility than other topologies. It covers multiple paths for messages within the network. If a particular router fails, then ZigBee self-healing mechanism will recover the network. Mesh topology is highly reliable and robust. The limitation of this topology is lower end-to end performance. Figure 4 show the network using mesh topology [7].

3) *Tree Topology*: A Tree topology consists of a Coordinator and this coordinator linked with its children which are basically the set of routers and end devices. A Router also connected with its children like another routers and end devices [7].

III. SIMULATION ENVIRONMENTS

The simulation of ZigBee network has been done by using OPNET modeler 14.5.The network area selected on the level of official network scale (45m*45m).The number of nodes used in the network are 36.Three scenarios are used, in which percentage of mobile nodes varies from 20 to 80 percent and using random waypoint mobility model. The speed of mobile nodes is 5 m/s. The nodes are randomly placed anywhere in the network. The simulation duration of the network is 300 sec .The maximum number of routers are 10 and maximum number of children are 10 whereas maximum depth of network is 5.In the first scenario tree topology is used and in second scenario mesh topology is used. In third scenario the combination of tree and mesh topologies is used to implement a hybrid topology.

IV. RESULTS AND DISCUSSIONS

Using the parameters PDR, DTR, DTS and network load following results have been observed.

A. *Packet delivery ratio (PDR) with respect to percentage of mobile nodes:*The results inFigure 4 are calculated by varying number of mobile nodes from 20% to 80% in the mesh, tree and hybrid topologies.Figure 1 shows that that mesh topology provides maximum PDR and tree topology provides minimum PDR. Whereas hybrid topology provides efficient PDR.

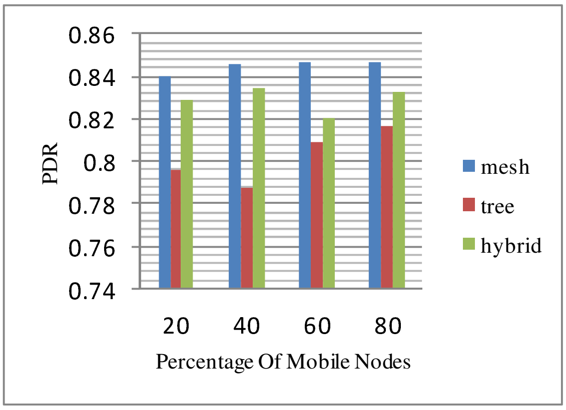


Fig.1. PDR with varying mobile percentage of node

B. *Data Traffic Received (DTR):* It represents the total traffic successfully received by the MAC from the physical layer in bits/sec. This includes retransmissions. The result of figures 2, 3, 4 and 5shows that tree topology provides maximum DTR whereas mesh topology in most of cases provides minimum DTR. The hybrid topology provides efficient DTR.

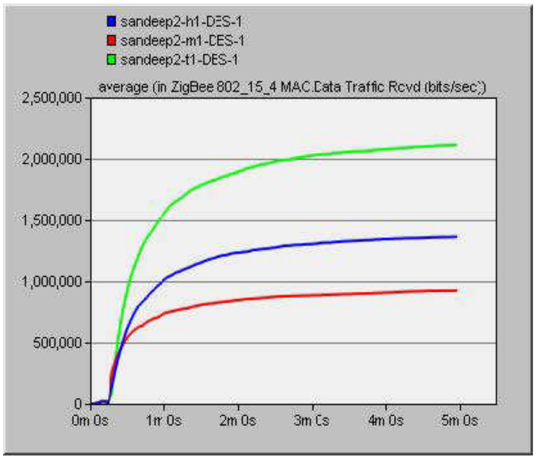


Fig.2.DTR of tree, mesh and hybrid topologies with 20% of mobile nodes

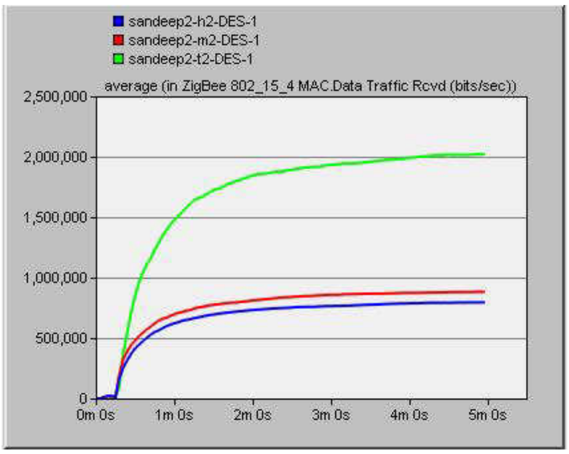


Fig.3.DTR of tree, mesh and hybrid topologies with 40% of mobile nodes

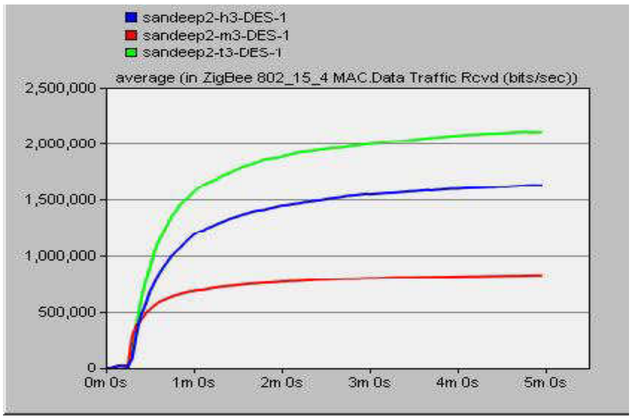


Fig.4. DTR of tree, mesh and hybrid topologies with 60% of mobile nodes

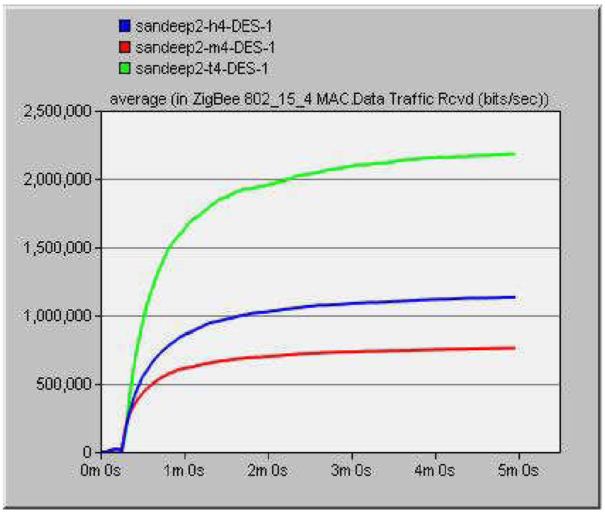


Fig.5.DTR of tree, mesh and hybrid topologies with 80% of mobile nodes

C. *Data Traffic Send:*These statistics record the amount of data transmitted by the network interface onto the physical layer. The result in figures 6, 7,8and 9shows that tree topology provides maximum DTS and mesh topology in most of cases provides minimum DTR. The hybrid topology provides efficient DTS.

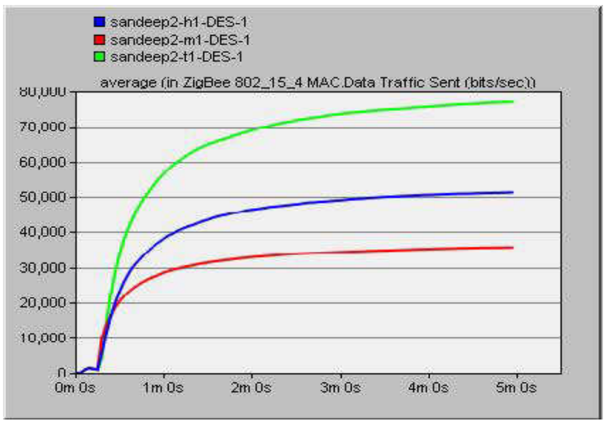


Fig.6.DTS of tree, mesh and hybrid topologies with 20% of mobile nodes

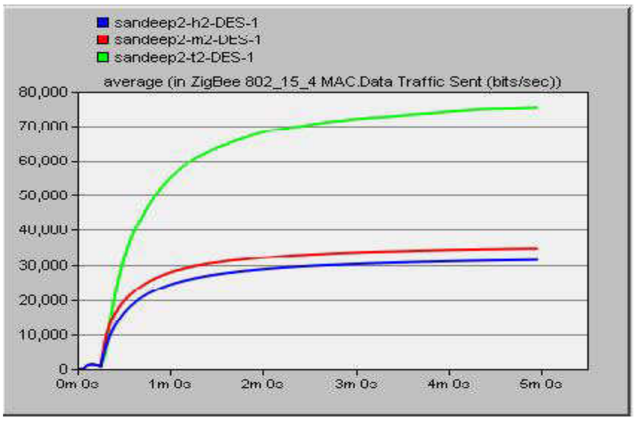


Fig.7.DTS of tree, mesh and hybrid topologies with 40% of mobile nodes

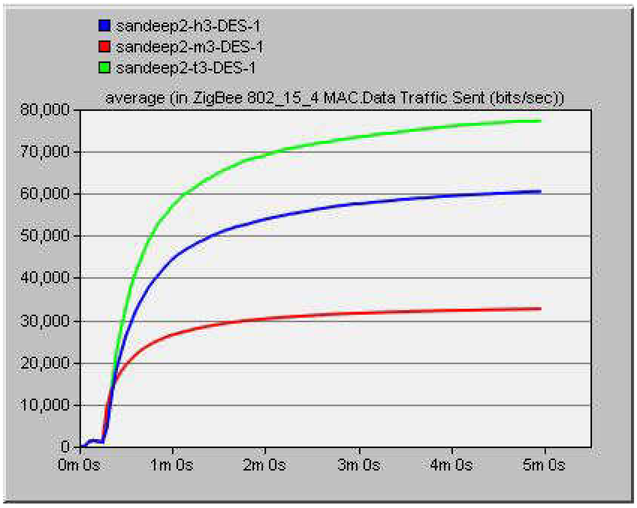


Fig.8.DTS of tree, mesh and hybrid topologies with 60% of mobile nodes

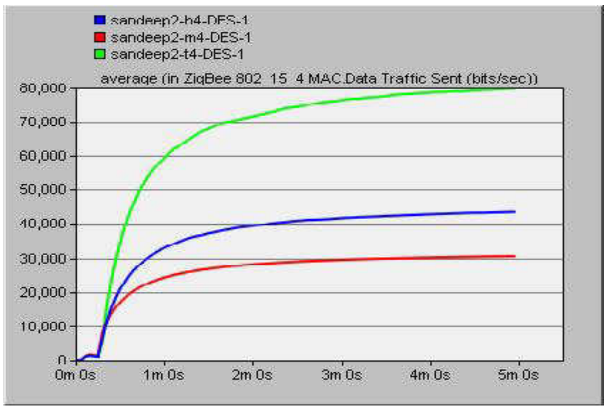


Fig.9.DTS of tree, mesh and hybrid topologies with 80% of mobile nodes

D. Network Load: It is defined as total load (in bits/sec) submitted to 802.15.4 MAC by all higher layers in all WPAN nodes of the network. The figures 10,11, 12 and 13 show that tree topology has the maximum network load and in most of cases mesh topology has minimum network load , whereas hybrid topology have efficient network load.

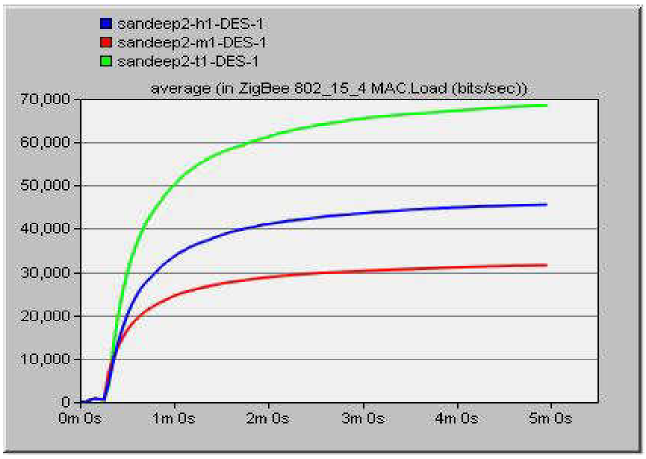


Fig.10.Network Load of tree, mesh and hybrid topologies with 20% of mobile nodes

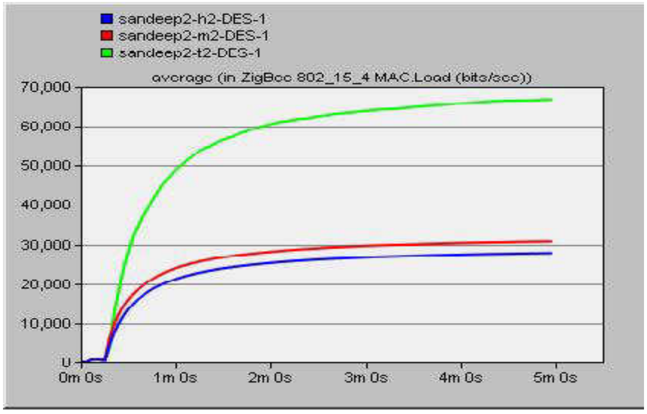


Fig .11Network Load of tree, mesh and hybrid topologies with 40% of mobile nodes

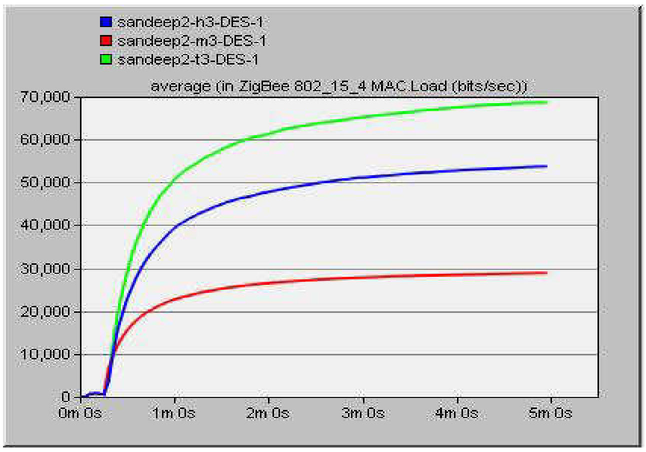


Fig.12.Network Load of tree, mesh and hybrid topologies with 60% of mobile nodes

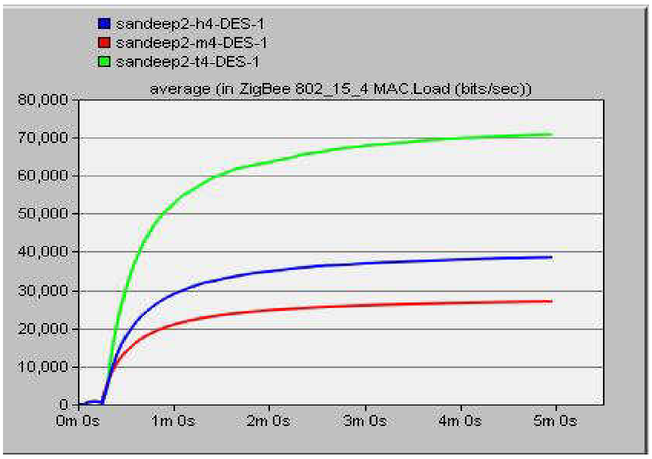


Fig. 13. Network Load of tree, mesh and hybrid topologies with 80% of mobile nodes

CONCLUSION

The results are examined using the parameters PDR, DTR, DTS and network load. The simulation results show that minimum PDR is attained by tree topology whereas maximum PDR is attained by mesh topology In case of hybrid topology efficient PDR are obtained. In case of DTR and DTS maximum DTR and DTS are achieved by tree topology and minimum DTR and DTS by mesh topology. Whereas hybrid topology provides efficient DTR and DTS in both cases .In case of network load maximum network load is provided by tree and minimum by mesh topology whereas hybrid topology offers efficient network load .For future this work may be extended to more hybrid topologies like star mesh and tree.

REFERENCES

[1] http://en.wikipedia.org/wiki/ZigBee.

[2] http://en.wikipedia.org/wiki/Wireless_sensor_network

[3] Francesca Cuomo, Emanuele Cipollone, Anna Abbagnale," Performance analysis of IEEE 802.15.4 wireless sensor networks: An insight into the topology formation process", Computer Networks 53 (2009) 3057–3075.

[4] Karandeep Malhi, Subhas Chandra Mukhopadhyay, Mathias Haefke, and Hartmut Ewald, "A Zigbee-Based Wearable Physiological Parameters Monitoring System", IEEE sensor journal, vol. 12, NO. 3 march 2012.

[5] Lovish Jaiswal, Jaswinder Kaur, Gurjeevan Singh, "Performance Analysis of TopologicalVariation in Personal Area Network using ZigBee Wireless Sensors", IJCST Vol.3, Issue 4, Oct- Dec 2012, ISSN: 2229- 4333(Print).

[6] Karunakar Pothuganti1 and Anusha Chitneni, A Comparative Study of Wireless Protocols: Bluetooth, UWB, ZigBee, and Wi-Fi, ISSN 2231-1297, Volume 4, Number 6 (2014), pp. 655-662.

[7] Ling-Jyh Chen, Tony Sun, Nia-Chiang Liang," An Evaluation Study of MobilitySupportin ZigBee Networks", J Sign Process Syst (2010) 59:111–122.

[8] Zavosh Abdollahzadeh Davani, Azizah Abdul Manaf," A Survey on Key Management of ZigBee Network",The international conference on E-Technologies and business on the Web (EBW2013) SDIWC.

[9] J. Pedro Amaro, Fernando J.T.E.Ferreira, RuiCortesão, Jorge Landeck, “Powering Wireless Sensor Networks Nodes for Complex Protocols on Harvested Energy”, CENTERIS 2012 – Conference on enterprise Information Systems HCIST 2012 – International Conference on Health and Social Care Information Systems and Technologies.

[10] Yuan-Yao Shih, Wei-Ho Chung, Pi-Cheng Hsiu, Ai-Chun Pang, “A Mobility-Aware Node Deployment and Tree Construction Framework for ZigBee Wireless Networks”, IEEE transaction on vehicular technology, vol. 62, no. 6, July 2013.

Effective Caching Techniques in Wireless Sensor Networks

Taranvir Kaur
IT Department,CTIEMT,
Shahpur, Jalandhar
taranvir.kaur15@gmail.com

Sajpreet Kaur
CSE Department, RBIENT,
Hoshiarpur
sajpreet.dhillon@gmail.com

***Abstract*—With the advancement in the microelectronics, Wireless Sensor Networks (WSN) has become a critical area for many applications. The small sensor nodes can be deployed to monitor the human inaccessible areas to check for any problems and helps to take appropriate actions. Energy efficiency is the critical issue in these networks. Most of the energy is spent on communicating the data between the nodes. Better routing protocols and other aggregation techniques have been worked upon to handle this issue. Apart from this, Caching of the sensed data at the intermediate nodes, from where it can be retrieved later on, also offers a good approach to extend the lifetime of the sensor network and provide better energy efficiency. Caching offers benefits of low data latency, increase in data availability and decrease in network traffic. This paper discusses the benefits of caching and various caching techniques.**

***Keywords*—WSN, cache, source, sink, nodes.**

I. INTRODUCTION

With the advent of MEMS technology, it has become possible to design small size sensor nodes that can be easily deployed under different conditions to monitor the environment [1]. They have found applications in various fields like patient monitoring[3], surveillance[4-5], greenhouse surveillance[6], industrial monitoring [7], traffic monitoring [8], habitat monitoring [9], cropping monitoring [10]. The sensor node consists of different sensors, memory, processor, battery, actuators and has communicating ability. The sensor nodes deployed in an area to sense and report the changes collectively forms a Wireless Sensor Networks (WSN). The sensor nodes run out of energy very quickly due to limited battery life and hence affects the efficiency of WSN. Much of the work is done in handling the energy efficiency of the WSN. Many routing protocols and aggregation techniques have been designed to handle the energy constraint. Caching can also be used as one of the approach that can yield fruitful results to achieve the energy efficiency in WSN.

II. CACHING IN WIRELESS SENSOR NETWORK

Caching aims at storing the sensed data in the intermediate nodes between the sink and source so that every time instead of asking from the source, the data can be retrieved from the nodes caching the data, this reduces the communication overhead and hence saves energy.

WSN consists of sensor nodes randomly deployed in an application area. Any event change sensed is communicated to the base station with the multihop communication between nodes. From the sink the data is sent to the user where the processing of data is done based on the application to take relevant actions. The sensed data needs to be delivered within the time, so that any real time action can be taken with the help of sensed data.

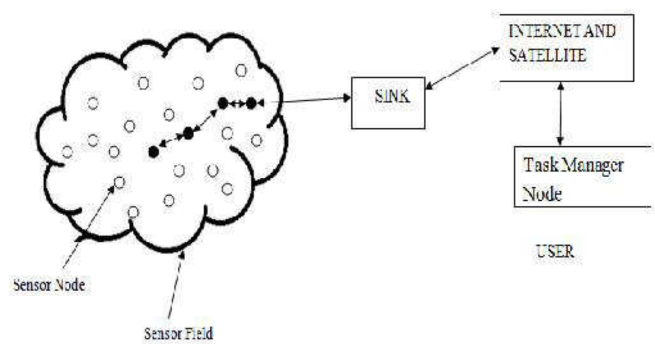


Fig 1. Architecture of sensor network [2]

As depicted in Fig 1, whenever sink injects the query into the network, the data is retrieved from the data source. The data is retrieved with the help of multihop communication between the sensor nodes along the path from the data source to sink. This communication involves lot of energy consumption which can be reduced with the help of caching.

Apart from energy efficiency, caching offers many other benefits. The caching of data increases the data availability and reduces the network traffic as data can be retrieved from the intermediate nodes instead of source. In case of failure of some nodes, the nearby nodes can cache the data of the failed nodes till they are replaced or repaired. Caching increases the Quality of Service of an application by making the data available from the cached nodes instead of source.

Various caching techniques have been used in WSN so far to utilize the benefits of caching in one or the other way. In the following section, the details of these caching techniques have been discussed.

III. VARIOUS CACHING TECHNIQUES

D.Nikos et al. proposed a cooperative caching scheme for Wireless Multimedia Sensor Networks (WMSN's) which deals with the multimedia content [11]. Apart from energy conservation, Quality of Service is one of the most indispensable requirements of such networks. WMSNs, which are mainly static and not mobile, there is a proposal for a cooperative caching policy which is based on the idea of exploiting the sensor network topology, so as to discover which nodes are more important than the others, in terms of their position in the network and/or in terms of residual energy. These so called "*mediator*" nodes are selected dynamically, so as to avoid the creation of hot-spots in the communication and the depletion of their energy. The mediators are not more powerful than the rest of the nodes, but they have some special role in implementing the cooperation among the sensors.

The parameter Node Importance Index (NI index) defines the importance of each node in a network by calculating the number of shortest path passing through that node. The higher the NI Index, higher the chances of a node becoming a mediator node. Each node calculates NI index of its 1-hop neighbor nodes and accordingly select few out of them as mediator nodes. The node notifies which of their neighbors are mediator nodes. When a node acquires the multimedia data that it has requested for, it caches it and broadcast a small index packet containing dataID, Time To Live (TTL) associated with item, its own remaining cache space and energy. Every mediator node stores this information to retrieve data later on.

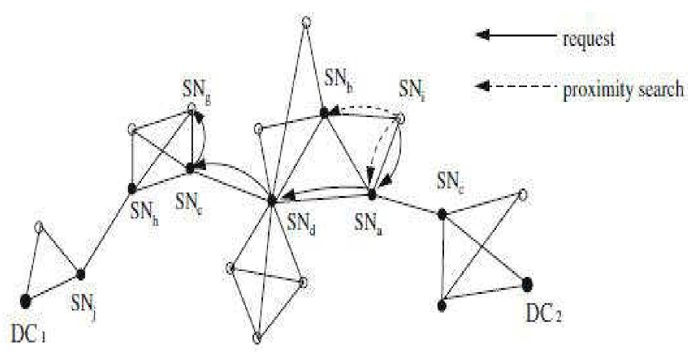


Fig 2. A request packet from node SNi being supplied by node SNg.[11]

As shown in Fig 2, the black nodes represent mediator nodes that will maintain information about the data items being cached by different nodes.

A suitable cache discovery mechanism and cache replacement policy is used to enhance the effectiveness of this cooperative caching protocol. Whenever a node requests for data, it is first searched in local cache, if found there a local hit occurs, otherwise (local miss) then the request is broadcasted to mediator nodes which look up the stored information to find the node that has cached the requested data item. If there exists any node that has cached the data, the node is responded with the node address that has stored data item else proximity miss is said to occur and data is retrieved from the data source.

Sometimes data item may be served by the non 1-hop neighbors lying in the path from node to the data source, in that case a remote hit is said to occur. This cache discovery mechanism keeps track of the remaining energy of nodes while making data retrieving decisions.

This protocol ensures better performance by exploiting the most central nodes in network to coordinate caching decisions.

T.P. Sharma et al. proposed a Cooperative caching scheme which utilizes dual radio based antennas to cache data at the nodes nearer to the sink [12]. This caching scheme exploits the cooperation of the nearby sensor nodes resulting in a larger cache known as *cumulative cache*. The entire sensor field is a square sized grid. The grid is formed such that the diagonal of cell shouldn't exceed the RH (High transmission range) of the antenna. Each cell of grid is considered as a cluster. The corner nodes of the each cluster act as Dissemination nodes (DN) those are responsible for forwarding the data using the RH.

Cumulative caching is done by making the Cache Zone utilizing the low power antenna (RL) of the sensor nodes. Cache zone is taken as region of radius RL/2 around each dissemination node on the data/query path. Choosing radius RL/2 makes all nodes in a cache zone to communicate in a single hop within the cache zone.

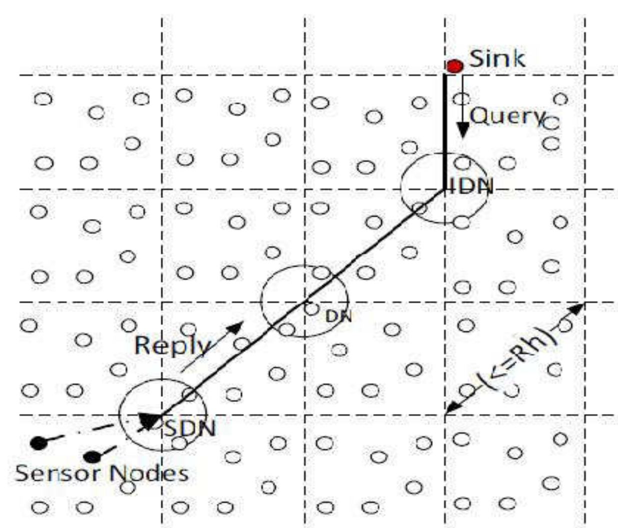


Fig 3. Query and data flow path in DRDD[12]

As shown in Fig 3, initially the sink stores the data in its local cache. Once the local cache is filled, sink looks for its IDN (Immediate Dissemination Node) to store the data. IDN stores the data in its local cache, once it is full; it utilizes the cache of the nodes in the cache zone to store data. This continues until all the nodes of all cache zones are filled. Suitable Cache Admission, Cache discovery and Cache replacement policies are used for effective cache management.

This scheme utilizes cache of the neighboring nodes of Cache zone to increase data proximity nearer to the sink and saves the energy used to access the data from the source. This scheme works well when used in 2-D geometry sensor field only and can't be applied to 3-D geometry sensor field as grid formation becomes difficult. The use of same DN's lying in the path of the sink to source, will lead to their failure due to energy depletion because of their continuous use.

J. Xu et al. proposed an algorithm that improves energy efficiency in WSNs by waiting caching [13]. It aims to improve energy efficiency by reducing network traffic. As shown in below Fig 4, the whole network is assumed to be divided into small clusters with a cluster head.

The cluster head is responsible for communication with the other cluster head and does not perform any sensing activity. Within a cluster, at a particular instant of time only cluster head and a node sensing the data will be active while remaining will be in sleeping mode and hence conserving energy. As depicted in Fig. 4 H2 (Cluster Head) will aggregate the data from node N2 and will forward it to Cluster Head H1. H1 will not forward data to base station until it receives data from lower cluster and N1. This is known as Waiting Cache. In this way network traffic is reduced and energy is conserved.

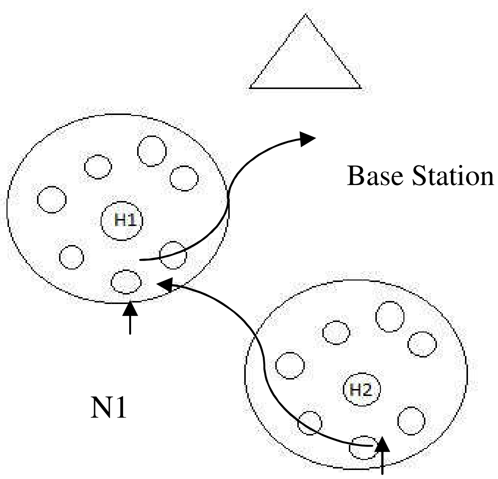


Fig 4. Cluster formation and data flow path [13]

Since the cluster heads may run out of en N2 › periodically the cluster head can be chosen. This Energy Efficient Waiting Cache Algorithm provides very high performance if cluster size is chosen carefully.

Md. A. Rahman et al. gave away to provide effective caching [14]. Apart from efficient data routing techniques, there are many factors that can improve the overall efficiency of the wireless sensor networks. The factors that can affect the working of the WSN are:-

- Data Negotiation between sink and the sensors – this scheme aims at sending the data to the sink only when the sensed data changes. The sink will cache the latest data from all the sensors. The cached data in the sink will be invalidated whenever any sensor becomes inactive.
- Developing Expectancy of Data change –this scheme proposes to avoid unnecessary sensing of data by the sensor nodes. The nodes should know when its data is going to change and should accordingly sense the data.
- Data Vanishing- Duplicate sensed data can be vanished by a routing node in between the source node and sink, and only single packet is send to the sink, as there can be more than one sensor which would have sensed the same data.

The advantage of this scheme is in exploiting the given new constraints that certainly gives better way to design protocols for WSNs.

Dae-Young Kim et al. proposed an active caching technique that aims at providing the desired communication reliability (CR) during the multihop communication over the sensor networks [15]. To handle the data losses there is two types of recovery mechanisms - end-to-end loss recovery (E2E) and hop-by-hop loss recovery (HBH). The E2E mechanism sends the lost packets from source whereas in HBH retransmission is done with the help of intermediate nodes. To provide 100% reliability E2E loss recovery mechanism is used which offers end-to-end delay because of retransmissions from the source. The end-to-end delay can be handled by storing data in the intermediate nodes by employing HBH transmissions but it demands huge memory requirements. Thus there generally occurs a tradeoff between the end-to-end delays and memory requirements of the nodes. The active caching aims at providing the flexible method of finding the intermediate nodes where caching of the data needs to be done to provide communication reliability demanded by the application.

Active caching approach is based on dynamic programming. The whole problem is divided into smaller sub problems.

Problem: $Ptx(H) > CR$.

Subproblem: $Ptx(h) > CR$, where $h = 1, 2, \cdots, H$.

The packet delivery rate $Ptx(H)$ during total hop counts H should be greater than the desired communication reliability

CR. To do that, the packet delivery rate *Ptx*(*h*) during hop counts *h* in each hop should be greater than the *CR*.

The key idea for solving the problem is to cache data packets if the probability of packet transmission does not satisfy the desired communication reliability. By solving the sub problems, we can solve the entire problem. Whenever a packet loss occurs between source and caching node, the packet is requested by caching node to the source node but if the packet loss occurs between caching node and destination node, the packet is retransmitted by the caching node. Thus Active caching proposes a mechanism to handle E(C) i.e. total retransmission counts and E(B) i.e. memory requirements of the networks deploying nodes, to satisfy the required Communication Reliability (CR).

Thus we can analyze that the main thrust of any Caching technique is the position of caching nodes in the network. The nodes caching the data should support the needs of different sinks with minimum energy requirements. Apart from caching nodes, Cache Admission, Cache Discovery and Cache Replacement are other issues that can affect the performance of any caching technique. Keeping in view all these factors, we can explore the caching techniques in which caching nodes are nearer to the sink node. The caching techniques deploying different arrangements of caching nodes around sink node can be worked upon to find the technique in which the better and improved energy efficiency is achieved. The different Cache Admission, Cache Discovery and Cache Replacement policies can also be explored to find the improved caching techniques.

IV. CONCLUSION

In this paper various caching techniques have been discussed that can be employed in WSN to offer different benefits.

Energy efficiency is the key issue that needs to be addressed in WSN. Caching when used effectively can increase the energy efficiency and lifetime of the network manifold. The data availability increases and network traffic reduces leading to bandwidth savings. Various other new caching techniques can also be proposed by exploiting the key aspects of caching to implement efficient WSN and to handle the constraints of WSN.

REFERENCES

[1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, "Wireless sensor networks: a survey", *IEEE Communications Magazine* , Vol. 38, Issue 4, 15 March 2002, pp. 393–422.

[2] Jennifer Yick, Biswanath Mukherjee, Dipak Ghosal, "Wireless sensor network survey", *Elsevier Computer Networks* , Vol 52, Issue No. 12, pp. 2292–2330,2008.

[3] N. Noury, T. Herve, V. Rialle, G. Virone, E. Mercier, G. Morey, A. Moro, T. Porcheron, "Monitoring behavior in home using a smart fall sensor", *IEEE-EMBS Special Topic Conference on Microtechnologies in Medicine and Biology*, pp. 607–610, October 2000.

[4] C.K. Chang and J. Huang, "Video surveillance for hazardous conditions using sensor networks", *in Proceedings of IEEE International Conf. on Networking, Sensing and Control*, Vol.2 , pp. 1008–1013, March 2004.

[5] A. Jain, Y.F. Wang and E. Y. Chang, "A collaborative camera system for surveillance", *in UCSB Technical Report*, Nov. 2004.

[6] Tao Chi, Ming Chen, Qiang Gao, "Implementation and Study of a Greenhouse Environment Surveillance System Based on Wireless Sensor Network", *The 2008 International Conference on Embedded Software and Systems Symposia*, 2008.

[7] L. Nachman, R. Kling, R. Adler, J. Huang and V. Hummel, "The intel mote platform: a bluetooth-based sensor network for industrial monitoring", *in IPSN 2005*, pp. 437–442, Apr. 25–27 2005.

[8] N. S. Love, I. Masaki, and B. K. P. Horn, "Efficient traffic monitoring", in MITITRC Report,2003.[Online]. Available:http://www.mtlweb.mit.edu/researchgroups/itrc/ITRC_report/MTL_2003_report/nsl_mtl_annual_2003.doc

[9] J.R. Polastre, "Design and implementation of wireless sensor networks for habitat monitoring", M.S. Thesis, U.C. Berkeley, 2003.

[10] M.A.Miskam, A. B. Nasirudin, I. A. Rahim, "Preliminary Design on the Development of Wireless Sensor Network for Paddy Rice Cropping Monitoring Application in Malaysia" , *ISSN*, Vol. 37, Issue No.4, pp. 649-657, 2009.

[11] D.Nikos, K.Dimitrios, M.Yannis, "Cooperative Caching in Wireless Multimedia Sensor Networks", *Third International Mobile Multimedia Communication Conferenc,*Vol 13, Issue 3-4, pp. 337-356, 2007.

[12] T.P. Sharma, R.C. Joshi, Manoj Misra, "Dual Radio based Cooperative Caching for Wireless Sensor Networks", *16th IEEE Conference*, pp.1-7, 2008.

[13] J. Xu, K. Li, Y. Shen, J. Liu, "An Energy-Efficient Waiting Caching Algorithm in Wireless Sensor Network", *International Conference on Embedded and Ubiquitous Computing*, Vol. 1, pp. 323 - 329, 2008.

[14] Md. A. Rahman, S. Hussain, "Effective Caching in Wireless Sensor Network", *21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07)*, Vol. 1, pp. 43-47 2007.

[15] Dae-Young Kim, Jinsung Cho, "Active Caching: A Transmission Method to Guarantee Desired Communication Reliability in Wireless Sensor Networks", *IEEE communications letters*, Vol. 13, Issue No. 6, JUNE 2009.

Dynamic WSN Nodes Based on Flexible Power Management Techniques

Preet Kanwal
MCP College
preetgf@gmail.com

Robin Chadha
CT Institutes
chadha_robin@yahoo.com

Jatinderpal Singh
CT Institutes
jhsingh11@gmail.com

Ashima
Guru Nanak Dev
University, Amritsar

***Abstract*—The most important aspect for wireless sensor networks (WSNs) is its Energy Efficiency because the consumable energy is limited. Each WSN node has its own power supply. Basically, the lifetime of the WSN node depends on its average power consumption. Therefore, an efficient supply of the WSN node can enhance the lifetime of it. The various components of a WSN node (microcontroller, transceiver, sensors) have different supply voltage ranges. The supply voltage of the node should be as low as possible to save as much energy as possible. Therefore, a voltage converter is needed to reduce the voltage of the battery. The minimum allowed supply voltage depends on the components that are active. The active components and consequently the minimum allowed supply voltage vary over time. Flexible Power Management Techniques (FPMT) can be used to adapt the supply voltage of the node.**

Keywords—WSN, FPMT, PU

I. INTRODUCTION

A WSN consists of a lot of sensor nodes. Each sensor node is an intelligent device which is responsible to forward messages through the network and to prepare the measured data for the transmission. These tasks are typically fulfilled by a microcontroller. The missing wired infrastructure causes the need of dedicated energy sources for each sensor node. Conventional batteries are cheap and provide a reliable supply for a certain period of time because each battery has a limited capacity. Therefore, the lifetime of battery powered sensor nodes is also limited. Energy harvesting systems (EHSs) can be used to extend the lifetime of the sensor nodes [1]–[3]. They utilize environmental energy sources. Energy harvesting devices (EHDs) convert the energy of the environmental sources into electrical energy. Typical EHDs are solar cells and thermo-generators. The converted energy must be stored because energy from the environment is generally unpredictable, discontinuous and unstable [4]. The available energy is often very low compared to the energy consumption of the sensor node. These are the reasons of the strict power constraints for WSNs.

The application areas of WSNs are manifold. Precision agriculture [5], [6], wildlife monitoring [7], [8], human health care [9] and structural health monitoring [10] are only a few examples. Typically, a sensor node is adapted to its special application area. Therefore, only the necessary sensors and components are embedded into the sensor node and significant energy can be saved. The needed sensors depend strongly on the application area and are often predetermined. Each sensor has its own power requirements and its own supply voltage range. Therefore, the components (especially the microcontroller) have to be adapted to the sensors to ensure proper functionality.

In many cases, the supply voltage of the sensor node is set to a fixed high value because only one component of the node needs this high voltage. The other components do not need such a high supply voltage. Therefore, lot of energy is wasted because of a single component.

Another case of wasting energy is the direct connection of the sensor node to the power unit (PU) e.g. battery. The voltage of almost all PUs depends on its energy level. Typically, the voltage decreases with decreasing energy level of the PU. Each sensor node has a specific supply voltage range wherein a proper functionality is guaranteed. Below the lower limit, the sensor node cuts out. Therefore, the voltage of the fully charged PU must be higher than this lower limit. The emerging difference is wasted energy. This difference depends on the lower limit of the supply range and the type of the PU.

II. RELATED WORK

As mentioned before, WSN nodes are often supplied by conventional batteries. The problem is the limited lifetime. It can be extended by reducing the power consumption of the sensor node. However, this extension is limited by the leakage currents of each battery. Figure 1 shows the expected lifetime of a single AA cell (Ni-MH) with a capacity of 2.6 Ah depending on the current consumption of the sensor node. A typical leakage current of such a cell is 30 A [11]. Therefore, the batteries have to be replaced after a certain period of time. This is called battery replacement problem as mentioned in [12]. The replacement is sometimes very difficult because the WSN nodes are deployed at hard-to-access locations. Hence, the replacement should be avoided if possible.

An EHS can extend the lifetime of a WSN node by utilizing environmental energy sources. A theoretical energy harvesting model is introduced in [13]. The model deals with the description of environmental energy sources and the energy sink (consumer). Two conclusions can be derived from this model. First, the EHS needs a buffer to bridge the discontinuous power output of the environmental energy source. There are different types of ESUs that can be used e.g. rechargeable batteries (nickel-cadmium NiCd, nickel-metal hydride NiMH, lithium ion LiIon or lithium-ion polymer LiPo) or double layer capacitors (DLCs).

Second, if the average power consumption of the WSN node is lower than the average harvested power of the environmental energy source and the ESU is well dimensioned then a perpetual operation is possible. This solves the battery replacement problem.

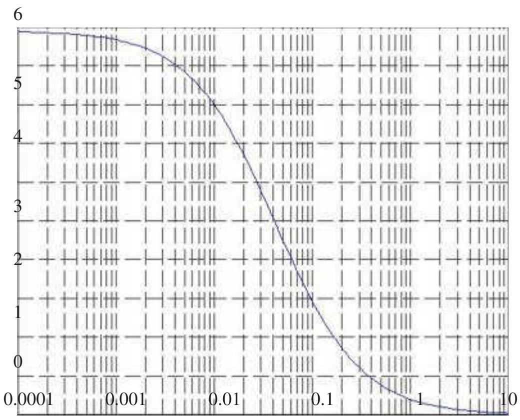


Fig. 1. Lifetime of a battery-powered system versus average power consumption using a simple battery model of a single AA cell (NiMH) with 2.6 Ah and a leakage current of 30 A [11].

The amount of harvestable energy depends on the type of the environmental energy source, on the power of the energy source and on the efficiency of the EHD. Table I shows a list of commonly used energy harvesting technologies and their power densities [2], [14]. It can be seen that the power density of the solar cell depends dramatically on the location. If the EHS is deployed outdoor, solar cells have the best power density. Therefore, the power consumption of the WSN node must be lower or the size of the EHD must be increased.

TABLE I. Power densities of different harvesting technologies [2], [14].

Harvesting technology	Power density	
Solar cells (outdoor at noon)	10-15 mW/cm^2	
Solar cells (indoor) [15]	10 W/cm2	
Piezoelectric (shoe inserts)	300	W/cm3
Vibration (small microwave oven)	100	W/cm3
Thermoelectric (10 C gradient)	40	W/cm3
Acoustic noise (100dB)	1	W/cm3

Dynamic voltage scaling (DVS) is commonly used to adapt the supply voltage of a processor depending on its clock frequency. The clock frequency can be reduced if the workload of the processor is low. In [16] the design issues for DVS are explained. In [17] they describe various algorithms that can be used for DVS. These algorithms reduce energy consumption bychanging processor speed and voltage at run-time depending on the needs of the applications running [17]. In [18] DVS isapplied to low power microprocessors. They use a low power Strong ARM embedded Linux platform running a multimedia application for evaluation. In [19] they describe the energy savings using dynamic frequency and voltage scaling. The reduction of the frequency causes linear energy savings. Additionally, the adaption of the supply voltage to the lower clock frequency of the processor causes quadratic energy savings. However, all the works only consider the power consumption of the processor core itself. Other components are not included in their calculations.In [20] they have implemented a dynamic voltage and frequency scaling platform using an MSP430 ultra-low power microcontroller from Texas Instruments and other components of the shelf (COTS). This approach is very interesting because they can vary the supply voltage of the MSP430 microcontroller. However, they do not consider the efficiency and the leakage currents of the voltage converter. The following section describes FPMT applied on the whole sensor node and the influence of the voltage converters.

III. ENERGY EFFICIENT SUPPLY OF WSN NODES

This section starts with the basics of the energy efficient supply of WSN nodes.As mentioned in the introduction, an energy efficient supply of a WSN node is very important and can extend the lifetime of such a node. Each sensor node typically consists of a microcontroller and some sensors.

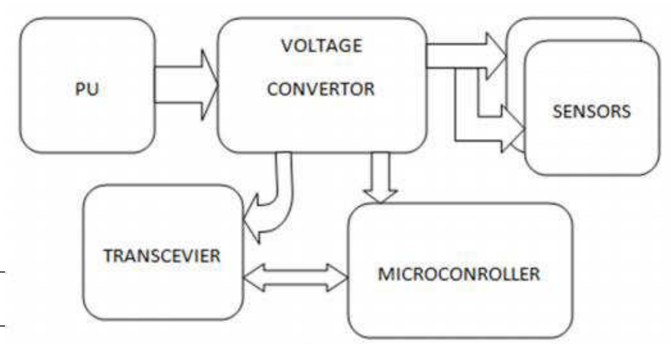


Fig. 2. Structure of a wireless sensor node (extended from [21]).

To guarantee an efficient supply of the WSN node, a conversion circuit is needed. Typically, this converter reduces the battery voltage to a constant value. Figure 2 shows the structure of a wireless sensor node. The converter is placed between the ESU (e.g. a battery or a DLC) and the rest of the sensor node. Usually, the whole sensor node is supplied by the same voltage. Multiple supply voltages are possible but it dramatically increases the complexity of the hardware because each voltage level needs its own converter and between each voltage level, some level shifter hardware is needed. Therefore, it is not useful to supply a sensor node with different voltage levels simultaneously in general.

Energy can be saved by switching off the components which are not needed currently by the application. This means the power supply of each component should be controllable by the microcontroller. Then, the driver software of a component or the application software itself can enable or disable the specific components. Therefore, a program or a part of a program has a list of the used hardware components L. The only component which is supplied all the time is the microcontroller. Hence it is always on the list L. The rest of the list depends on the components that are currently needed by the application (transceiver, sensors). Each of the components has its own supply voltage range with a lower limit $V_{sup,min,component}$. This limit is the lowest possible supply voltage (LPSV) of a component. The list L and the LPSVs $V_{sup,min,component}$ of the components result in a list of the LPSVs $L_{LP\ SV}$. Due to the fact that all hardware components of a sensor node are supplied by the same voltage, the minimum allowed voltage can be determined as shown in Equation 1.

$$^{V}sup,min,node = {}^{max}({}^{L}LP\ SV\ ) \qquad (1)$$

The list of the needed hardware components L depends on the application or even on the part of the application which is currently executed by the microcontroller. Therefore, the list changes over the time and consequently the list $L_{LP\ SV}$ and the minimum allowed supply voltage $V_{sup,min,node}$. To be as efficient as possible, the supply voltage of the node should be equal to the minimum allowed supply voltage: $V_{CC} = V_{sup,min,node}$. This variation depends on the application and on the embedded hardware components of the node. The following section describes the assumptions for further considerations.

TABLE II Hardware components and their supply voltage range

Hardware Component	Supply Range	$V_{sup,min,component}$
Microcontroller	1.8 V to 3.6 V	1.8 V
Temperature Sensor	3.3 V to 3.3 V	3.3 V
Transceiver	2.4 V to 3.6 V	2.4 V

IV. SCENARIO

A typical energy saving method used in WSNs is duty cycling (DC). The sensor nodes are only active for a short period of time t_{active}. During this time, the node collects the information of its environment, preprocesses the data and transmits the data. During the rest of the time t_{sleep}, the sensor node is placed into a sleep state to save as much energy as possible. Thus, one interval T consists of the active period t_{active} and the sleep period t_{sleep}.

An MSP430F1611 from Texas Instruments is used as microcontroller. This microcontroller has a supply voltage range from 1.8 V to 3.6 V [22]. To be able to use the full supply range, the clock frequency of the microcontroller has to be limited to 4MHz. The active mode current I_{AM} is calculated according to the datasheet. First, the current consumption at a clock frequency of 1 MHz and the given supply voltage V_{CC} is calculated as shown in Equation 2. Therefore, the current consumption at 1 MHz and 3 V is used ($I_{AM,1MHz,3V} = 500$ A). Then, the current consumption at the given clock frequency f can be calculated as shown in Equation 3).

$$^{I}AM,1MHz\ (^{V}CC\ )\ =\ ^{I}AM,1MHz,3V\ +$$

$$210\ \ A/V \cdot (V_{CC} - 3\ V) \qquad (2)$$

$$I_{AM}\ (V_{CC}\ , f\ )\ =\ I_{AM,1MHz}\ (V_{CC}\ ) \cdot f\ [MHz] \qquad (3)$$

The sleep mode current I_{sleep} (low power mode 3, only the 32 kHz timer is active) is interpolated using the two values given in the datasheet [22] (Equation 4).

$$I_{sleep}\ (V_{CC}\ ) = 1.3\ \ A \cdot (V_{CC} - 2.2\ V) \cdot {}^{2.6\ A-1.3\ A} \qquad (4)$$

$$= 3\ V - 2.2\ V$$

The WSN node should be used to measure the temperature of the environment. Therefore, the precise temperature sensor TMP05B from Analog Devices is embedded. It has a typical accuracy of 0.2 % if the supply voltage is in the range of 3.135 V to 3.465 V . The full supply voltage range is from 3.0 V to 5.5 V. However, to guarantee the best accuracy, the supply voltage of the temperature sensor should be 3.3 V. Furthermore, the sensor node uses the MRF24J40MB 2.4 GHz, IEEE802.15.4 transceiver module from Microchip. It has a supply voltage range from 2.4 V to 3.6V. Table II summarizes the embedded components and their supply voltage ranges.

Furthermore, it is assumed that the sensor node is in a sleep state (t_{sleep}) during 85 % of the interval T . Then the sensor node wakes up and performs some measurements (t_{meas}) during 5 %. After that, the measured data is preprocessed (t_{comp}) during 5 % of the interval. Therefore, no other hardware components are needed. Finally, the sensor node sends the results through the network (t_{comm}) during the last 5 % of the interval.

As shown in Fig. 2, a voltage converter is placed between the PU and the components of the sensor node. The minimum supply voltage of the WSN node is 3.3 V and the maximum supply voltage is 3.6 V. This is a very small range and there is no battery type that operates well in this range. It would be possible to use a battery, which is not fully loaded but this would impair dramatically the usable capacity of the battery. Therefore, it is not possible to supply the node directly. The voltage converter reduces the battery voltage to a constant value. However, a little voltage drop $U_{drop,min}$ remains at the converter. The relation is shown in Equation 5.

$$^{U}battery,min = {}^{U}drop,min + {}^{U}sup,node \qquad (5)$$

Therefore, the minimum voltage of battery $U_{battery,min}$ must be higher than the supply voltage of the node $U_{sup,node} = 3.3$ V. For further calculations, a minimum voltage drop of $U_{drop,min} = 0.2$ V of the converter is assumed [able III shows five possible battery types to supply the WSN node. All values are from The first column shows the type of the battery and the second column shows the number of needed cells which have to be connected in series to achieve the minimum supply voltage of the node. The third column lists the voltage values of the fully charged cell (or cells) and the fourth column shows the average voltage during a full discharge process. Finally, the last column shows the estimated usable capacity of the battery concerning the supply voltage of the node and the minimum voltage drop of the converter.

The minimum voltage of the battery is $U_{battery,min} = 3.5$ V. The table shows that the LiIon and the LiPo rechargeable batteries have the best usable capacity. The alkaline battery is not rechargeable and is only given as reference. The low usable capacity is caused by the faster drop of the cell voltage compared to the other types.

TABLE III. Possible battery types for a supply of the wsn node [27].

Battery	Needed	Start	Avg. Discharge	Usable
Type	Cells	Voltage	Voltage	Capacity
Alkaline	3	4.5 V	3.6 V	65 %
NiMH	3	4.14 V	3.75 V	89 %
NiCd	3	4.44 V	3.78 V	92 %
LiIon	1	4.1 V	3.76 V	95 %
LiPo	1	4.1 V	3.8 V	95 %

V. VOLTAGE CONVERSION TECHNIQUES

As mentioned before, energy can be saved if the voltage of the battery is reduced using a voltage converter. Basically, there are two different types of voltage converters which are explained in the following two sections.

1). *Linear Voltage Regulators:* This type provides a constantoutput voltage by adapting its internal resistance. So called low-dropout (LDO) regulators are linear voltage regulators which can operate with a low difference of the input and the output voltage.

The dependency of the current (blue trace) and the power consumption (red trace) of the microcontroller on the supply voltage during active mode at 4 MHz. The values are simulated using Equation 2. It is a linear dependency of the current consumption on the supply voltage. When using an LDO regulator, the difference between the battery voltage V_{Bat} and the supply voltage of the microcontroller V_{CC} drops at the regulator V_{LDO} (see Equation 6).

$$^{V}Bat = {}^{V}LDO + {}^{V}CC \qquad (6)$$

The current through the microcontroller is equal to the current through the LDO (Equation 7).

$$^{I}LDO = {}^{I}\ C \qquad (7)$$

The power consumption of the microcontroller and the LDO can be calculated as shown in Equation 8.

$$\begin{aligned} P\,(V_{CC}\,) &= {}^{I}LDO \cdot {}^{V}Bat \qquad (8) \\ &= {}^{I}AM\,({}^{V}CC\,) \cdot {}^{V}Bat \end{aligned}$$

It can be seen that the power consumption depends linearly on the supply voltage. The reason is that the input current of the LDO regulator is the same as the output current and therefore the current of the microcontroller.

The dependency of the current (blue trace) and the power consumption (red trace) of the microcontroller depending on the supply voltage during sleep mode (low power mode 3, only the 32 kHz timer is active). Due to the fact that the sleep current depends also linear on the supply voltage the resulting power consumption has a linear dependency, too.

2). *Step Down Converter*: A step down converter, also called buck converter, provides a constant output voltage by using internal switching elements. Typically a coil or a capacitor is used in combination with the switches. The output voltage can be controlled by varying the timing of the switching elements (pulse-width modulation).

The dependency of the current (blue trace) and the power consumption (red trace) of the microcontroller depending on the supply voltage during active mode at 4 MHz. The values are simulated using Equation 2. When using a step down converter, the difference between the battery voltage V_{Bat} and the supply voltage of the microcontroller V_{CC} drops at the converter V_{Buck} (see Equation 9).

$$V_{Bat} = V_{StepDown} + V_{CC} \qquad (9)$$

However, the average input current of the converter is unequal to the output current and therefore, it is different from the current through the microcontroller (Equation 10).

$$^{I}Buck,input\_={}^{I}Buck,output = {}^{I}\ C \qquad (10)$$

The ideal step down converter has no losses and therefore, the input power of it is equal to the output power. This means that the input power depends only on the power consumption of the microcontroller as shown in Equation 11.

$$\begin{aligned} P_{IN}\,(V_{CC}\,) &= P_{OUT}\,(V\ CC) \qquad (11) \\ &= I_{AM}\,(V_{CC}\,) \cdot V_{CC} \end{aligned}$$

It can be seen that the power consumption depends quadratically on the supply voltage. The reason is that the input current of the step down converter is lower than the output current on the supply voltage, too.

VI. FLEXIBLE POWER MANAGEMENT TECHNIQUE

FPM technique can be applied to reduce the power consumption of the hardware. It is assumed that only a few messages are sent. Therefore, the energy of the transmissions is neglected at the following calculations. This can be done, because the main focus of this work is on the possible energy savings using CADVS and not on a correct energy estimation of an application running on the sensor node. Furthermore, it isassumed that the transceiver and the temperature sensor can be switched off completely.

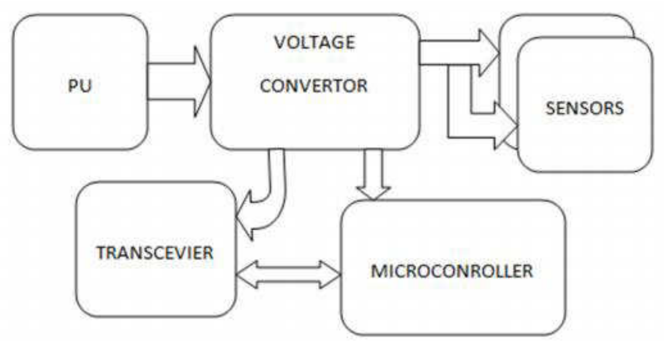


Fig. 3 Block diagram of sensor node

This is necessary, because they have very high sleep currents (compared to the microcontroller). Fig. 3 shows the resulting block diagram of the sensor node. TPS22921 from Texas Instruments are assumed as power switches.

Table IV shows an overview of the current consumption and power consumption of the single components during the different subintervals.

TABLE IV Power consumption of the components of the sensor node

Component	Sleep	Active Measure	Active Computation	Active Commun.
V_{CC}	1.8 V	3.3 V	1.8 V	2.4 V
Microcontroller	1170 nW	7.43 Mw	1.786 mW	3.6 mW
Transceiver	0	0	0	60 mW
Temp. Sensor	0	1.221 mW	0	0
Power Switches	306 nW	1327 nW	306 nW	621 nW

As mentioned before, two different techniques can be used to convert the supply voltage of the WSN node. The following figures and tables show the possible power savings using an LDO regulator with fixed output voltage (3.3V), an LDO regulator with variable output voltage (CADVS), a step down converter with fixed output voltage (3.3 V) and a step down converter with variable output voltage (CADVS).

The voltage of the battery is assumed to be V_{bat} = 3.77 V. This is the mean of the average discharge voltages of the rechargeable batteries . The power savings are calculated using the LDO regulator with fixed output voltage as reference, because it is the simplest supply configuration .

It is also important to consider the quiescent current of the voltage converters. The LDO regulator TPS78001 from Texas Instruments has a quiescent current of $I_{q,LDO}$ = 500nA. The quiescent current of the step down converter TPS62120 also from Texas Instruments is $I_{q,stepdown}$ = 11 A . Furthermore, the efficiency of the TPS62120 is about 90 %. To adjust the output voltage a digital potentiometer is needed. The quiescent current of the ISL22313 from Intersil is $I_{q,dpot}$ = 1 A. The results of this calculation are shown in Table VIII.

Fig. 4 shows the power trace of the WSN node using the four different supply configurations. Three intervals are plotted. The white trace shows the minimum possible supply power using ideal components without conversion losses and leakage currents.

The reason is the energy wasting of the step down converter during the sleep subinterval. Analyzing the savings during each subinterval, it can be seen that the step down converter has much higher power savings at higher power consumptions and lower supply voltages.

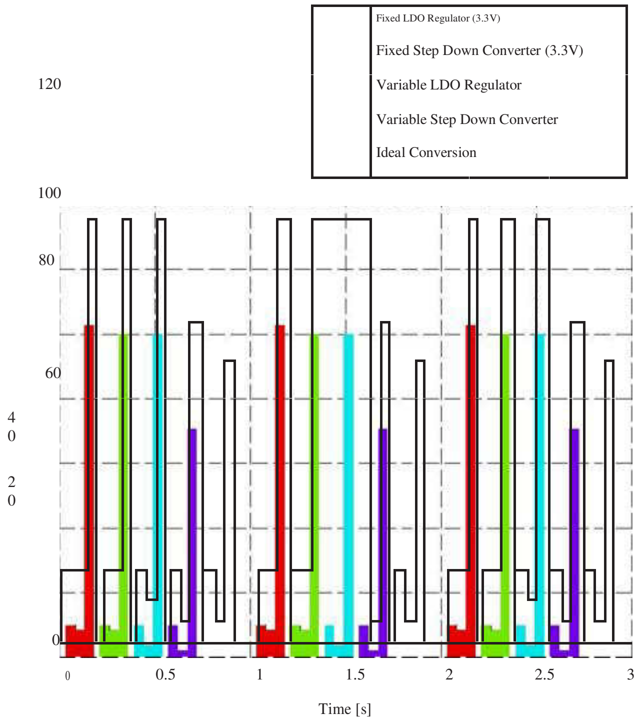


Fig. 4. Power consumption of the WSN node with four different supply configurations and ideal conversion.

The average power savings depends on the DC of the WSN node. Therefore, the sleep duration is varied and the durations of the active subintervals are kept constant. The shorter the sleep duration, the higher is the DC. It can be seen that at a higher DC the step down converter using CADVS provides the best results. At a lower DC, the LDO regulator using CADVS provides the best results. However, there are applications with a dynamic DC. This means the DC is adapted to the current circumstances. The best solution is to combine an LDO regulator and a step down converter. Only the voltage converter with the better power savings is active.

VII. CONCLUSION AND FUTURE WORK

It has shown different voltage conversion techniques along with their advantages and disadvantages regarding the energy saving system. A realistic scenario has been developed to show how DVS can be used to save up to 31.5 % of the energy compared to a constant voltage supply using the introduced scenario while achieving the same end-user performance. Furthermore, it showed the energy savings dependency on the DC of a WSN node in which the leakage currents and the conversion efficiency of the four different voltage conversion techniques are considered. In future, different voltage conversion techniques with real components will be evaluated by implementing the prototype.

REFERENCES

[1] R. Min, M. Bhardwaj, S.-H. Cho, N. Ickes, E. Shih, A. Sinha, A. Wang, and A. Chandrakasan, "Energy-centric enabling tecumologies for wire-less sensor networks," Wireless Communications, IEEE, vol. 9, no. 4,2002.

[2] V. Raghunathan, A. Kansal, J. Hsu, J. Friedman, and M. Srivastava, "Design considerations for solar energy harvesting wireless embedded systems," in Proceedings of the 4th international symposium on Infor-mation processing in sensor networks. IEEE Press, 2005, p. 64.

[3] M. Rahimi, H. Shah, G. Sukhatme, J. Heideman, and D. Estrin, "Studying the feasibility of energy harvesting in a mobile sensor network," Robotics and Automation, 2003. Proceedings. ICRA '03. IEEEInternational Conference on, vol. 1, pp. 19 – 24 vol.1, sep. 2003.

[4] A. Janek, C. Trummer, C. Steger, R. Weiss, J. Preishuber-Pfluegl, and M. Pistauer, "Simulation based verification of energy storage architectures for higher class tags supported by energy harvesting devices," Microprocessors and Microsystems, vol. 32, no. 5-6, pp. 330– 339, 2008, dependability and Testing of Modern Digital Systems. [On-line]. Available: http://www.sciencedirect.com/science/article/B6V0X-4S80XHN-2/2/22459f6e78fe4b7c4111c8af88fa7fb4

[5] K. Langendoen, A. Baggio, and O. Visser, "Murphy loves potatoes: experiences from a pilot sensor network deployment in precision agri-culture," Parallel and Distributed Processing Symposium, 2006. IPDPS2006. 20th International, p. 8 pp., apr. 2006.

[6] N. Watthanawisuth, A. Tuantranont, and T. Kerdcharoen, "Microclimate real-time monitoring based on zigbee sensor network," Sensors, 2009IEEE, pp. 1814 –1818, oct. 2009.

[7] P. Juang, H. Oki, Y. Wang, M. Martonosi, L. S. Peh, and D. Rubenstein, "Energy-efficient computing for wildlife tracking: design tradeoffs and early experiences with zebranet," in ASPLOS-X: Proceedings of the10th international conference on Architectural support for programming languages and operating systems. New York, NY, USA: ACM, 2002, pp. 96-107.

[8] A. Lindgren, C. Mascolo, M. Lonergan, and B. McConnell, "Seal-2-seal: A delay-tolerant protocol for contact logging in wildlife monitoring sen

[9] P.K. Lorincz, B.-r. Chen, G. W. Challen, A. R. Chowdhury, S. Patel, Bonato, and M. Welsh, "Mercury: a wearable sensor network platform for high-fidelity motion analysis," in SenSys '09: Proceedings of the 7thACM Conference on Embedded Networked Sensor Systems. New York,NY, USA: ACM, 2009, pp. 183–196.

[10] R.N. Xu, S. Rangwala, K. K. Chintalapudi, D. Ganesan, A. Broad, Govindan, and D. Estrin, "A wireless sensor network for structural monitoring," in SenSys '04: Proceedings of the 2nd international con-ference on Embedded networked sensor systems. New York, NY, USA:ACM, 2004, pp. 13–24.

[11] C. Enz, N. Scolari, and U. Yodprasit, "Ultra Low-Power Radio Design for Wireless Sensor Networks," in IEEE Int. Workshop on Radio-Frequency Integration Technology, 2005, pp. 1–17, (Invited Keynote).

[12] M. Minami, T. Morito, H. Morikawa, and T. Aoyama, "Solar biscuit: A battery-less wireless sensor network system for environmental mon-itoring applications," The 2nd International Workshop on NetworkedSensing Systems, 2005.

[13] A. Kansal, D. Potter, and M. B. Srivastava, "Performance aware task-ing for environmentally powered sensor networks," in SIGMETRICS'04/Performance '04: Proceedings of the joint international conference on Measurement and modeling of computer systems. New York, NY,USA: ACM, 2004, pp. 223–234.

[14] P. He, Q. Cui, and X. Guo, "Efficient solar power scavenging and utilization in mobile electronics system," Green Circuits and Systems(ICGCS), 2010 International Conference on, pp. 641 –645, jun. 2010.

[15] S. Roundy, D. Steingart, L. Frechette, P. Wright, and J. Rabaey, "Power sources for wireless sensor networks," in Wireless Sensor Networks, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2004, vol. 2920, pp. 1–17.

[16] T. D. Burd and R. W. Brodersen, "Design issues for dynamic voltage scaling," in Proceedings of the 2000 international symposiumon Low power electronics and design, ser. ISLPED '00. NewYork, NY, USA: ACM, 2000, pp. 9–14. [Online]. Available: http://doi.acm.org/10.1145/344166.344181

[17] T. Simunic, L. Benini, A. Acquaviva, P. Glynn, and G. De Micheli, "Dynamic voltage scaling and power management for portable systems," in Proceedings of the 38th annual Design Automation Conference, ser. DAC '01. New York, NY, USA: ACM, 2001, pp. 524–529. [Online]. Available: http://doi.acm.org/10.1145/378239.379016

[18] J. Pouwelse, K. Langendoen, and H. Sips, "Dynamic voltage scaling on a low-power microprocessor," in Proceedings of the 7th annualinternational conference on Mobile computing and networking, ser.MobiCom '01. New York, NY, USA: ACM, 2001, pp. 251–259. [Online]. Available: http://doi.acm.org/10.1145/381677.381701

[19] A. Sinha and A. Chandrakasan, "Dynamic power management in wire-less sensor networks," Design Test of Computers, IEEE, vol. 18, no. 2, pp. 62 –74, Mar./Apr. 2001.

[20] H.C. Powell, A. T. Barth, and J. Lach, "Dynamic voltage-frequency scaling in body area sensor networks using cots components," in proceedings of the fourth international conference on body area networks, ser. BodyNets '09. ICST, Brussels, Belgium. Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009, pp. 15:1–15:8. [Online]. Available: http://dx.doi.org/10.4108/ICST.BODYNETS2009.6015

[21] D. Puccinelli and M. Haenggi, "Wireless sensor networks: applications and challenges of ubiquitous sensing," Circuits and Systems Magazine,IEEE, vol. 5, no. 3, pp. 19–31, 2005.

[22] Texas Instruments, "Msp430f15x, msp430f16x, msp430f161x mixed sig-nal microcontroller," www.focus-ti.com, SLAS368F, 2009.

Evolution of Motives and Trends of DDoS: A Review

Avneet Dhingra
Department of Computer Science and Engineering
Punjab Technical University
Jalandhar, India

Monika Sachdeva
Department of Computer Science and Engineering
SBS College of Engineering and Technology
Ferozepur, India

***Abstract*— Internet, one of the greatest inventions of modern times, has enabled entirely new form of activities and social interaction. Business opportunities have widened, thus leading to professional development. But apart from shrinking the world and providing a vast amount of information, the malicious attackers use it to create havoc and inconvenience to authorized users. One of the ways to disrupt the services is through Distributed Denial-of-service attacks. Since last few years, it has been most expensive security problem for major network operators and E-Commerce websites. These attacks can cause financial overheads and losses up to millions of dollars to the victim. This paper reviews the history and evolution of such attacks since 1990s till March, 2014. The change in the motives behind these attacks and trends in technologies used over the time period, to bring down victim websites, has been discussed. The trend line at the end sums up the trends in attack and motives over the years.**

Keywords*— *DDoS attacks, motives, trends, bandwidth, resource

I. Introduction

One of the greatest inventions of the century, the Internet, has widened the opportunities for business and professional developments. It offers an enormous amount of useful information, thus leading to its growth. As per Wikipedia report, the number of internet users globally rose from 394 million to more than 7.1 billion in 2013. Apart from plethora of knowledge it provides, it presents some threats too.

Hackers and worms are constantly on a lookout for vulnerabilities in the security of the computers to inject malicious code to attack the one or all the three components of Internet security- Availability, Confidentiality and Integrity. Cyber attacks by malicious party, aiming to disrupt a website on the Internet, are called an *Availability based attacks.* Such attacks aim to exhaust the resources of the victim, make the victim inaccessible to legitimate users and are known as Denial of Service (DoS) attacks. If the source of attack is more than one system, the attack is known as Distributed Denial of Service (DDoS) attack.

DDoS attacks are a popular way to impact people, organizations, and even nations in malicious ways. These are used as a form of retribution, as a means for activists to further their causes, and even as a strategy for a military to create an advantage during warfare. They are capable of having a devastating effect [3].

In this paper, we discuss the trend line of DDoS attacks in last 15 years. Section II of this paper, discusses about DoS and DDoS attacks and their classification as a Bandwidth depletion attack and resource depletion attack. Section III explains the reasons why the internet is susceptible to DDoS attacks. Section IV gives the history of DDoS attacks along with the changing trends and motive. The study has been conducted from year 1996 (beginning of DDoS) till March 2014. This section gives the DDoS attacks in chronological order. The paper concludes in Section V.

II. DoS & DDoS Attacks

The DoS attack uses a single machine's resources to exhaust the resources of another machine, to prevent it from functioning normally. A way to increase the efficiency of a DoS attack, while evading detection and blocking, is to split the attack load among numerous machines simultaneously [6]. Such coordinated attacks are called Distributed Denial of Service attack, or DDoS, - the most effective form of DoS today. DDoS attacks are often executed using massive botnets and compromised servers. The attackers never legitimately control their attacking machines; rather, they infect thousands of computers spread across the world with specialized malware in order to gain unauthorized access to such machines.

Traffic volume generated by DDoS can exceed 10 Gbps [Arbor 2005]; the IP source traceback is extremely difficult as the source is widely distributed; and traffic seems to be legitimate and it is difficult to filter attack traffic without disturbing legitimate traffic. All these features (Peng et al) make it difficult to mitigate such attacks.

According to Specht and Lee, DDoS attacks can be divided into basically two classes: *Bandwidth Depletion Attacks* and *Resource Depletion Attacks*. The main aim of attacking in bandwidth depletion attacks is to congest the victim system's bandwidth. The flood attacks like UDP flood and ICMP ECHO packets send large volume of packets via zombies to the victim system and slow it down, crash the system or saturate the network bandwidth. A DDoS amplification attack like Smurf attack, Fraggle attack, DNS amplification attack and NTP amplification attack aim to amplify and reflect the attack by broadcasting IP address feature found on most routers.

Resource Depletion attacks tie up network resources so that they are not available for legitimate users. They do it by sending malformed packets or packets that misuse network protocol communications. TCP SYN attack is an example of such an attack. It exploits the weakness of the 3-way handshake in the TCP connection setup [2].

III. Reasons for the High Frequency of Attacks

DDoS attacks pose a great threat despite of being careful and spending all the efforts and resources to secure the network. There could be a number of reasons for the same. These have been explained well by Mirkovic, Houle & Weaver and Mitroktsa & Doilegeris [2, 32, 4].

First and foremost, the Internet is comprised of resources like memory, bandwidth and processor, which are limited. This limitation becomes the target of attacks.

Internet lacks central authority to control all networks that are part of it. Thus, the security policy cannot be centralized. It is interdependent. Even though the network within an organization may be defended appropriately, still it is not fully secure until its outer network also takes the required measures. There is no authority to check the authentication and integrity of the addresses thus making traceability difficult. This encourages IP spoofing.

Moreover, a large number of DDoS tools are readily available online which facilitates the attackers and makes their job easy.

IV. Changing Motives and Trends of DDoS

Over the years, DDoS has evolved from packet throwing scripts targeting individual computer to operations using millions of machines via botnets/zombies. The frequency of such attacks has also increased manifold. DDoS first appeared in the mid '90s as a way of focusing attack traffic on a single site. In 1996, Panix's computers were flooded by an average of 150 SYN packets per second (pps) [7]. Early DDoS tools like Trinoo, emerged from the underground using customer protocols to communicate in late 90's. The attacks were from multiple sources towards a single site. Smurf attack was used to spoof the IP address.

The year 2000 and 2001 saw DDoS attacks compromising windows systems. Worms like code red and Trojan horse programs sent via email (Love Letter, Anna Kournikova) exploited known vulnerabilities and rendered a large number of systems out of service. Attacks began to be financially motivated. Amazon, Buy.com, CNN and eBay suffered DDoS attack. Yahoo! also became the victim and was rendered inaccessible for 3 hours. It is estimated that it suffered a loss of business and advertising revenue that amounted to $500,000. During DDoS attack CNN.com's users went down to below 5% of normal volume [9].

Fastest cyber attack in history was observed on Jan 2003. DDoS became a potent weapon against sites with limited resources to defend their sites or pay over usage fees from attack-related bandwidth consumption. EBay was taken offline by a DDoS attack involving 20,000 computers, causing damage of at least $5,000 [8]. After mid-2000s, DDoS tools became widely adopted by hackers and motive shifted from financial gains to being politically motivated. Government websites started being affected. In 2008, the website for Georgian President Mikheil Saakashvili was hit with a DDoS attack. In 2009, a crowd sourced, PHP-scripted DDoS attack took down several pro-Ahmadinejad websites during the protests of 2009 Iranian election, demonstrated its potential use in political activism.

Arbor Networks, in its 2009 Worldwide Infrastructure Security Report, states that the size of the largest DDoS attacks had increased steadily, from 400 Mbps in 2002 to 49 Gbps. This report also identified a shift away from purely Bandwidth Attacks to Application Attacks. The arbor's report states that Network Attacks were 45% of total DDoS attacks and Application Attacks were 49%. Different forms of DDoS emerged in this decade. The recruit participants, known as *Volunteer DDoS,* springed up. Revenge became the cause/motive of DDoS. There was an increasing concern about smaller attacks that relied less on bandwidth.

During 2010, there was increase in large bps/pps attacks year on year (YOY). From 2009-2010, 319% increase in the number of monitored attacks was greater than 10 Gbps. Figure1 shows this trend in the size of DDoS attack over the decade. Over 102% increase YOY in attack size shows the resurgence of brute force and volumetric attack techniques through the decade.

An extremely important finding in Arbor's 2011 Worldwide Infrastructure Security Report, is the use of cyber attacks to make a political and social statement. 35 % of attacks reported were motivated politically. The largest DDoS attack reported during the survey period was 60 Gbps, in contrast with the 100 Gbps attack reported the previous year. This suggested that the trend was moving towards *smaller but more frequent attacks intended at the application layer* [10].

The most infamous cyber-gang, Anonymous, hit government and finance companies like Visa and MasterCard worldwide. Operation Payback, in September 2011, is a series of DDoS attacks organized by members of Anonymous against a number of major entertainment websites, including the Recording Industry Association of America and the Motion Picture Association of America.

As per Arbor's report, in the Year 2012, Advanced Persistent Threats (APTs) became a top concern for the service providers. Size of volumetric attacks was plateaued. Attack on application layer evolved. Multi-vector attacks (combinations of attack methods designed to pass through the defense system an organization has in place) made an appearance. The attackers moved on to more sophisticated, diverse and frequent DDoS attacks to achieve their goals.

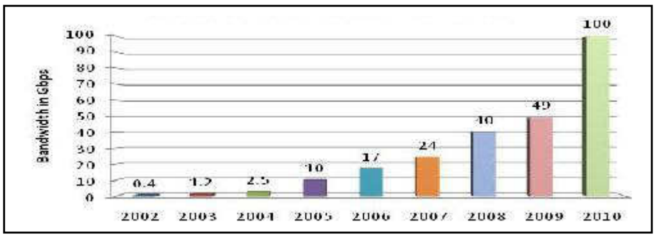


Fig.1. DDoS Attack size over time (source: Arbor Networks) [38]

According to Prolexic report for the year 2012, the average bits per second, or bandwidth, of a typical DDoS attack increased dramatically- by more than 400%, the average packets per second also increased an enormous 710%. Table I gives the major DDoS attacks experienced during 2012.

Year 2013 experienced a tremendous increase in the size of DDoS attacks. As per Prolexic quarterly reports, attack methodologies changed significantly, shifting towards amplification and reflection attacks. This led to launch of powerful and well orchestrated attacks with the help of fewer resources.

NSFOCUS report 2013 revealed that "hackernomics" was "driving an overarching DDoS trend of *smaller, shorter and repeated* attacks".

The motives included extortion, political activism, and distraction from information theft and, for hackers, testing and showing their skills. There was an increase in the total number of DDoS attacks by 33% (approx.) when compared with previous years. Also infrastructure attacks and application attacks saw the rise by 23% and 79 % (approx.). Figure 2 shows the percent growth in Attacks from 2011 to 2012 and 2012 to 2013. The data given is consolidated from Prolexic Quarterly reports for the specified years.

In the first half of 2013, of the 90 major DDoS attacks that occurred worldwide, 39 (43 %) targeted banks. This was a result of the Operation Ababil campaign. Enterprises and Government were assaulted in 19 (21 %) and 26 (29 %) major DDoS events, respectively. Internet Service Providers (ISPs) and Non-profit organizations (NPOs) fell victim to these attacks.

According to NSFOCUS mid-year DDoS Threat Report 2013, 93.2% of attacks were less than 30 minutes in duration. A total of 79.8% of all attacks were 50Mbps or less. 63.6% of victims were attacked more than once. Table II shows the major DDoS attacks through the year 2013.

Between November 30, 2013, and February 27, 2014, 81% of all recorded network attacks used two or more attack methods [16]. Also, it was observed that attackers distracted defenders by attacks, to look for vulnerabilities to exploit.

TABLE I. MAJOR DDOS ATTACKS EXPERIENCED DURING 2012

Date	Details
Sept. '12	• Wordpress sites, using expired plugin, were compromised to attack another website to bring it offline. • Application layer attacks [13]. • DNS Amplification
Dec. 10, '12	• 10 major banks of US including JP Morgan, Bank of America, were attacked by hacktivists group named Iz ad-Din al-Qassam cyber fighters. • Steal intellectual property. • DNS Amplification
Dec. 24, '12	• $900,000 was netted out of account of company ascent builders by launching a DDoS attack against its bank. • Zeus Trojan variant.

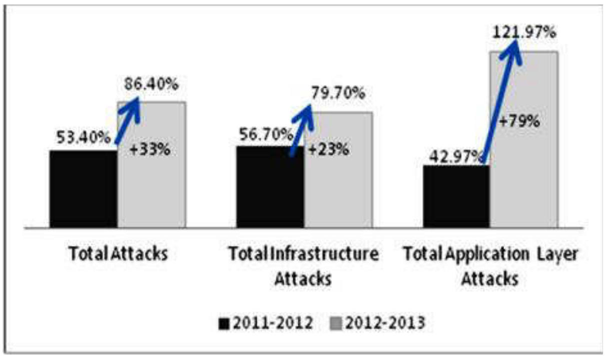


Fig.2. Percent growth in DDoS Attacks for year 2011- 2012 and 2012 - 2013 (inferred from: Prolexic and Arbor [27, 36, 37])

TABLE II. MAJOR DDOS ATTACKS EXPERIENCED DURING 2013

Date	Peak attack bandwidth	Industry	Attack duration	Method Used
Mar 6, '13	130 Gbps	Finance, Banking, Insurance	14.6 days	UDP, SYN and DNS Floods
Mar 6, '13	309 Gbps	Anti-spam organization Spamhaus- biggest cyber attack till date.	Around 10 days	DNS floods
Apr 4, '13	144 million pps	Finance, Banking, Insurance	17.5 hrs	UDP and SYN Floods
Sep 6, '13	111 Gbps	Media and Entertainment	20.6 hrs	Get and SYN floods DNS and UDP fragments
Dec 13, '13	179 Gbps	Media and Entertainment	5.4 hrs	UDP, SYN, NTP reflection

The year 2014 started with significant DDoS activities. Figure3, obtained from the information collected by Incapsula [19], shows the trend of attacks in the year 2013 and 2014. 2014 sees the rise of NTP amplification attacks along with the increased use of multi-vector tactics.

On Feb 2014, attack larger by 30 % as compared to largest attack of 2013 was faced by Cloudfare's customer. It had to endure the network traffic of 400Gbps. On February 20, Internet registration firm, Namecheap, was overwhelmed temporarily by simultaneous attack on 300 of the websites registered with it. Also in the same month, Bitly, an organization which creates shortened addresses for websites like Twitter, was also knocked out briefly. According to Incapsula, “the total *amount* of DDoS traffic in 2014 has jumped to 240 %” [18].

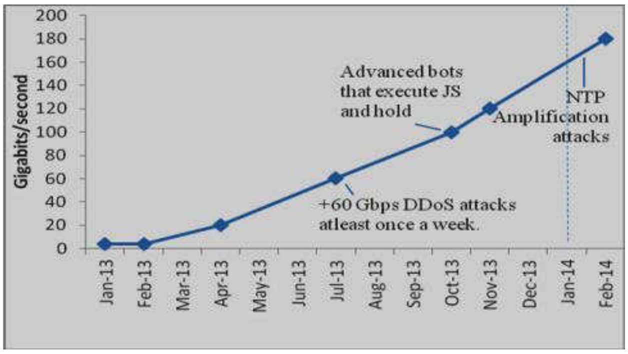


Fig. 3. 2013: Increase in peak attack volumes and bot sophistication.

2014: Addition of NTP DDoS Attack [19]. (Source: Incapsula)

Moreover, 21 percent of the bots attacked more than 50 targets a month. Of these, more than 80 percent were multi-vector, a common tactic that criminals use to ensure their aims are met. Table III shows the major DDoS attacks experienced in 2014 till March.

TABLE III. MAJOR DDOS ATTACKS EXPERIENCED IN 2014 TILL MARCH

Date	**Details**	**Motives/Technology**
Jan, ‘14	Gaming sites attacked [10]	NTP amplification
Feb, ‘14	Meetup.com was hammered after it refused to pay $300 as ransom	
Feb, ‘14	400 Gbps attack hit Cloudfare’s customer	NTP amplification
Mar, ‘14	162000 WordPress installations were hijacked to attack another site	Application layer attacks. Less traffic more effect.
Mar, ‘14	An attack launched by pro-Russian Ukrainians against NATO domains. These included Primary NATO domain a Co-Op cyber Defense Center and parliamentary assembly	Politically motivated

V. CONCLUSION

As we can observe, the trend in Denial of Service attacks changed significantly over the years. The attackers became more knowledgeable, thus increasing the complexity of the attacks. The present era is about Advanced Persistent Threats (APT). Traditional DDoS attacks were relatively simple but brutal, directly overwhelming the targets with requests by controlling a huge network of Zombies. By the year 2013-14, attackers migrated to much sophisticated attack techniques. The motive of most attacks is to disrupt, not to destroy, and in order to put forward their demands. Attack technologies have the capability of posing greater harm. DDoS attacks have been used as veil to carry out APT attacks or for other malicious purposes. DNS and NTP amplification attacks remain a very real and serious threat to the Internet at large. These can far exceed the bandwidth levels available till date. Figure 4 gives the trend line which sums up the evolution of DDoS attacks and shows the trends and motives of attackers since 1996.

With an ever increasing threat from DDoS having potential not only to cause enormous financial setback, but also influence politically, IT organizations need to re-think their security strategies and enhance their defenses to protect themselves from the malicious agents.

Contributions of paper:

- Provides chronological information related to DDoS attacks and its changing trends and motives.
- Provides a thrust to enhance knowledge and the understanding of DDoS.

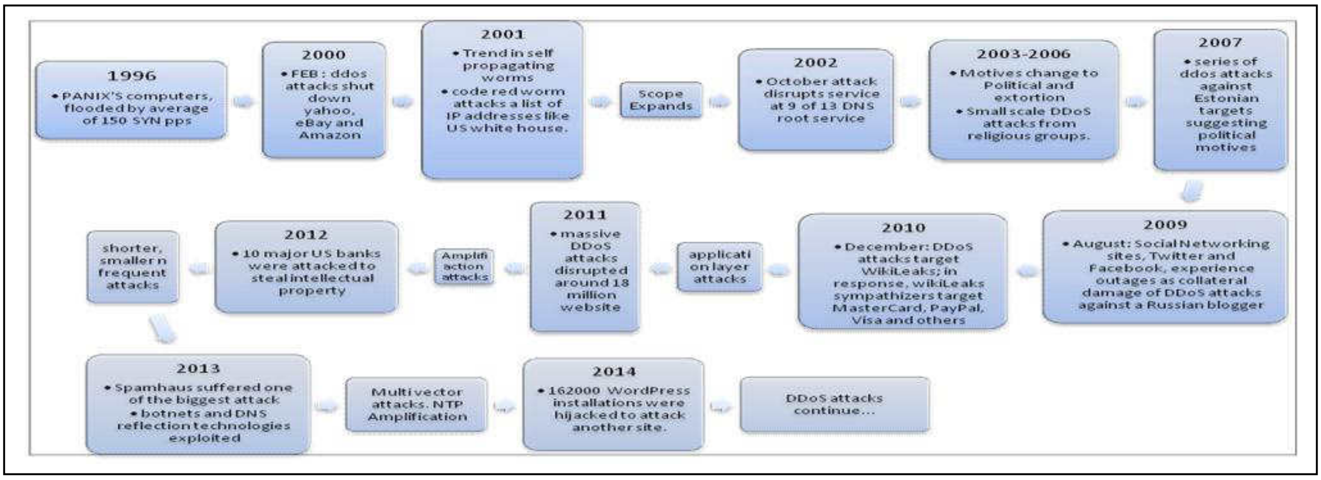


Fig. 4. DDoS Trendline: 1996- March, 2014

REFERENCES

[1] Radware.com, Radware DDoS Survival handbook, 2013.

[2] J. Mirkovic, P. Reiher, " A taxonomy of DDoS attack and DDoS defense mechanisms," ACM SIGCOMM Computer Communications Review, Volume 34, Number 2: April 2004,pp:198-202.

[3] B. Harris, E. Konikoff, P. Petersen, "Breaking the DDoS chain," CMU-ISR-MITS-2, August 2013,pp:178-184.

[4] A. Mitrokotsa, C. Dougligeris, "DDoS attacks and defense mechanisms: classification and state-of-the-art," Computer Networks44 (2004), 2003, pp:643-646,

[5] L. Garber, "Denial of Service attacks rip the internet," Computer journal of IEEE, vol. 33, no. 4,2000, pp: 12-17 .

[6] E. Messmer, Network World, "Massive denial-of-service attacks pick up stream, new nefarious techniques," http:// www. Networld. com/news/2014/011414-denial-of-service-277699. html

[7] L. Constantin," DDoS attacks against US banks peaked at 60 Gbps," IDG news service, at Networkworld. com, 2012,pp:345-349.

[8] A.M. Freed, "DDoS attacks tripled in bandwidth magnitude 2013," at tripwire.com, 2013.

[9] S. Gibson, "The strange tale of the denial of service attacks against grc.com," http://grc.com/dos/grcdos.htm, 2007.

[10] Steve Ragan, "DDoS traffic up 240% in 2014," http://www.csonline.com/article/2137053/malware-cybercrime/ report-ddos-traffic-up-240-in-2014. html, 2014.

[11] Imperva, 2013-2014, " DDoS Threat Landscape Report".

[12] A. Peter, "DDoS cyber attacks get bigger, smarter, more damaging," at uk.reuters.com, 2014.

[13] J. NAZARIO, "Politically motivated denial of service attacks," Arbor Networks, United States.

[14] D. Anstee, " DDoS attack trends through 2010, infrastructure security report and ATLAS initiatives," Arbor Networks, United States.

[15] E. Zuckerman, H. Roberts, R. McGrady, J. York, J. Palfrey, " Distributed denial of service attacks against independent media and human rights sites," The Berkman Center for Internet and Society at Harvard University, December 2010,pp:987-991.

[16] M. Sachdeva, G. Singh, K. Kumar, K. Singh, " DDoS incidents and their impact: a review," The International Arab Journal of Information Technology, Vol. 7, No. 1, January 2010.

[17] D. Mahajan, M. Sachdeva, "DDoS attack prevention and mitigation techniques- a review," International Journal of Computer Applications (0975 – 8887) Volume 67– No.19, April 2013.

[18] M. Sachdeva, G. Singh, K. Kumar, K. Singh, " A distributed approach to defend web service from DDoS attacks," International Journal of Computer Science and Security (IJCSS), Volume (5): Issue (3): 2011

[19] ITworld.com, "CERT Hit by DDoS Attack for a Third Day," http://security.itworld.com/4339/IDG010524CERT2/pfindex.html, 2007.

[20] Cert. Gov. au, " Managing denial of service (DoS) attacks," Summary Report for CIOs and CSOs, December 2009.

[21] P. Vixie, G. Sneeringer, M. Schleifer, "Events of 21-Oct-2002," Event Report, ISC/UMD/Cogent, November 24, 2002.

[22] M. Handley," Internet denial of service considerations," Internet draft, IETF, 2004.

[23] K.J. Houle, and G.M. Weaver (2001), "Trends in denial of service attack technology," Technical Report v1.0. CERT and CERT coordination center, CMU, Pittsburgh, PA.

[24] A.D. Householder, A. Manion, L. Peasant, " Managing the threat of denial of service attacks," SEI, CMU, Pittsburg, PA.

[25] S. Specht, R. Lee, " Taxonomies of distributed denial of service networks, attacks, tool, and countermeasures," Technical Report CE-L2003-03, May 2003,pp:675-680.

[26] M. Bhutan, H. Kashyap, D. Bhattacharyya, J. Kalita, "Detecting distributed denial of service attacks: methods, tools and future directions," 2012.

[27] B. Kleyman, "Preventing a Botnet Attack on Your Data Center," at datacenterknowledge.com,2013,pp:431-435.

[28] C.V. Zhou, C. Leckie, S. Karunasekera, " A survey of coordinated attacks and collaborative intrusion detection," Computers and Security ,2010, pp: 124-144.

Review of QoS Based Scheduling and Congestion Control in Mobile WiMAX

Goldendeep Kaur
Dept. of Computer Science and Engineering
Guru Nanak Dev University Regional Campus,
Jalandhar, Punjab
goldenchugh@gmail.com

Jyoteesh Malhotra
Guru Nanak Dev University Regional Campus,
Jalandhar, Punjab

***Abstract*—Mobile WiMAX (Worldwide Interoperability for Microwave Access) is being treated as the most promising and potential broadband wireless technology. And the popularity rate has been surging to newer heights as the knowledge backed service era unfolds steadily. Especially Mobile WiMAX is being projected as a real and strategic boon for developing counties such as India due to its wireless coverage acreage is phenomenally high. The IEEE 802.16e standard based Mobile WiMAX system will be investigated for the purpose of Quality of Service provisioning. It is necessary to provide Quality of Service (QoS) guaranteed with different characteristics. As a possible solution the scheduling algorithms and congestion control will be taken into main consideration and the present well known algorithms will be described .In this paper, we have highlighted the following critical issues for Mobile WiMAX technologies.**

Keywords*—Mobile WiMAX, QoS, MAC, Scheduling, TCP Congestion Control.

I. INTRODUCTION

"Mobile WiMAX" refers to a rapidly growing broadband wireless access solution built upon the IEEE 802.16e-2005 air interface standard. It is equally applicable to fixed, portable and mobile applications. The Mobile WiMAX air interface utilizes Orthogonal Frequency Division Multiple Access (OFDMA) for improved multipath performance in non-line-of-sight (NLOS) environments and high flexibility in allocating resources to users with different data rate requirements. The fundamental premise of the IEEE 802.16e media access control (MAC) architecture is QoS. Mobile WiMAX QoS features enable operators to optimize network performance depending on the service type (e.g., voice, video, and gaming) and the user's service level. Mobile WiMAX is expected to deliver significant improvements over Fixed WiMAX which makes it even more attractive for fixed deployments. Since the packet scheduling theme is mainly done in the MAC layer of the OSI protocol stack, the main task will be to present a brief background overview on the scheduling procedure related MAC entities; as it will also briefly explain the WiMAX Physical (PHY) Layer interrelated functionalities which are involved in the scheduling process.

A. *IEEE 802.16 PHY Layer*

The WiMAX physical layer is based on orthogonal frequency division multiplexing. OFDM is the transmission scheme of choice to enable high-speed data, Big files, video, Deficit Round Robin (DRR), and multimedia file and is used by a variety of commercial broadband systems.

There are three types of subcarriers:

- Null Subcarrier: used as guard band or DC carriers.
- Pilot Subcarrier: used for channel estimation and channel tracking.
- Data Subcarrier: carrying the data information.

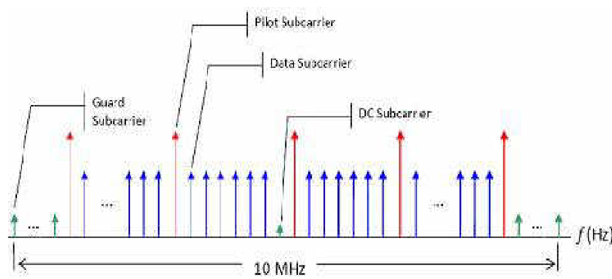


Fig.1. OFDM subcarrier structure

Fundamentally, OFDMA is OFDM with the application of Subchannelization and Time Division Multiple Access (TDMA). Figure 1 could be modified to show Subchannelization for OFDMA, as illustrated in Figure 2:

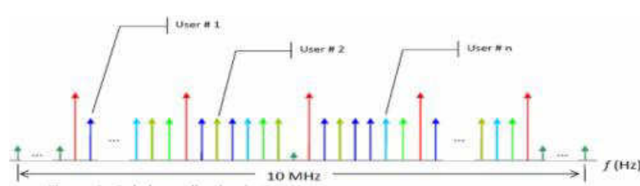


Fig.2.Subchannelization in OFDM

Note, that different colours mean different users. Unlike OFDM, here in OFDMA data streams from different users are multiplexed, rather than using the whole band for a single user per one symbol frame.

B. IEEE 802.16e MAC LAYER

The main functions at the MAC layer that have significant impact on the performance of WiMAX systems are: scheduling, automatic repeat request (ARQ) and hybrid ARQ (HARQ), fragmentation and packing, and packet header suppression (PHS). Among these, the scheduler is possibly the most significant and complex. If an inefficient scheduler is used, it is possible to get meaningless results. The MAC layer concentrates more on the congestion control and scheduling process, which is the process of requesting and providing bandwidth to users according to the Quality ofService (QoS) .

These QOS service types can be classified as:

- UGS: This service is designed to support real-time uplink service flows that transport fixed size data packets on a periodic basis.
- ertPS: This service is designed to support real-time uplink service flows that transport variable size data packets on a periodic basis.
- rtPS: This service class is for variable bit rate (VBR) real time traffic such as MPEG.
- nrtPS: This service class is for non-real-time VBR traffic with no delay guarantee. Only minimum rate is guaranteed.
- BE: Most of data traffic falls into this category. This service class guarantees neither delay nor throughput.

The remainder section of this paper is organized as follows: Section 2 describes related work; Section 3 describes the various scheduling algorithms studied by various authors; Section 4 describes congestion algorithms being studied; Section 5 describes the open issues; and finally section 6 concludes the paper.

II. RELATED WORK

In [1] N. Abu Ali analyze several scheduling algorithms to meet QoS requirements especially for real time transmission with the delay and delay jitter constraints. The proposed algorithm is hierarchical in which eight queues are defined according to their direction (uplink and downlink) and the service classes (ertPS, rtPS, nrtPS, and BE). In [2], T R Andel analyze Scheduling algorithms for rtPS Qos . He proposed three UL scheduling algorithms that provide a balancing scheme in satisfying the QoS constraints in WiMAX. Simulation results show that the three proposed algorithms (AFQLA, AFSMD, and AFDW) provide a reasonable network satisfaction with different approaches where the aim of designing and implementing those scheduling algorithms is that to maintain the QoS constraints, not to provide the highest possible throughput or achieve minimum delay in serving the rtPS SSs. In [3], J Chen analyze Modified Weighted Fair Scheduling Algorithm. MWFQ algorithm is only able to provide good performance for RTP(VoIP) protocol, because this algorithm is made based on the protocol type, with its prioritization order is RTP, CBR and then TCP. For throughput, MWFQ algorithm provides good performances for UGS, RTPS and eRTPS. For packet loss, MWFQ algorithm provides smaller percentage of packet loss.

In [4], C Wand presented various Congestion control algorithm. A new TCP hybrid congestion control mechanism is proposed as the aims for improving TCP performance such as loss, optimize utilization congestion window size and maximize throughput over wireless network especially in mobile WiMAX. In [5], A fair Connection Admission Control Scheme has been analyzed. FCAC is proposed to implement service fairness for users in 802.16e networks, which considers a common scenario that users are limited to the fixed amount of bandwidth (called as “legal bandwidth” in this paper) by ISP. In the FCAC scheme, the user’s legal bandwidth and the system’s available bandwidth are used as the admission control criterion.In [6], Bandwidth and delay guaranteed call admission control scheme has been analyzed. The key issue is to provide an efficient call admission control scheme that satisfies both bandwidth and delay guarantee to the admitted connections. In [7], A distributed Scheduling Algorithm has been analyzed. A scheduling algorithm for mobile WiMAX networks that provides more chances to delay sensitive traffic to be serviced while still maintaining fairness for various traffic types. In this paper we have discussed various scheduling and congestion control algorithms that have been analyzed by various authors.

III. SCHEDULING ALGORITHMS

Traditionally, the First Come First Served (FCFS) scheme had been used for packet scheduling. Packets coming from all the input links were enqueued into a First In First Out (FIFO) memory stack, then they were dequeued one by one on to the output link. In the present time different queues are specified to non similar packets forachieving packet classification. In this case scheduling should be done.

SCHEDULING ALGORITHM AIMS**:** The main aims behind the concept of packet scheduling could be simply defined by four points:

- The first and foremost goal of a scheduling algorithm is to be able to share the total system bandwidth fairly.
- The algorithm should be able to guarantee the minimum bandwidth per SS. This consequently leads to the separation of different demanding SSs.
- To be capable of meeting packet drop guarantees.

- To be able to assure latency guarantees.

A. Various Scheduling Algorithms

1. Round Robin Scheduling:

   Round Robin as a scheduling algorithm is considered the most basic and the least complex scheduling algorithm. Basically the algorithm services the backlogged queues in a round robin fashion.

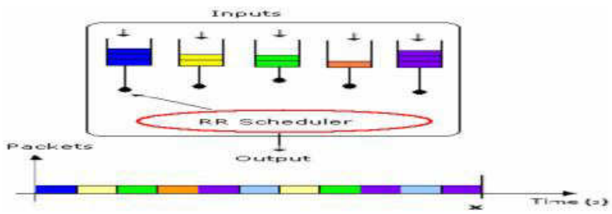


Fig.3. Round Robin Scheduling

2. Weighted Round Robin:

   WRR assigns a weight to each queue. The weight of an individual queue is equal to the relative share of the available system bandwidth. This means that, the number of packets dequeued from a queue varies according to the weight assigned to that queue.

3. Deficit Round Robin:

   In DRR scheduling, every queue is accompanied by a deficit counter which is initially set to the quantum of the queue. A quantum is a configurable amount of credits (in bits/bytes) given to a queue whenever it is served. Quantum should represent the idle amount of bits/bytes a queue may require.

4. Modified Deficit Round Robin:

   The algorithm depends on the DRR scheduling fundaments to a great extent, however, in MDRR the quantum value given to the queues is based on the weight associated with them

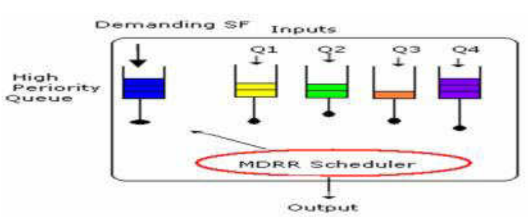


Fig.7.MDRR

Scheduling Algorithms for RTPS QOS Class

1. Alternative factor specified maximum data algorithm:

   It mainly lies on specifying a maximum data to send for each user.

2. Alternative factor with delay aware algorithm:

   It presents another alternative by prioritizing users according to their delay the priority of users who did not receive any allocation for a long time is increased.

3. Alternative factor with queue lentgh aware algorithm:

   While the AFSMD and AFDW focus on enhancing the throughput and delay respectively they do not take fairness in consideration. To consider fairness, this algorithm modifies by giving priority for each user according to his/her queue length (data to send), where the higher data size user has the higher priority.

4. Alternative factor with mix of snr and queue aware algorithm:

   This approach combines the three modifications presented in the previous subsections to present a new approach. It aims to find a trade-off between the throughput maximization of the AF and delay satisfaction of the AFQLA. This is achieved by giving priority to each user according to α of queue length (data to send) and $(1-\alpha)$ of SNR value for his assigned subchannel.

Channel Unaware Schedulers:

1. First in first out algorithm:

   This algorithm is based on first come first served basis. It forms a queue, the service type that comes first will be the first one to be served.

2. Priority based algorithm:

   In order to guarantee the QoS to different classes of service, priority-based schemes can be used in a WiMAX scheduler. For example, the priority order can be: UGS, ertPS, rtPS, nrtPS and BE, respectively. Or packets with the largest delay can be considered at the highest priority. Queue length can be also used to set the priority level, e.g., more bandwidth is allocated to connections with longer queues.

3. Delay threshold priority queing algorithm:

   Delay Threshold Priority Queuing (DTPQ) was proposed for use when both real-time and non real-time traffic are present. A simple solution would be to assign higher priority to real-time traffic but that could harm the non realtime traffic. Therefore, urgency of the real-time traffic is taken into account only when the head-of-line (HOL) packet delay exceeds a given delay threshold.

4. Earliest deadline first algorithm:

   Earliest Deadline First (EDF) is the basic algorithm for scheduler to serve the connection based on the deadline.

5. Largest weighted delay first algorthm:

   Largest Weighted Delay First (LWDF) chooses the packet with the largest delay to avoid missing its deadline**.**

6. Weighted fair queuing algorithm:

   WFQ is an approximation of General Processor Sharing (GPS). WFQ does not make the assumption of infinitesimal packet size. Basically, each connection has its own FIFO queue and the weight can be dynamically assigned for each queue. The resources are shared in proportion of the weight. For data packets in wired networks with leaky bucket, an end-to-end delay bound can be provably guaranteed.

B. Congestion Control Algorithm

The IEEE 802.16e standard provides Quality of Service (QoS) guarantees for users based on the conception of connection-oriented service. Various algorithms have been presented.

1. Fair connection admission control algorithm:

A Fair Connection Admission Control scheme (FCAC) has been proposed, in which when the BS receives a connection request from a user, first it considers whether the users remaining bandwidth is enough to admit the connection request. If not it denies the request. Otherwise it checks whether the system's available bandwidth is enough to admit the request. In general if these two conditions are satisfied the connection request can be accepted.

2. Bandwidth and delay guaranteed call admission control algorithm:

An efficient Call Admission Control (CAC) scheme for IEEE 802.16e Mobile WiMAX that satisfies both bandwidth and delay guarantee to the admitted connections has been proposed in this paper. The proposed CAC scheme provides higher priority to Handoff connections, because it is more annoying to drop an ongoing connection than blocking a newly originated connection. The BS upon receiving the request will check whether it can provide the required QoS for that connection, if the request was accepted and at the same time verifies whether the QoS of all the ongoing connections can be maintained.

3. TCP congestion control mechanism:

TCP congestion control mechanism was implementing the first of three congestion control principles such as slow-start, congestion avoidance, fast retransmits known as TCP congestion principles. Furthermore Reno are introduced which is implementing the all TCP congestion principles fast recovery after packet loss experienced but it lack in handle if two more segments are lost in the current window, fast recovery cannot retransmit all the lost packet until retransmit timeout

IV. ISSUES AND CHALLENGES

From the literature surveyed in ithe previous section there are still pros and cons of the various algorithms which are discussed below:

Table I.COMPARISON OF CHANNEL UNAWARE SCHEDULERS

SCHEDULING	PROS	CONS
FIFO	Fast and simple	Cannot meet Qos requirements.
Round Robin	Very simple	Cannot meet Qos requirements.
Weighted Round Robin	Simple; meets the throughput guarantee.	Unfair(variable packet size)
Priority	Simple; meets the delay guarantee.	Some flows may starve.
Delay Threshold Priority Queuing	Trades-off the packet loss rate of rtPS and average data throughput of nrtPS	Lower throughput.
Earliest Deadline First	Meets the delay guarantee	Non work conservative.
Largest Weighted Deadline First	Guarantees the minimum throughput.	N/A

V. OPEN ISSUES

Based on the work done by numerous authors there are still some issues which needs immediate attention. These issues are discussed below:

1. In the FCAC scheme, the user's legal bandwidth and the system's available bandwidth are used as the admission control criterion. Furthermore, in order to improve the system's fairness on bandwidth allocation to users, the bandwidth requirements of other users are also taken into account to decide whether to admit a user's connection request.

2. There are no proposed Call Admission Control schemes in literature, which satisfy both bandwidth and delay guarantees which are efficient.

3. For future studies, there are several things that can be carried out, such as some QoS parameters that are still not perfect, for example improvement on delay for CBR and TCP, throughput for nRTPS and BE**.**

VI. CONCLUSION

In this paper, we provided an extensive survey of recent scheduling proposals and congestion control algorithms for WiMAX and discussed key issues and design factors. The goals of the schedulers are basically to meet QoS guarantees for all service classes, to maximize the system goodput, to maintain the fairness, to minimize power consumption, to have as less a complexity as possible and finally to ensure the system scalability. To meet all these goals is quite challenging since achieving one may require that we have to sacrifice the others. Optimization for WiMAX scheduler is still an ongoing research topic. There are several holes to fill in, for example, polling mechanism, backoff optimization, overhead optimization and so on. Future research on scheduling should consider the use of these

characteristics. The use of Multiple Input Multiple Output with multiple antennas to increase the bandwidth makes the scheduling problem even more sophisticated. Various congestion control algorithms aims at improving performance such as loss, optimize utilization congestion window size and produce high throughput in mobile WiMAX.

REFERENCES

[1] N. Abu Ali, P. Dhrona, H. Hassanein, "A performance study of scheduling algorithms in point-to-multipoint WiMAX networks",

[2] T. R. Andel and A. Yasinsac, "A performance study of uplink scheduling algorithm,"IEEE Computer, vol.39, no.7, pp.48-54, July 2006.

[3] J. Chen, W. Jiao, H. Wang, "A Service Flow Management Strategy for IEEE 802.16 ," Proc. IEEE International Conf. on Communications, (ICC 2005), Seoul, Korea, May 16-20, 2005.

[4] J. Chen, C. Wang, C. Chang, S. Liu, J. Guo, W. Lien, J. Sum, and C. Hung, "A Scalable and TCP Friendly Congestion Control algorithms," Workshop on ns-2 (WNS2): the IP network simulator, Piza, Italy, 10 October 2006.

[5] G. Chu, D. Wang, and S. Mei., Dynamic Admission Control and Qos for 802.16 Mobile WiMAX BWA System," Proc. IEEE conf. on Communications, Circuits, and Systems, Chengdu, China, June 29-July 1, 2002.

[6] C. Cicconetti, A. Erta, L. Lenzini, and E. Mingozzi, "Performance Evaluation of the IEEE 802.16 MAC for QoS Support," IEEE Transactions on Mobile Computing, vol.6, no.1, pp.26-38, Jan. 2007.

Performance Analysis of Wireless VoIP under Pervasive Conditions

Monica Sharma
Dept. of CSE
CTIT
Jalandhar, India

Harjit Pal Singh
Dept. of ECE
CTIEMT
Jalandhar, India

Jasbinder Singh
Dept. of CE
MD College
Sri Ganganagar

Manju Bala
Dept. of CSE
CTIEMT
Jalandhar, India

***Abstract*- Internet Protocol is used as a communication platform for VoIP (Voice over Internet Protocol). Before the transmission of VoIP it is required to do signal processing for reliable & efficient performance over the computer and the communication platform.Speech quality is critically important to the update of the service. Signal quality gets deteriorated at network layer due to problems like delay, jitter or packet loss. The work presented in this paper presents the comparative analysis of speech quality for narrowband and wideband VoIP for different modulations schemes such as DBPSK, DQPSK, QAM64. The VoIP simulations are conducted for G.729A and AMR-WB speech codecs at different packet loss rates in wireless environment. The signal quality of wireless VoIP system for different modulations are analyzed and it is observed that signal quality of the wireless VoIP system decreases as the packet loss increases. The significant quality of the signal is observed for DBPSK modulation scheme.**

Keywords- Digital Signal Processing, Narrowband VoIP, Wideband VoIP, Signal Analysis.

I. INTRODUCTION

Voice transmission over Internet is known as Voice over Internet Protocol (VoIP). It is used for making voice calls with the help of Internet where the voice packet is transmitted using the internet protocol network. With the advent of VoIP it gained popularity due to its low cost as compared to the traditional Public Switched Telephone Network (PSTN). The traditional network required a dedicated line for voice communication where majority of the bandwidth is wasted whereas in case of VoIP it is transmitted over already established IP network [1]. Another advantage that makes VoIP better than PSTN is that in VoIP voice is transmitted in digital form which enables it to provide more features than PSTN [3].

VoIP involves digitization of voice streams and transmitting the digital voice as packets over the IP based packet networks like the internet. Figure1. shows the basic steps involved in the working of VoIP system. The analog voice signal is sent to the encoder that digitizes it with the help of an encoder where echo cancellation and voice activity detection algorithms are implemented on the digitized signal. The signal is compressed using various ITU-T codecs such as G.711 [4], G.729 [5] and G.723.1[6]. The codec is selected on the basis of various factors such as bandwidth availability and quality of signal.

The encoded speech is then packetized and each packet includes the headers at the various protocol layers such as RTP 12 bytes [7], UDP 8 bytes, IP 20 bytes and the payload composed of the encoded speech for certain duration [8], depending on the codec used.

Then the packets are routed to the destination through the internet. At the receiver side the voice packets are depacketized, decoded and converted back to the voice signals for the receiver. At the application layer RTP and Real-Time Control Protocol (RTCP) were designed to support real-time applications. For VoIP transmission User Datagram Protocol (UDP) is preferred over Transmission Control Protocol (TCP) as it doesn't require acknowledgement for the delivery of the packets at the receiver side in comparison to real time transmission.

The next section presents the brief description of the coders used for VoIP simulations. Performance evaluation results for various coders with different modulation schemes calculating BER is discussed in section III. IP network modelling is discussed in next section. The simulation results are shown in section V. The last section draws the conclusion and future work.

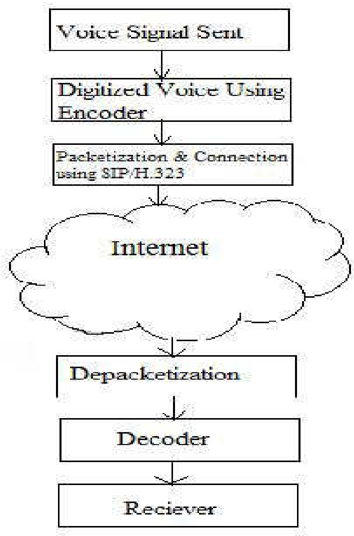


Fig. 1 Working of VoIP

II. CODEC

A. G.729A coder

The CS-CELP coder is based on a code-excited linear prediction (CELP) coding model. In this model the throat and mouth are modelled as a linear filter and voice is generated by a periodic vibration of air exciting this filter. The locally decoded signal is compared against the original signal and the coder parameters are selected such that the mean-squared weighted error between the original and reconstructed signal is minimized.

The CS-ACELP coder is designed to operate with an appropriately band limited signal sampled at 8000Hz. The input and output samples are represented using 16-bit linear PCM. The coder operates on frames of l0ms corresponding to 80 samples. For each frame, the speech signal is analysed to extract the parameters of the CELP model which include linear-prediction filter coefficients, adaptive and fixed-codebook indices and gains. At the decoder, these parameters are used to retrieve the excitation and synthesis filter parameters. The speech is reconstructed by filtering through the short term synthesis filter. After computing the reconstructed speech, it is further enhanced by post-filter [4].

B. AMR-WB Coder

Adaptive multi-rate wideband (AMR-WB) [5, 6] was standardized by 3GPP in 2000 for use in 3G mobile systems and was also adopted by ITU in 2001 as their latest wideband coder, G.722.2. It operates at 9 varying modes of 6.6 kbps up to 23.85 kbps bit rates and describes the detailed mapping from input blocks of 320 speech samples in 16 bit uniform PCM format to encoded blocks of 132, 177, 253,285, 317, 365, 397,461 and 477 bits and from encoded blocks of 132, 177, 253,285, 317, 365, 397,461 and 477 bits to output blocks of 320 reconstructed speech samples. AMR-WB offers even lower bit rate compressions as well as the ability to quickly adapt to varying compression as per the network conditions. The bandwidth is automatically conserved when network congestion is high. When the congestion returns to a normal mode, a lower-quality bit rate is restored.

III. ANALYSIS OF VoIP CODEC

The original speech signal is compressed through G.729A, AMR-NB & AMR-WB speech coders and then, the compressed speech signal is modulated to transmit over the wireless fading channels. On the receiver side, the signal is demodulated and decoded and the measurement is performed through bit error rate (BER) calculations and MOS scores obtained through PESQ & PESQ-WB measurements.

A. *BER for different Modulation schemes for VoIP Coders:*

The effect of the SNR on the VoIP speech coders has been tested for different modulation schemes such as Differential Binary Phase Shift Key (DBPSK), Differential Quadrature Phase Shift Keying (DQPSK and Quadrature amplitude modulation (QAM) with and without convolutional channel coding. The value of SNR is varied 0 to 20. The performance evaluation results for various coders with different modulation schemes are given in Fig. 2 to Fig. 10.

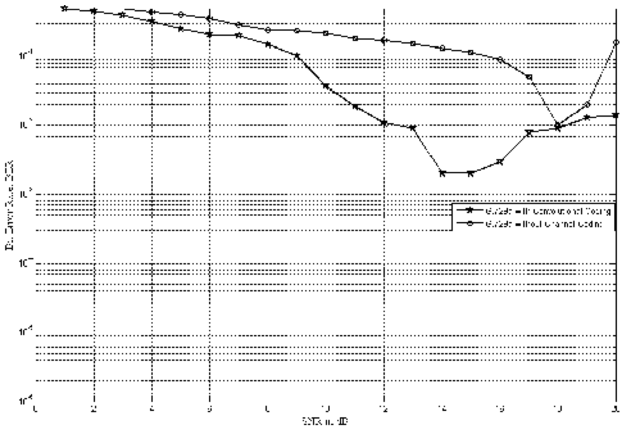


Fig. 2 DBPSK (G.729a)

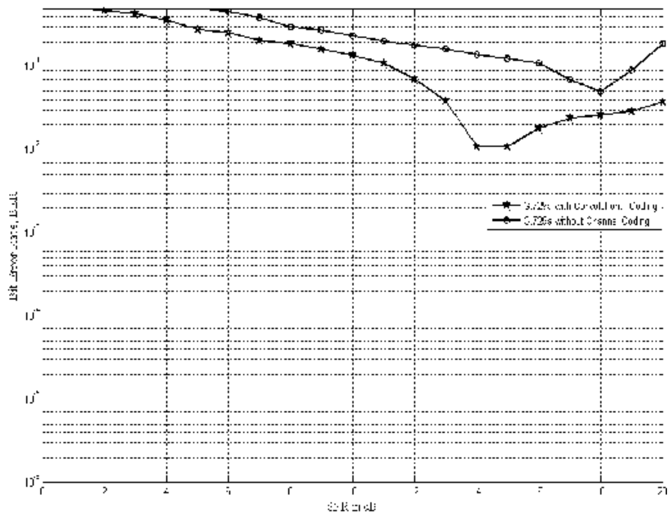


Fig. 3 DQPSK (G.729a)

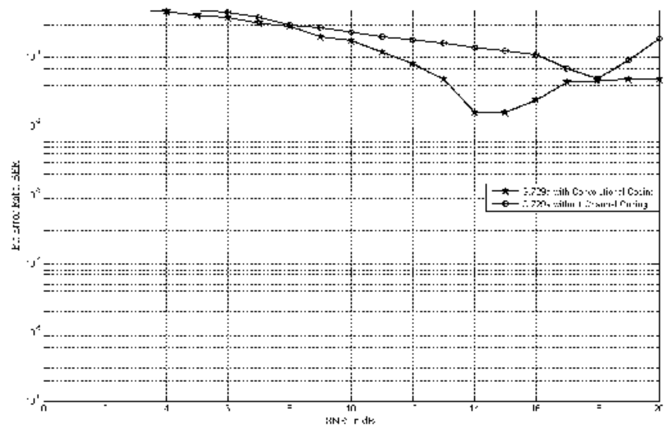


Fig. 4 QAM64 (G.729a)

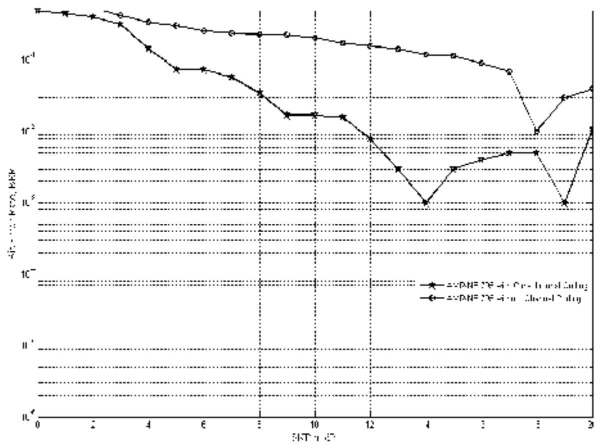


Fig. 5 DBPSK (AMR-NB)

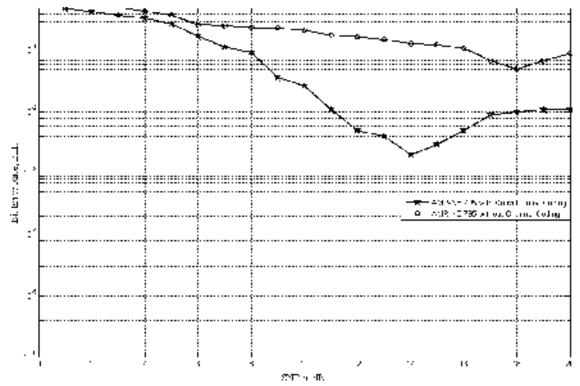


Fig. 6 DQPSK (AMR-NB)

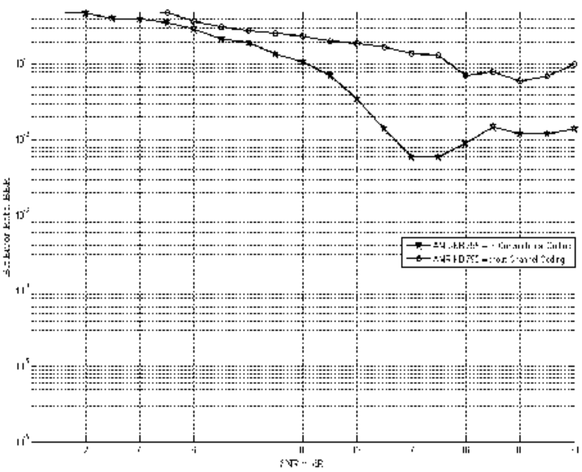


Fig. 7 QAM64 (AMR-NB)

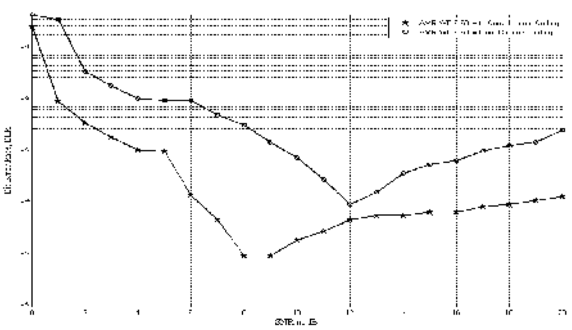


Fig. 8 DBPSK (AMR-WB 6.6 kbps)

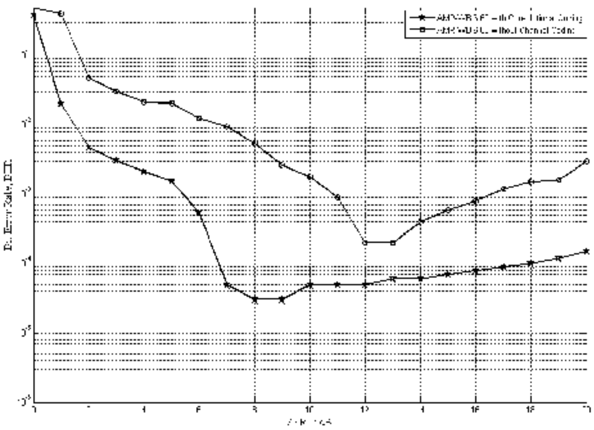


Fig. 9 DQPSK (AMR-WB 6.6 kbps)

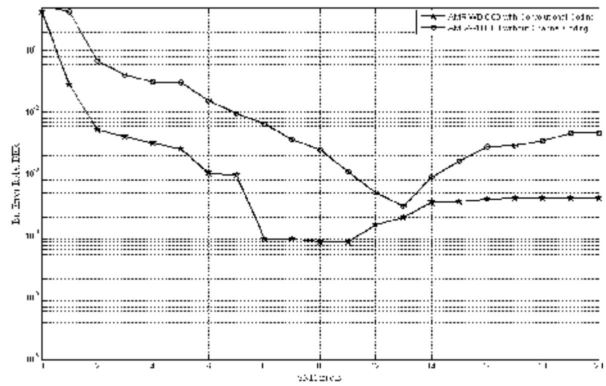


Fig. 10 QAM64 (AMR-WB 6.6 kbps)

IV. IP NETWORK MODELING

The simulation of VoIP system was performed where each packet contains one frame. Packet losses are not independent on a frame-by-frame basis, but appear in bursts. The packet loss can be approximated by Markovian loss model such as Gilbert model, as discussed by Bolot in [7]. Thus simulation of IP network was performed by using a 2-state Gilbert-Elliot Model. The model has two states reflecting whether the previous packet is received or lost. The state “0” represents that a packet being correctly received and state “1” represents that a packet being lost. The Gilbert-Elliot model is shown in Fig 11.

Let p be the transition probability for the network model to drop a packet given that the previous packet is delivered i.e. the probability for network model to go from state “0” to state “1”. Let q is the probability for the network model to drop a packet given that the previous packet is dropped, i.e. the probability for the network model to stay in state “1”. This probability is also known as the conditional loss probability.

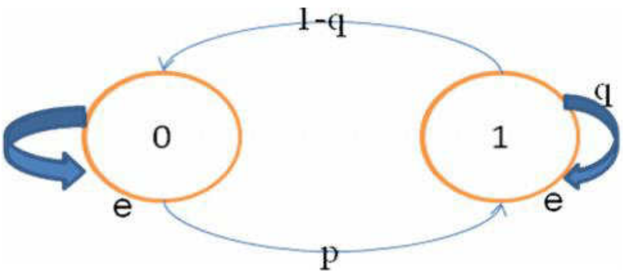


Fig.11 Gilbert-Elliot Model

Let $p0$ and $p1$ denote the probability of the network model to be in state 0 and 1.The probability for a packet to be dropped regardless whether the previous packet is delivered or dropped i.e. the unconditional loss probability is exactly the probability for the network model to be in state 1 ($p1$).

$$p0 = \frac{q}{p+q} \; ; \quad p1 = \frac{p}{p+q} \tag{1}$$

The transition matrix is given as

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \tag{2}$$

A. *VoIP SIMULATIONS*

The VoIP simulations were carried out for making comparison between narrowband speech coding and wideband speech coding algorithms. The speech samples for simulations were taken from [8]. The network impairments were introduced into VoIP frames with the modelling of the IP network through above discussed Gilbert-Elliot model. The speech signal of VoIP system is degraded at different packet loss rates (PLR). The simulations were conducted at different packet loss rates as discussed in Table.1 and then the spectral analysis of the degraded VoIP speech signal is then carried using MATLAB and MOS (Mean Opinion Score) is found out as discussed in table 2-4. The simulation environment is presented in Fig. 15.

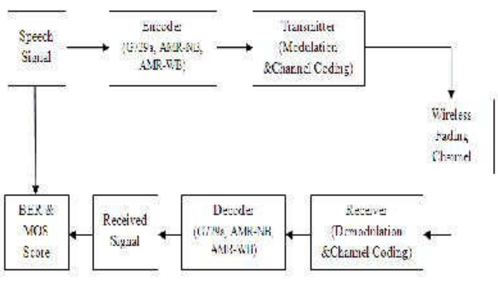


Fig. 12 Simulation Environment

Table I. Simulated Loss Rates

PLR (%)	p	q	e
3	0.97	0.03	0.005
5	0.95	0.05	0.005
10	0.90	0.10	0.005
15	0.85	0.15	0.005
20	0.80	0.20	0.005

Table II. VoIP Simulation Result for G.729A

PLR(%)	**MOS Scores**		
	DBPSK	**DQPSK**	**QAM64**
3	**2.44**	**2.30**	**2.05**
5	**2.35**	**2.10**	**2.00**
10	**2.15**	**2.02**	**1.95**
15	**2.09**	**1.95**	**1.85**
20	**1.98**	**1.80**	**1.70**

Table III. VoIP Simulation Result for AMR-NB

PLR(%)	**MOS Scores**		
	DBPSK	**DQPSK**	**QAM64**
3	**2.52**	**2.38**	**2.15**
5	**2.44**	**2.22**	**2.09**
10	**2.29**	**2.11**	**1.97**
15	**2.15**	**2.01**	**1.80**
20	**2.07**	**1.92**	**1.67**

Table 4 VoIP Simulation Result for AMR-WB

PLR(%)	**MOS Scores**		
	DBPSK	**DQPSK**	**QAM64**
3	**3.52**	**3.25**	**2.98**
5	**3.44**	**3.11**	**2.83**
10	**3.30**	**3.01**	**2.68**
15	**3.10**	**2.90**	**2.44**
20	**3.05**	**2.78**	**2.29**

B. *Discussion*

The voice quality assessment has been performed through on the decoded speech signal, which is degraded with fading effect during the transmission through wireless channels. The results show that the wireless system works well with convolutional channel coding along with various narrowband and wideband speech codecs. The BER obtained for G.729A & AMR-NB (7.95 kbps) with DBPSK is 10^{-3}, with DQPSK & QAM64 is 10^{-2} at SNR= 14 dB with convolutional coding. For wideband speech coders, the measured BER is 10^{-5} for DBPSK and 10^{-4} for DQPSK & QAM64. It is observed from the performance evaluation results that the DBPSK modulation is better for both narrowband and wideband VoIP speech coders, since as the modulation rate increases, the performance decreases. But in comparison of the narrowband G.729A & AMR-NB and wideband AMR-WB coders, the wideband coders works well in the wireless fading scenarios.

V. CONCLUSION

The performance of the wireless VoIP was analyzed for different modulations schemes under pervasive environment. The narrowband and wideband VoIP coders were tested in wireless fading environment with DQPSK, DBPSK & QAM64 modulation schemes and it was observed that the VoIP coders were best with DBPSK modulation schemes in wireless environment. Further, the performance of the wireless VoIP was analyzed through Gilbert-Elliot model. It was observed from the VoIP simulation results that DBPSK modulation worked well since MOS scores were best in case of DBPSK as compared to DQPSK and QAM64 for both narrowband and wideband coders. In future, the study would be extended to improve the signal quality of wireless VoIP system

REFERENCES

[1] J.Walter Goralski and C.Matthew Kolon, IP Telephony.McGraw-Hill, 1st Edition (2000).

[2] H P Singh, S Singh, J Singh, S A Khan, "VoIP: State of art for global connectivity—A critical review", Journal of Network and Computer Applications, Vol.37, No.1, 2014, pp:365-379

[3] U.Varshney, A.Snow,M.McGivern and C.Howard," Voice over IP", Communications of the ACM, January 2002.

[4] ITU, ITU-T G.729: CS-ACELP speech coding at 8 kbit/s, ITU 1998.

[5] 3GPP TS 26.171," Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; General description", 2012

[6] B.Bessette, R.Salami, R.Lefebvre, M.Jelinek and co-authors," The adaptive multirate wideband speech codec (AMR-WB)", IEEE transactions on speech and audio processing, vol. 10, No. 8, Nov 2002

[7] J.C.Bolot," End-to-end frame delay and loss behavior in the internet", in Proc. ACM SIGCOMM, Sept. 1993, 99.289-298

[8] Open Speech Repository.

http:// www.voiptroubleshooter.com/open_speech/.

[9] Johan D.Proakis, Dimitris G. Manolakis " Digital Signal Processing :Principles, algorithm and applications, Third edition" june2000.

[10] S Singh, H P Singh, J Singh, "Spectral Analysis of Speech Quality in VoIP for G.729A and AMR-WB Speech Coders", in Proceedings of 2nd IEEE International Conference on Computational Intelligence, Communication Systems and Networks, 2010, pp:182-186

A Survey on Wireless Sensor Networks

Gagandeep Kaur
Electronic Communication and Engineering
Baba Farid Collage of Engineering, Bathinda
Bathinda, India
gagansabo@gmail.com

Poonam Arora
Electronic Communication and Engineering
Baba Farid Collage of Engineering, Bathinda
Sirsa, India
poonambfcet@gmail.com

***Abstract*—Recent advancement in wireless communication and electronics has enabled the development of low cost sensor networks. The spread use of wireless sensor networks (WSNs) in a variety of fields, ranging from the military to environmental protection, has drawn the attention of researchers to this type of networks. In this paper, we focus on the problem of designing a energy efficient for wireless sensor networks. In large wireless sensor networks, low energy consumption is a major challenge. Hence, energy conservation takes on additional importance. Many techniques are proposed for energy saving, Clustering is one of them. In this work, we have to reduced sensor energy conservation and enhance the network throughput and network lifetime.**

Keywords—Wireless sensor networks, military application, environmental protection, Network throughput, Network lifetime.

I. INTRODUCTION

The creation of small devices capable of sensing data from its surroundings and transmitting them through the wireless medium has allowed many interesting applications to be implemented. Networks composed by these devices are called wireless sensor networks (WSNs), and their organization in networks has enabled applications to be ubiquitous. Some recent examples from the literature can be seen in several fields. In the medical field, WSNs can be used to monitor patient conditions at home, saving time and making it easier to take emergency actions. In industry, WSNs enhance the efficiency of processes by allowing timely fault correction [1].

A wireless sensor networks a collection of nodes are organized in cooperative network. These nodes are the sensor nodes which is communicate over the wireless medium. The wireless medium may either of radio frequencies, infrared or any other medium, of course, having no wired connection. These nodes are deployed in a random fashion and they can communicate among themselves to make an ad-hoc network [2]. If the node is not able to communicate with other through direct link, i.e. they are out of coverage area of each other, the data can be send to the other node by using the nodes in between them. This property is referred as multi-hoping [2]. Wireless sensor networks have the following characteristics [3] :- 1)**.**All sensor nodes use the direct transmission or multi-hop transmission to communicate with the base station because all sensor nodes are immobile. 2).Sensor nodes sense conditions at different locations at a fixed rate and always have data to send to the base station. 3).Sensor nodes can revise the transmission power of wireless transmitter according to the distance. 4).The sensor nodes are organized into a group is called cluster. Cluster head perform data aggregation and BS receives compressed data. 5).The lifetime of wireless sensor network is the total amount of time before the first sensor node runs out of power. The wireless sensor node, being a microelectronic device, can only be equipped with a limited energy. In some application scenarios, the recharge or replacement of energy resources might be impossible. Sensor node lifetime shows a strong dependence on battery lifetime. If the sensor nodes may get die due to limited battery then the chances of network breakage increases and if we want to collect the data (temperature, humidity etc) of that particular area we cannot get that data. Also, in a multihop ad-hoc sensor network, each node plays the dual role of data originator and data router. The malfunctioning of a few nodes can cause significant topological changes and might require rerouting of packets and reorganization of the network. Hence, energy conservation takes on additional importance. Many techniques are proposed for energy saving, Clustering is one of them. In this technique, the clusters are formed by clustering of the grouping nodes. The cluster heads are elected periodically such that members of a cluster can communicate with their cluster heads. These cluster heads send data received from its members to a base station. The multi clustering can also be used. The cluster head should have to be rotated for the balancing of energy and then there will be equal load on every node. The energy consumption can be reduced.

II. SENSOR NETWORKS COMMUNICATION ARCHITECTURE

The sensor nodes are usually scattered in a sensor field as shown in Figure 1. Each of these scattered sensor nodes has the capabilities to collect data and route data back to the sink and the end users.

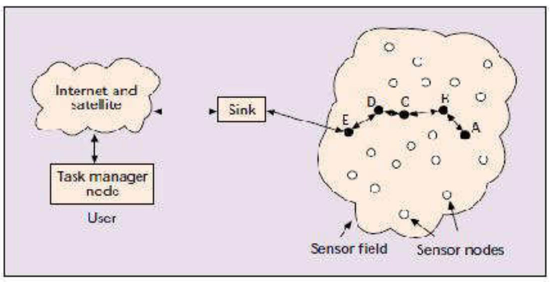


Fig.1. Wireless sensor network[1]

Data are routed back to the end user by a multihop infrastructure less architecture through the sink. The sink may communicate with the task manager node via Internet or Satellite [4].

In the present work the whole network is distributed in clusters. The cluster heads can communicate to each other by using the Destination Sequenced Distance Vector (DSDV) routing algorithm. All the members of the cluster give their data to the cluster head and cluster head forward it to the other cluster head until the data do not reach its destination. In the whole network the path between cluster heads is fixed. The path cannot be changed until all the sensor nodes do not die means their battery goes to down. In this case some intermediate nodes will die earlier than other nodes. Then the path is break down between source and destination. Here due to path breakage the packet loss increases, the packet do not reach at the destination. Packet retransmission is also increases the whole network becomes useless. A new network is configured for complete the communication. To configure the new network again become the clusters and cluster heads it takes too much time and consume energy may be the network do not complete the communication.

The CSMA/CA is the protocol which is used for the channel sensing in the wireless networks. As, the mobile ad hoc networks is the self configuring type of network in which the mobile nodes can leave or join the network when they want, In such type of network clocks of mobile devices are not synchronized which reduced the efficiency of the CSMA/CA. The overall networks throughput will be reduce and end to end delay will be increased as due to the reduction in the efficiency of the CSMA/CA packet collision will be there in the network. In the present work the sensor nodes are not synchronous to each other. The packet collision occurs due to the mismatch of timing. Again the packet loss and packets do not reach to their destination. The whole scenario of sensor network without synchronous of sensor nodes. here two nodes are the source and two destination, they followed the same path for data transfer. Cluster 1's source send the data to cluster 4's destination and cluster 3's source send the data to cluster 2's destination. Here the clock synchronous is not present between the mobile nodes.

To solve the problem of clock synchronize in wireless sensor networks, the scheme of adaptive packets has been used. In this scheme, the wireless nodes will use the adaptive packets to take the channel access. The main problem exists in such scheme is when the adaptive packets collide with each other. The wireless sensor nodes keep on waiting without any reason. This will reduce the network efficiency and reliability.

III. APPLICATIONS OF SENSOR NETWORKS

Sensor networks may consist of many different types of sensors such as seismic, low sampling rate magnetic, thermal, visual, infrared, acoustic and radar, which are able to monitor a wide variety of ambient conditions that include the following [5]:

- temperature,
- humidity,
- movement,
- lightning condition,
- pressure,
- soil makeup,
- noise levels,
- the presence or absence of certain kinds of objects,
- mechanical stress levels on attached objects, and
- the current characteristics such as speed, direction,and size of an object.
- Soil moisture
- Wind speed and direction
- Rainfall
- Sunshine
- Level of CO2

Sensor nodes can be used for continuous sensing, event detection, event ID, location sensing,and local control of actuators. The concept of micro-sensing and wireless connection of these nodes promise many new application areas. We categorize the applications into military, environment, health, home and other commercial areas.

A. *Military application*

Sensor networks can provide variety of services to military and air force like information collection, battlefield surveillance, intrusion detection and attack detection. Since sensor networks are based on the dense deployment of disposable and low-cost sensor nodes, destruction of some nodes by hostile actions does not affect a military operation as much as the destruction of a traditional sensor, which makes sensor networks concept a better approach for battlefields.

1) Intrusion detection:

Sensor network can be used as a 2-phase in Intrusion Detection System. Instead of using mines intrusion can be detected by establishing sensor network in that area. Mines are dangerous to civilians so instead sensor nodes sense the detection and alarm the army. The response to prevent intrusion is now decided by the military.

2) Enemy Tracking and target classification:

Objects moving with significant metallic content can be detected using specially designed sensors. So enemies can be tracked and civilians are ignored. This system specially helps in detecting armed soldiers and vehicles.

3) Battlefield surveillance:

Critical areas and borders can be closely monitored using sensor networks to obtain information about any enemy activity in that area. This provides quick gathering of information provides time for quick response.

4) Battlefield damage assessment:

Sensor networks can be deployed after the battle or attacks to gather information of damage assessment.

5) Detection of NBC attacks:

Sensor networks can be used as Nuclear, Biological and Chemical warning system. If any nuclear biological or chemical agents can be detected by sensors and embedded alert system can now send a warning message. It provides the military critical response time to check the situation and prevent possible attacks which can save lives of many.

6) Targeting system:

Sensors can be embedded in weapons. Exact information about the target like distance, angle can be collected and sent to the shooter. So sensors can be collaborated with the weapons for better target assessment [6].

B. Environmental applications

Some environmental applications of sensor networks include tracking the movements of birds,small animals, and insects; monitoring environmental conditions that affect crops and livestock; irrigation; macro instruments for large-scale Earth monitoring and planetary exploration; chemical/biological detection; precision agriculture; biological,Earth, and environmental monitoring in marine, soil, and atmospheric contexts; forest fire detection; meteorological or geophysical research; flood detection; bio-complexity mapping of the environment; and pollution study.

1. Forest fire detection:

Since sensor nodes may be strategically, randomly, and densely deployed in a forest, sensor nodes can relay the exact origin of the fire to the end users before the fire is spread uncontrollable. Millions of sensor nodes can be deployed and integrated using radio frequencies/optical systems.

2. Flood detection :

An example of a flood detection is the ALERT system deployed in the US. Several types of sensors deployed in the ALERT system are rainfall, water level and weather sensors. These sensors supply information to the centralized database system in a pre-defined way [7].

3) Pollution Monitoring:

WSNs can be deployed for monitoring the level of pollution and warning generation. Air Pollution Monitoring Systems are deployed in cities like London and Brisbane to monitor the level of pollutants. These sensor networks look for amount of poisonous gases and these statistics are studied to analyze if pollution has increased and take actions to check pollution.

C. Health applications

Some of the health applications for sensor networks are providing interfaces for the disabled; integrated patient monitoring; diagnostics; drug administration in hospitals; monitoring the movements and internal processes of insects or other small animals; telemonitoring of human physiological data; and tracking and monitoring doctors and patients inside a hospital.

D. Smart Home applications

Sensor networks can be used for Home Automation application. Sensors can be configured with home appliances like Washing Machine, Air conditioners, vacuum cleaners and oven etc. This will make easy handling of these devices and now appliances can be accessed via Internet and can be controlled remotely.

E. Applications to Automobiles

Even sensors can be employed in vehicles which provide advance tracking mechanism for vehicles as well as tracking for theft vehicles by Police. Applications of WSN to vehicles are:

1).Detecting and monitoring car thefts:

Sensor nodes are being deployed to detect and identify threats within a geographic region and report these threats to remote end users by the Internet for analysis.

2*) Vehicle tracking and detection:*

Sensor nodes can be embedded in vehicle designs which is connected to WSN, the vehicle can be tracked with the help of sensor networks.

3) Traffic Control:

Sensor networks have been used for vehicle traffic monitoring and control for quite a while. Most traffic intersections have either overhead or buried sensors to detect vehicles and control traffic lights. Furthermore, video cameras are frequently used to monitor road segments with heavy traffic, with the video sent to human operators at central locations.

E. Other commercial applications

Some of the commercial applications are monitoring material fatigue; building virtual keyboards; managing inventory; monitoring product quality; constructing smart office spaces; environmental control in office buildings; robot control and guidance in automatic manufacturing environments; interactive toys; interactive museums; factory process control and automation; monitoring disaster area; smart structures with sensor nodes embedded inside; machine diagnosis; transportation; factory instrumentation; local control of actuators; detecting and monitoring car thefts; vehicle tracking and detection; and instrumentation of semiconductor processing chambers, rotating machinery, wind tunnels, and anechoic chambers.

VI. CONCLUSION AND FUTURE WORK

Its applications are not limited to Military Services but extend to vast area of human activities. The flexibility, fault tolerance, high sensing fidelity, low-cost and rapid deployment characteristics of sensor networks create many new and exciting application areas for remote sensing .In future work we have enhance the energy efficiency and increase the network throughput and network lifetime.

REFERENCES

[1] Lucas D. P. Mendes, Joel J. P. C. Rodrigues , Athanasios V. Vasilakos , and Liang Zhou “Lifetime Analysis of a Slotted ALOHA-based Wireless Sensor Network using a Cross-layer Frame Rate Adaptation Scheme”IEEE,2011

[2] K.Ramesh and Dr. K.Somasundaram “A Comparative Study of Cluster-Head Selection Algorithms in Wireless Sensor Networks” International Journal of Computer Science & Engineering Survey (IJCSES) November 2011.

[3] E. Ilker Oyman and Cem Ersoy, “Multiple Sink Network Design Problem in Large Scale Wireless Sensor Networks” IEEE, 2011.

[4] I.F. Akyildiz, W. Su*, Y. Sankarasubramaniam, E. Cayirci, “Wireless Sensor Networks: A Survey Broadband And Wireless Networking Laboratory, School of Electrical and Computer Engineering, Georgia institute of Technology, Atlanta, GA 30332, USA Received 12 December 2001; accepted 20 December 2001.

[5] D. Estrin, R. Govindan, J. Heidemann, S. Kumar, Nextcentury challenges: scalable coordination in sensor networks,ACM MobiCom’99, Washingtion, USA, 1999, pp. 263–270.

[6] Ms. Sunita, Jyoti Malik, Suman Mor " Comprehensive Study of Applications of Wireless Sensor Network" International Journal of Advanced Research in Computer Science and Software Engineering 2 (11), November- 2012, pp. 56-60.

[7] P. Bonnet, J. Gehrke, P. Seshadri, Querying the physical world, IEEE Personal Communications (October 2000) 10–15.

Review on Energy Efficient WSNs Protocols

Baljinder Kaur
Dept. of Computer Science & Engg.
DAV Institute of Engineering & Technology
Jalandhar, India
bkkhinda@gmail.com

Parveen Kakkar
Dept. of Computer Science and Engg.
DAV Institute of Engineering & Technology
Jalandhar, India
parveen.daviet@gmail.com

Abstract--This paper is a review on the critical problem of network lifetime of WSNs. Due to limited battery of sensor nodes, energy efficiency found to be main constraint of limited life of WSNs. Therefore the main focus of the present work is to find the ways to minimize the energy consumption problem and how one can enhance the network stability period and life time. Many researchers have proposed different kind of the protocols to enhance the network lifetime but still much improvement can be done further to enhance the network lifetime further. The overall objective of this paper is to evaluate the gaps in existing clustering techniques of WSNs. This paper has also evaluated the issues which have been neglected in the field of the WSNs.

KEYWORDS—WSNS, NETWORK LIFETIME, STABLE PERIOD, CLUSTERING, LEACH, DEEC.

I. INTRODUCTION

Wireless Sensor Networks (WSNs) are networks that consist of nodes also called sensors which are deployed in a region. These sensors work with each other to sense various types of physical information from the atmosphere. In various significant fields WSNs are very helpful like environmental traffic, military surveillance, area monitoring, air pollution monitoring, wastewater monitoring, pressure etc [1] [6]. Current WSNs is working on the problems of low-power communication, computation and energy storage.

All sensor nodes process data and transmit it to base station also called sink. In WSNs nodes are battery constrained due to limited energy [2] [3] [4]. So use of the battery in efficient way becomes critical issue. A number of protocols play an important role to reduce energy consumption. Direct communication and multi-hop data transmission used initially. But due to limited power of sensor nodes these techniques don't work effectively.

Energy is very critical issue in WSN, because of limited energy in sensor nodes, so to conserve energy clustering technique was introduced, in which out of thousands of nodes few nodes become cluster head and they manage the entire network.

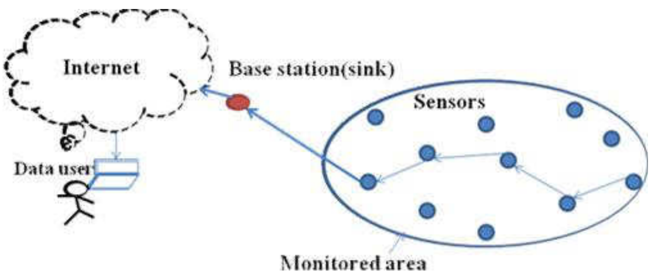


Fig.1. wireless sensor network

Cluster head is a node which is responsible for maintain cluster, collect data from nodes in the cluster and communicating with sink. By using clustering methodology it has been observed that there is large amount of energy that has been saved. In static clustering method some rules were followed to elect a cluster head, once a cluster is formed and cluster head is elected, the cluster was statically operated until the head node dead.

Because cluster head node have more responsibility so rapid decrease in energy in the Cluster head node. The death time of head node was too early in static clustering technique. So there was a need required the W. Heinzelman [4] proposed a protocol based on adaptive clustering technique he named it LEACH.

II. CLUSTERING

Clustering [2] [4] [5] is a technique where nodes are arranged into clusters that are useful in achieving energy efficiency. All nodes belonging to the same cluster send their data to cluster head. The main function of cluster head is to provide efficiently data communication between sensor nodes and the base station. So the cluster head should have high energy as compared to other nodes. CH aggregates data and sends aggregated data to BS where the end-user can access the data.

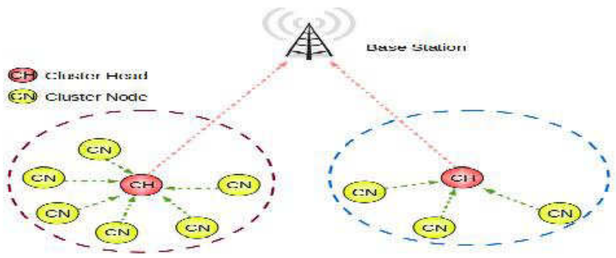


Fig.2. cluster formation in WSN [1]

III. DIFFERENT CLUSTERING SCHEMES

In Wireless sensor network clustering can be done in two types of network i.e. homogeneous and heterogeneous.

All sensor nodes with identical energy level are known as homogeneous WSNs. Low-Energy Adaptive Clustering Hierarchy (LEACH) [4], LEACH–centralized (LEACH-c) [5], Power Efficient Gathering in Sensor Information Systems (PEGASIS) [6] Hybrid Energy-Efficient Distributed Clustering (HEED) [15] protocols are widely used for homogeneous WSNs.

In heterogeneous WSNs all sensor nodes have dissimilar energy level and fewer energy nodes died first than the high energy sensor nodes. Stable Election Protocol (SEP) [7], Distributed Energy Efficient Clustering (DEEC) [8], Stochastic Distributed Energy-Efficient Clustering (SDEEC), Developed DEEC (DDEEC) [10] and Enhanced DEEC (EDEEC) [11] are well known heterogeneous WSNs protocols.

Low-Energy Adaptive Clustering Hierarchy (LEACH) [4] used for homogeneous protocol in which all the nodes have same energy level. Each round starts with set-up phase and followed by the steady phase.

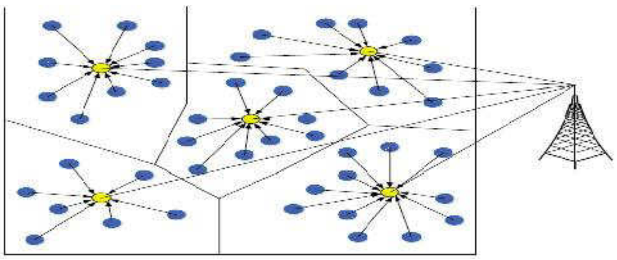


Fig.3. LEACH Clustering Hierarchy [3]

In first phase i.e. setup phase cluster creation and CH selection was take place. CHs selection choice is prepared by the each node select a random number between 0 and 1Threshold T (n) = P/ {1-P(r mod 1/P)} calculated to check a node has chance to become CHs for current round. In this P is desire percentage of CHs, r is the number of current round. If node contains value less than T (n) it becomes CHs for current round and cannot be CHs for the next1 / P rounds. Therefore probability of remaining nodes must be increased. After this in steady phase CHs node receives all data from local nodes compress it and send it to the sink. LEACH is an effective technique to reduce energy dissipation, enhanced network lifetime.

Distributed Energy Efficient Clustering (DEEC) [11] is used for heterogeneous WSNs. In DEEC, the CHs chosen by a possibility based on the ratio between the remaining energy of every node and the average energy of the WSNs. The round number of the rotating period for every sensor node is dissimilar to its initial and remaining energy. The sensor nodes with maximum initial and remaining energy will have more chance to become the CHs than normal nodes.

In LEACH every node has chance to become a CHs after 1/p rounds. All the nodes cannot same remaining energy when sensor network evolves so, the energy will be not well distributed and the low-energy nodes will finish earlier than the high-energy nodes. DEEC uses initial and remaining energy and provides good performance in the networks containing normal and advanced nodes.

Developed Distributed Energy-Efficient Clustering (DDEEC) [15] allows to balance the cluster head selection overall WSNs nodes following their remaining energy. DDEEC uses same method for estimation of average energy and CH selection algorithm based on remaining energy as applied in DEEC.

After some period advanced nodes having same remaining energy like normal nodes. Although, DEEC continues to punish the advanced nodes so this is not best way for energy distribution, because after this advanced nodes are continuously be a CH and they expire faster than normal nodes. To avoid this problem DDEEC presents a threshold residual energy .When advanced and normal nodes energy level less than threshold residual energy then both types of nodes use same probability to become cluster head. Therefore, CH selection is balanced and more efficient in DDEEC.

Enhanced Distributed Energy Efficient Clustering (EDEEC) [16] scheme uses the idea of DEEC with addition of super nodes and expands it into three level heterogeneity WSNs. It includes three types of nodes i.e. normal, advanced and super nodes with their probabilities based on initial energy.

Enhanced Developed Distributed Energy Efficient Clustering (EDDEEC) [2] scheme is used for three-level heterogeneous WSNs. It uses same method for CH selection based on initial, remaining energy level of the nodes and average energy of network as in DEEC. In EDEEC after some rounds, some super and advance nodes have same remaining energy level as normal nodes due to continually CH selection. Therefore it continues to penalize advance and super sensor nodes. Same issue with DEEC, it also continues to penalize just advance nodes and DDEEC is limited only for two-level heterogeneous networks.

To remove this unbalanced problem in three-level heterogeneous WSNs EDDEEC changes in function which described in EDEEC for calculating probabilities of normal, advance and super nodes. These modifications are based on absolute remaining energy level that is the value in which advance and super sensor nodes have similar energy level as in case of normal nodes. Using absolute remaining energy all kinds of nodes has identical probability for CH selection.

IV. LITERATURE SURVEY

Heinzelman et al. (2000) [4] has proposed Low-Energy Adaptive Clustering Hierarchy (LEACH) protocol for homogeneous WSNs that has all nodes of same energy level. In LEACH every node to become a cluster head depends on its individual probability. It equally assigned the energy load among the sensor nodes by use of randomized alternation of cluster-head.

Chandrakasan et al. (2002) [5] has proposed LEACH-Centralized protocol. LEACH did not provide the guarantee about the position and number of group leader nodes. So LEACH-C is more efficient because in this base station has global information of the energy level and location of all the nodes, so it created better clusters.

Lindsey et al. (2002) [6] has presented an improved scheme Power-Efficient Gathering in Sensor Information Systems also called PEGASIS. PEGASIS is a chain-based protocol instead of cluster-based protocol. Each node communicates only with its neighbor node and transmitting data to the sink, so it reduced the amount of energy spent per cycle. Distance of the nodes that broadcast data are fewer in PEGASIS as compare to a CH in LEACH so, it saves energy.

Smaragdakis et al. (2004) [7] has proposed Stable Election Protocol also called SEP. It was made for the two-level heterogeneous networks that contain two types of sensor nodes. According to their own energy each node is able to become cluster head, based on individual selection probability. SEP improved the stability time, which is defined as the time period before death of the first node.

Qing et al. (2006) [8] has presented a new protocol Distributed energy-efficient clustering (DEEC) for heterogeneous WSNs. Cluster-heads selection depends on the ratio between remaining energy of each node and the average energy of WSNs. In this protocol, node containing more energy has more probability to be a cluster head. DEEC does not need full energy information at every time though the choice of cluster head.

Elbhiri et al. (2009) [9] has proposed Distributed Energy Efficient Clustering also called SDEEC for the heterogeneous WSNs. It permits to balance the cluster head choice overall network nodes follow their remaining energy. It has optimized the intra-clusters communication to prolong network lifetime and give better performance than the SEP and DEEC.

Elbhiri et al. (2010) [10] has proposed a new heterogeneous protocol called Developed Distributed Energy-Efficient Clustering (DDEEC). This choice of protocol always selected the advanced nodes, when their remaining energy decrease and be converted into the series of the normal sensor nodes. It optimized the cluster head choice by following their remaining energy and it performed better than DEEC.

Saini et al. (2010) [11] has presented an improved protocol Enhanced Distributed Energy Efficient Clustering (EDEEC). EDEEC comes up with three kinds of sensor nodes that can be used to increase the lifetime and stability of the WSNs. It enhances the heterogeneity, energy level and received more data packets at BS than SEP.

Alla et al. (2011) [12] has proposed a new technique Balanced and Centralized Distributed Energy Efficient Clustering also called BCDEEC. In this BS guarantees that the high energy nodes suitable first entrance for cluster heads to get better lifetime of WSNs and saving average energy It gives the better performance than SEP and DEEC.

Miao et al. (2012) [13] has proposed a performance analyses of low-energy adaptive clustering hierarchy routing algorithm (LEACH), distributed energy efficient clustering algorithm (DEEC) and Stable Election Protocol (SEP). LEACH does not perform well in heterogeneous environment. SEP is composed of two types of nodes according to the initial power. In DEEC High initial and remaining energy nodes will have more probability to be CH than lower energy nodes. Thus, DEEC can improve the stability period and Lifetime of WSNs.

Qureshi et al. (2012) [14] has tested the performance of Distributed Energy- Efficient Clustering, Threshold DEEC, Developed DEEC and Enhanced DEEC. It tested these schemes under a number of different condition hold high level heterogeneity to low level heterogeneity. DEEC and DDEEC performed well in three level heterogeneous network hold high energy level variation among nodes. EDEEC and TDEEC performed well in all heterogeneous state containing low energy level variation among.

V. GAPS IN EARILER WORK

By conducting the literature survey it has been found that the every WSNs protocol has some limitations; i.e. no one is perfect in every case and most of the existing literature has neglected one of the following:

1. The Most of the existing researchers has neglected the use of the distance between the sensor node and the base station while selecting the cluster head.

2. The optimum numbers of clusters in every round are not consistent in LEACH as well as in DEEC variants.

3. The use of the fuzzy cost is also ignored in many of the cases.

VI. COMPARITATIVE ANALYSIS

Table 1 has shown the comparison among various WSNs protocols based upon first node dead time and last node dead time. The network dimensions are of 100* 100 and initial energy is .5 J other variables are based on as in [2] [8].

Table I. Comparison among various WSNs protocols

Protocol	Network Type	First node died	Last node died
LEACH	Homogeneous	1000	1450
SEP	Heterogeneous	1350	1800
LEACH-E	Homogeneous	1450	4200
DEEC	Heterogeneous	969	5536
DDEEC	Heterogeneous	1355	5673
EDEEC	Heterogeneous	1432	8638
EDDEEC	Heterogeneous	1777	8638

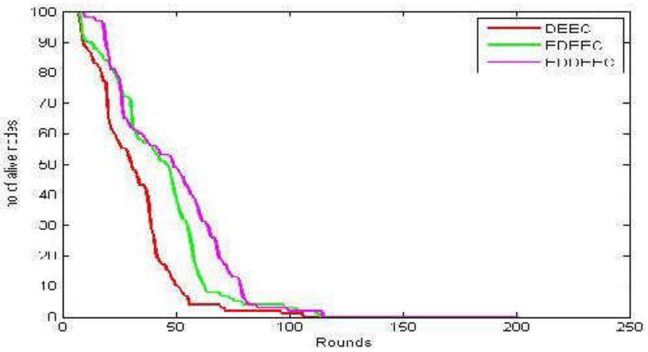


Fig.4. Alive nodes during network lifetime

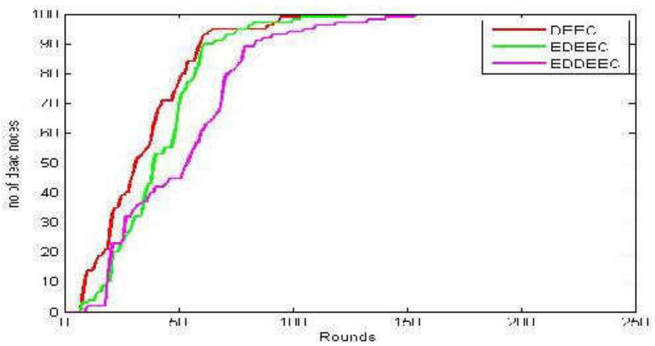


Fig.5. Dead nodes during network lifetime

Figure 4 and 5 has shown that EDDEEC gave better results than DEEC and EDEEC in terms of alive and dead nodes. The comparison among various protocols has shown that EDDEEC has shown quite significant results over the available protocols.

VI. ANALYTICAL SOLUTION

One can improve the performance of the EDDEEC using fuzzy based cluster head selection. The EDDEEC has used different probability function for selecting the best cluster head by using the residual energy and average energy of the network. But EDDEEC has neglected the distance between base station and cluster. The fuzzy cost will be evaluated on the basis of the residual energy and the node centrality. The fuzzy cost will be dynamic in nature as it is evaluated in each round. Thus will provide more better results due to its adaptive nature i.e. will change as the residual energy changes. The main advantage of this solution is that the optimum numbers of clusters are formed in every round, which is almost impossible in LEACH and also not guaranteed in EDDEEC.

The possible steps to remove the short comings of the earlier work are:-

Step 1: Initialize the WSNs with required parameters like nodes position; sink position, initial energy of each kind of nodes etc.

Step 2: for every node i repeat the following steps until all nodes become dead.

Step 3: Select cluster head using following equations i.e. normal (eq.1), advance (eq.2), super nodes (eq.3) and for all types of nodes having same remaining energy (eq.4). Here is Fuzzy cost and is Tabsolute.

$\frac{}{(1-}$ for normal nodes if $>$ (1)

$\frac{1}{(1-}$ for advance node if $>$ (2)

$\frac{}{(1-}$ for super nodes if (3)

$c\frac{1}{(1-}$ for all nodes if (4)

Where Fuzzy cost will be evaluated using the Algorithm 1.

Step 4: Evaluate the energy dissipation and update the remaining energies it. Where distance will be evaluated using eq. 5 [2] and updating of energy will be based upon the eq. 6 [13] and eq. 7 [13].

, (5)

$E_{Tx}(l,d) =$, $d <$ (6)

$E_{Tx}(l,d) = l$, d (7)

Algorithm 1. Fuzzy cost calculation

1. First Cycle

1. Sink selects cluster head according to the weighted probability function (Padv for advance nodes, Psup for super nodes, Pnrm for normal nodes) and broadcast the CH_message

2. Cluster formation will be done and data transform will be take place

3. Each sensor node computes the remaining energy and sensor node centrality and sends the values to sink through Cluster Head

4. End

2. Normal Cycle
1. Fuzzy cost will be calculated by Sink using remaining energy and sensor node centrality
2. Sink chooses the Cluster Head based on the value of fuzzy cost and broadcast the CH_message
3. Cluster formation will be done and data transform will be take place
4. Each sensor node computes the sensor node centrality and remaining energy and sends the values to sink through Cluster Head
5. End

VIII. CONCLUSION

It has been found that the most of the existing researchers has worked hard to prolong the network lifetime. Many modifications have shown significant improvement over the existing protocols like LEACH. But it is also found that the most of the researchers has neglected many issues of WSNs. That means no technique is best for every case. The use of Fuzzy theory in cluster head selection has been neglected by the most of existing researchers. The analytical solution has shown that it will come up with improvisation in network lifetime.

Therefore in near future we will use fuzzy cost for cluster head selection to enhance the results further. Fuzzy cost will be evaluated on the basis of the residual energy and the node centrality. No implementation is considered in this research work in near future we will use suitable simulation tool to implement the different WSNs protocols.

REFERENCES

[1] A. Kashaf, N. Javaid, Z. A. Khan and I. A. Khan “TSEP: Threshold-sensitive Stable Election Protocol for WSNs”, 10th International Conference on Frontiers of Information Technology (FIT), 2012, pp.164-168.

[2] N.Javaid, T. N. Qureshi, A. H. Khan, A. Iqbal, E. Akhtar, and M. Ishfaq, "EDDEEC: Enhanced Developed Distributed Energy-Efficient Clustering for Heterogeneous Wireless Sensor Networks.", International Workshop on Body Area Sensor Networks, 2013, pp.914-919.

[3] Muneer Alshowkan, Khaled Elleithy and Hussain Alhassan, "LS-LEACH: A New Secure and Energy Efficient Routing Protocol for Wireless Sensor Networks", 17th International Symposium on Distributed Simulation and Real Time Applications (DS-RT), 2013, pp. 215-220.

[4] Heinzelman, Wendi Rabiner, Anantha Chandrakasan, and Hari Balakrishnan, “Energy-efficient communication protocol for wireless microsensor networks”, in Proc. 33rd Hawaii IEEE International Conference on System Sciences, 2000, pp.1-10

[5] Wendi Rabiner Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan, “An Application-Specific Protocol Architecture for Wireless Microsensor Networks,” IEEE Transaction on Wireless Comm., vol. 1, no. 4, pp. 660- 670, 2002.

[6] Stephanie Lmdsey and Cauligi S. Raghavendra, “PEGASIS: Power-efficient gathering in sensor information systems”, In Aerospace conference proceedings, 2002, pp. 1125-1130.

[7] Georgios Smaragdakis, Ibrahim Matta and Azer Bestavros, “SEP: A Stable Election Protocol for clustered heterogeneous wireless sensor network”, In Second International Workshop on Sensor and Actor Network Protocols and Applications (SANPA), 2004, pp.1-10.

[8] Li Qing, Qingxin Zhu and Mingwen Wang, “Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks”, Computer communications, vol. 29, no. 12, pp. 2230-2237, 2006.

[9] B. Elbhiri, R. Saadane and D. Aboutajdine, “Stochastic Distributed Energy-Efficient Clustering (SDEEC) for heterogeneous wireless sensor networks”, ICGST-CNIR Journal, Vol.9, Dec, pp.11-17, 2009.

[10] Brain.Elbhiri, Saadane Rachid, Sanaa El fkihi and Driss Aboutajdine, “Developed Distributed Energy-Efficient Clustering (DDEEC) for heterogeneous wireless sensor networks”, 5th International Symposium on Communications and Mobile Network, 2010, pp.1-4.

[11] Parul Saini and Ajay.K.Sharma, “E-DEEC- Enhanced Distributed Energy Efficient Clustering Scheme for heterogeneous WSN”, 1st International Conference on Parallel, Distributed and Grid Computing (PDGC), 2010, pp. 205-210

[12] Ben Alla, Abdellah Ezzati, Ahmed Mouhsen , Abderrahim Beni Hssane and Moulay Lahcen Hasnaoui, "Balanced and Centralized Distributed Energy Efficient Clustering for heterogeneous wireless sensor networks", 3rd International Conference on Next Generation Networks and Services, 2011, pp.39-44.

[13] Yu Miao, Bai Guang-wei and Shen Hang, "Performance Study of Routing Mechanisms in Heterogeneous WSNs”, International Conference on Computer Science & Service System (CSSS), 2012, pp. 971-974.

[14] T. N. Qureshi, N. Javaid, M. Malik, U. Qasim and Z. A. Khan, “On Performance Evaluation of Variants of DEEC in WSNs”,7th International Conference on Broadband, Wireless Computing, Communication and Applications (BWCCA), 2012, pp.162-169.

[15] Ossama Younis and Sonia Fahmy, “HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks”, IEEE Transaction on Mobile Computing, vol.3, no.4, pp .366 -379, 2004.

A Survey of Routing Issues in Heterogeneous Wireless Sensor Networks

Kuljeet Kaur
Dept. of CSE
GNDU, RC Jalandhar
kuljeetkaur032@gmail.com

Jyoteesh Malhotra
Dept. of CSE
GNDU, RC Jalandhar
jyoteesh@gmail.com

***Abstract*—Heterogeneous wireless sensor networks (H-WSN) are formed by small sensor nodes communicating over wireless links without using a fixed network infrastructure. Sensor nodes have a limited transmission range, and their processing and storage capabilities as well as their energy resources are also limited. In the H- WSN in routing is used to communication. Some time in routing many problems can occur. In this paper survey on routing issues and their solutions are covered. Routing protocols for wireless sensor networks are responsible for maintaining the routes in the network but in the routing protocols at the time some problems are occur like energy efficiency, fault tolerant, wrong route etc. In the H-WSN routing issues in routing algorithm regarding issues, protocols design issues can occur.**

***Keywords*— Wireless sensor network, H-WSN, Routing protocols and its comparison, routing issues.**

I. INTRODUCTION

A wireless sensor network (WSN) is distributed sensors to monitor physical or environmental conditions, such as temperature, sound, pressure, etc. and to cooperatively pass their data through the network to a main location. The more modern networks are bi-directional, also enabling control of sensor activity. [1] Sensors are small with limited processing and computing resources, and they are inexpensive compared to traditional sensors. These sensors nodes can sense, measure, and gather information from the environment and, based on some local decision process. They can transmit the sensed data to the user. [1]

Heterogeneous wireless sensor network (heterogeneous WSN) consists of sensor nodes with different ability, such as different computing power and sensing range. [2] H-WSNs allow for a variety of operating environments, and hence are useful for many practical applications. [3]. H-WSN where sensors have the same capabilities in terms of computation, communication, memory, power supply, reliability, etc. [4].

Routing for WSNs in energy efficiency and network capacity are two of the most important issues in wireless sensor networks. Many to one communication paradigm are widely used in regard to sensor networks. Sensor nodes send data to a common sink for processing. Sensor networks can be divided in to two classes as event driven and continuous dissemination networks. In event-driven networks, data is sent when an event occurs. In continuous dissemination networks, every node periodically sends data to the sink. The routing is mainly divided into flat based, location based and Hierarchical based routing. [5]

The paper has been organized as follows: in section II classification of routing protocols for heterogeneous wireless sensor networks and in this section describe the comparison of these protocols .In the section III Challenges of routing in the H-WSN, the section IV is related work of these issues. The section V in covered the issues of H-WSN and VI is open issue and future scope. The section VII is the conclusion.

II. CLASSIFICATION OF ROUTING PROTOCOLS IN H-WSN

In heterogeneity sensor network architecture, there are two types of sensors namely line-powered sensors which have no energy constraint, and the battery-powered sensors having limited lifetime, and hence should use their available energy efficiently by minimizing their potential of data communication and computation. . In this section we discuss uses of heterogeneity in WSNs to present a few routing protocols.

A. Information-Driven Sensor Query (IDSQ):

In IDSQ protocol, first step is to select a sensor as leader from the cluster of sensors. This leader will be responsible for selecting optimal sensors based on some information utility measure.

B. Cluster-Head Relay Routing (CHR):

CHR routing protocol uses two types of sensors to form a heterogeneous network with a single sink: a large number of low-end sensors, denoted by L-sensors, and a small number of powerful high-end sensors, denoted by H-sensors. Both types of sensors are static and aware of their locations using some location service. Land H-sensors are uniformly and randomly distributed in the sensor field. The CHR protocol partitions the heterogeneous network into groups of sensors each being composed of L-sensors and led by an H-sensor. Within a cluster, the L-sensors are in charge of sensing the underlying environment and forwarding data packets originated by other L-sensors toward their cluster head in a multi hop fashion. The H-sensors, on the other hand, are responsible for data fusion within their own clusters and forwarding aggregated data packets originated from other cluster heads toward the sink in a multi hop fashion using only cluster heads. While L-sensors use short-range data transmission to their neighboring H-sensors within the same cluster, H-sensors perform long-range data communication to other neighboring H-sensors and the sink. [6]

The table shows the comparison in the routing protocols.

Table I. Represents Classification and Comparison of routing protocols

Routing protocols	classification	Power usage	Data aggregation	scalability	Query based	Over head	Datadelivery model	QoS
SPIN	Flat/Data centric	Limited	Yes	Limited	Yes	Low	Event driven	No
DD	Flat/Data centric/Dst-initiated	Limited	Yes	Limited	Yes	Low	Demand driven	No
RR	Flat	Low	Yes	Good	Yes	Low	Demand driven	No
GBR	Flat	Low	Yes	Ltd	Yes	Low	Hybrid	No
CADR	Flat	Limited		Ltd.	Yes	Low	Continu-ously	No
COUGAR	Flat	Limited	Yes	Limited	Yes	High	Query driven	No
ACCQUIRE	Flat/datacentrc	Low	Yes	Limited	Yes	Low	Complex query	No
LEACH	Hierarchical/Dst-initiated/ node-centric	High	Yes	Good	No	High	Cluster-head	No
TEEN& APTEEN	Hierarchical	High	Yes	Good	No	High	Cluster-head	No
PEGASIS	Hierarchical	Max	No	Good	No	Low	Chains based	No
VGA	Hierarchical	Low	Yes	Good	No	High	Good	No
SOP	Hierarchical	Low	No	Good	No	High	Continously	No
GAP	Hierarchical/ Location	Limited	No	Good	No	Mod	Virtual grid	No
SPAN	Hierarchical/ Location	Limited	Yes	Limited	No	High	Continously	No
GEAR	Location	Limited	No	Limited	No	Mod	Demand driven	No
SAR	data centric	High	Yes	Limited	Yes	High	Continously	Yes
SPEED	Location/ data centric	Low	No	Limited	Yes	Less	geographic	Yes

III. CHALLANGES OF H-WSN IN ROUTING

A. *Data aggregation*

This technique has been used to achieve energy efficiency and data transfer optimization in a number of routing protocols. Signal processing methods can also be used for data aggregation. Sensor nodes might generate significant redundant data; similar packets from multiple nodes can be aggregated so that the number of transmissions would be reduced [6]. A multi keyword routing protocol for HWSNs they provides an efficient query interface for locating data; queries are routed only along paths with nodes matching the query. The keywords are aggregated via a novel use of lattices to reduce network cost. [7]

B. *Quality of Service (QoS)*

The quality of service means the quality service required by the application, it could be the length of life time, the data reliable, energy efficiency, and location-awareness, collaborative-processing. These factors will affect the selection of routing protocols for a particular application.[6] Sensor networks have specific requirements on energy saving, data-oriented communication, and inter connection between non-IP and IP, therefore sensor network dedicated routing protocols may be required, for energy efficient routing scheme. In WSN there are the routing protocols that minimize the used energy. [8]

C. *Fault Tolerance*

Some sensor nodes may fail or be blocked due to lack of power, have physical damage or environmental interference. The failure of sensor nodes should not affect the overall task of the sensor network. This is the reliability or fault tolerance issue. Fault tolerance is the ability to sustain sensor network functionalities without any interruption due to sensor node failures. [3]

D. Node/Link Heterogeneity

The existence of heterogeneous set of sensors raises many technical issues related to data routing. Some applications might require a diverse mixture of sensors for monitoring temperature, pressure and humidity of the surrounding environment, detecting motion via acoustic signatures, and capturing the image or video tracking of moving objects. Hierarchical protocols designate a cluster- head node different from the normal sensors. These cluster heads can be chosen from the deployed sensors or can be more powerful than other sensor nodes in terms of energy, bandwidth, and memory. In the sensor node link problems can occur in the deployed situation. [9]

E. Routing stability and coverage

It is a very important issue in addition to energy, bandwidth etc. In WSNs, each sensor node obtains a certain view of the environment. A given sensor's view of the environment is limited both in range and in accuracy; it can only cover a limited physical area of the environment.

IV. RELATED WORK

Considering the challenges and constraints highlighted in previous section of heterogeneous wireless sensor networks. The work done inorder to resolve these issues by various authors has been complied in this section.

Xiaofeng Han Xiang Cao Errol L. Lloyd Chien-Chung Shen [3] (2007) Described the definition of H-WSN and covered the fault tolerance issue.

Chun-Hsien Wu, Yeh-Ching Chung [2] (2007) surveyed the heterogeneous wireless sensor network in network sensors are used with different- different sensing range.

Jennifer Yick, in [1] (2008) described the wireless sensor network. In the WSN sensors are used, these sensors are small with the limited processing. The cost of these sensors is high as compare to traditional sensors.

Matt Spear, Karl Levitt in [7] (2008) covered the issues of quality of services the protocols are minimized the used energy.

Mansoor-uz-Zafar Dawood in [8] (2009) covered the issue of routing of H-WSN in quality of services issue and their solution. The issue of solutions is it minimize the total consume energy. It increases the total lifetime. It increases the productivity. If all traffic is routed through the minimum energy path to the destination, the nodes along that path will run out of batteries quickly rending other nodes useless due to the network partition even if they do have available energy.

Shio Kumar Singh, M P Singh, and D K Singh [6] (2010) described the quality of services issue and defines about cluster head relay routing.

Jasmine Norman in [5] (July 2010) described the heterogeneous routing and how to communicate the heterogeneous routing with one node to other.

IEEE ICC in [4] (2012) presented the heterogeneous wireless sensor network. This paper searched the H-WSN in increase the productivity.

Miloud Bagaa, Yacine Challal, Adlen Ksentini, Abdelouahid Derhab, and Nadjib Badache [10] (2014) Covered the solution of the data aggregation the data aggregation principle and focused on features and requirements that should be implemented by data aggregation scheduling protocols. The surveyed solutions aim to reduce the data latency, increase the data accuracy and aggregation freshness with a good distribution of nodes' waiting time (WT).

Jamal N. Al-Karaki [9] covered the issues of node and link heterogeneity.

V. ISSUES OF ROUTING IN H-WSN

Various authors have discussed about the routing issues of H-WSN in the survey as given above. Table given below summarizes these issues.

Table II.Comparison

Paper name	Issues	Remarks
Fault-tolerant Relay Node Placement in Heterogeneous Wireless Sensor Networks (2007)	Fault tolerance	A minimum number of relay nodes to achieve diverse level of fault tolerance, Ongoing work is performance ratios of the approximation algorithms.
iBubble: Multi-keyword Routing Protocol for Heterogeneous Wireless Sensor Networks(2008)	Data aggregation	iBubble bridges many routing and energy problems prevalent in WSNs, and provides a simple, uniform solution.
Routing Protocols in Wireless Sensor Networks - A Survey(2010)	Quality of services and data aggregation	More powerful and efficient protocols are required. Two imp. Research directions are 3D and 2D when designing some protocols.
Routing Techniques in Wireless Sensor Networks	Node/Link Heterogeneity	Improve the connection of nodes. Future research energy and bandwidth requirements.

VI. OPEN ISSUES AND FUTURE SCOPE

Based on the extensive literature survey done in section no V following issues has been extracted and they are given below:

In the wireless sensor networks in when the connect sensor nodes those time some problems occurs like limited energy supply, limited computing power, and limited bandwidth of the wireless links connecting sensor nodes. These problems are comes in some deployment nodes. In the routing of sensor network time and location synchronization problem can occur. The reconfiguration is very important for lifetime network.

Many applications of heterogeneous wireless network will be designed which have expected to have good routing techniques consume less energy so that battery lifetime can be maximized. Many researchers are working to remove the routing issues in wireless sensor networks but still many issues exist which need to be solved in future.

VII. CONCLUSION

As given above various authors have discussed the various routing issues of H-WSN like data aggregation, node\link heterogeneity , quality of services, reliability ,fault tolerance ,sensor nodes etc. In these issues mainly energy efficiency problem and the path selection etc are covered. In this paper literature survey was reviewed to find out related quality of services and data aggregation problems. The quality of services in sensor nodes consumes energy more equitably, they continue to provide connectivity for longer, and the network lifetime increases. The data aggregation reduces the data latency and increase the data accuracy. Here solutions for some of the challenges are given but these challenges are not solved properly, still they have some problems.

REFERENCES

[1] Jennifer Yick, Biswanath Mukherjee, Dipak Ghosal "Wireless sensor network survey", (2008).

[2] Yeh-Ching Chung, "Heterogeneous Wireless Sensor Network Deployment And Topology Control Based on Irregular Sensor Model", IEEE Volume4459, (2007).

[3] Xiaofeng Han Xiang Cao Errol L. Lloyd Chien-Chung Shen, "Fault-tolerant Relay Node Placement in Heterogeneous Wireless Sensor Networks", IEEE INFOCOM (2007).

[4] Xiao Chen, Zanxun Dai,"A Layer-Based Routing Protocol for Heterogeneous Wireless Sensor Networks", IEEE ICC Ad - hoc and sensor networking Symposium (2012).

[5] Jasmine Norman, J. Paulraj Joseph "HLAOD –A Cross Layer Routing Protocol for Pervasive Heterogeneous Wireless Sensor Networks Based On Location", IJCSI International Journal of Computer Science Issues, Vol. 7 Issue 4, No 7, (July 2010).

[6] Shio Kumar Singh, M P Singh, and D K Singh, "Routing Protocols in Wireless Sensor Networks - A Survey", IJCSES Vol. 1, No. 2 (November 2010).

[7] Xiaoming Lu, Matt Spear, Karl Levitt and S. Felix Wu, "iBubble: Multi - keyword Routing Protocol for Heterogeneous Wireless Sensor Networks", IEEE INFOCOM (2008).

[8] Mansoor – uz -za far Dawood,"Designing of energy efficient routing protocol for Wireless Sensor Network (WSN) Using Location Aware (LA) Algorithm" Journal of information of communication technology vol. 3, no. 2 (2009).

[9] Jamal N. Al-Karaki, "Routing Techniques in Wireless Sensor Networks - A survey", ICUBE initiative of Iowa State University, Ames IA 50011.

[10] Miloud Bagaa, Yacine Challal, Adlen Ksentini, Abdelouahid Derhab, and Nadjib Badache, "Data Aggregation Scheduling Algorithms in Wireless Sensor Networks: Solutions and Challenges" IEEE Communications survey & Tutorials.

Simulative Analysis of Spectral Parameters for Wireless VoIP

Sansar Chand
Dept. Of ECE
CT Institutions
sumisansar@gmail.com
Punjab, India

Harjit Pal Singh
Dept. Of ECE
CT Institutions
Punjab, India

Jasbinder Singh
Dept. of CE
MD College
Sri Ganganagar

Monica Sharma
Dept. Of CSE
CT Institutions
Punjab, India

Abstract: The transmission of Voice over Internet protocol (VoIP) requires signal processing for efficient performance over the communication networks. Speech quality, as perceived by the users of VoIP telephony, is critically important. Signal quality in wireless VoIP is deteriorated by network layers, which includes delay, jitter and packet loss. In this paper comparative analysis of speech quality for narrowband and wideband VoIP over wireless network is presented. The VoIP simulations are conducted for G.729A and AMR-WB speech coders at different packet loss rates. The comparative results of various speech analysis algorithms for both the narrowband and the wideband VoIP signals are presented.

Keywords- Digital Signal Processing, Narrowband VoIP, Wideband VoIP, Signal Analysis.

I. INTRODUCTION

Voice over IP is an advanced technology which is used to transmit voice media over the internet using internet protocol (IP) [1]. This technique provides us enhanced features such as low cost as compared to the traditional Public Switched Telephone Network (PSTN). VoIP systems cost as much as half that of traditional PSTN system in the field of voice communication. It is due to the efficient use of bandwidth requiring fewer long-distance trunks between switches [2].

Voice over IP has the ability to provide voice and data convergence at the application layer. QoS is one of the parameter that is used for the implementation in voice over IP. The quality of voice is the characteristic of the IP network and is affected by the parameters such as, packet delay, packet loss and congestion. The network traffic and interference causes jitter which is the variation in arrival time between packets, which results in harmful effects on the quality of VoIP. Due to the real time application for interactive speech transmission, it is impossible for the receivers to request the sender to retransmit the packets lost. When voice packets do not arrive before their playout time, they are considered as lost and cannot be played when they are received. One of the most typical problems in such networks is the packet loss. Even a single lost packet may generate audible distortion in the decoded speech signal. To reduce the effect of packet loss on perceived speech quality, the lost packets have to be regenerated at the receiver using packet loss concealment algorithms [3].

In the section II the brief description of the coders used for VoIP simulations is presented. The IP modelling is discussed in the section III. The spectral analysis algorithms are presented in the section IV. The simulation and spectral analysis results are shown in section V. The last section draws the conclusion and future work.

II. VoIP CODERS

A. G.729A coder

The CS-CELP coder is based on a code-excited linear prediction (CELP) coding model. In this model the throat and mouth are modelled as a linear filter and voice is generated by a periodic vibration of air exciting this filter. The locally decoded signal is compared against the original signal and the coder parameters are selected such that the mean-squared weighted error between the original and reconstructed signal is minimized. The CS-ACELP coder is designed to operate with an appropriately band limited signal sampled at 8000Hz. The input and output samples are represented using 16-bit linear PCM. The coder operates on frames of l0ms corresponding to 80 samples. For each frame, the speech signal is analysed to extract the parameters of the CELP model which include linear-prediction filter coefficients, adaptive and fixed-codebook indices and gains. At the decoder, these parameters are used to retrieve the excitation and synthesis filter parameters. The speech is reconstructed by filtering through the short term synthesis filter. After computing the reconstructed speech, it is further enhanced by post-filter [4].

B. AMR-WB Coder

Adaptive multi-rate wideband (AMR-WB) [5, 6] was standardized by 3GPP in 2000 for use in 3G mobile systems and was also adopted by ITU in 2001 as their latest wideband coder, G.722.2. It operates at 9 varying modes of 6.6 kbps up to 23.85 kbps bit rates and describes the detailed mapping from input blocks of 320 speech samples in 16 bit uniform PCM format to encoded blocks of 132, 177, 253,285, 317, 365, 397,461 and 477 bits and from encoded blocks of 132, 177, 253,285, 317, 365, 397,461 and 477 bits to output blocks of 320 reconstructed speech samples. AMR-WB offers even lower bit rate compressions as well as the ability to quickly adapt to varying compression as per the network conditions. The bandwidth is automatically conserved when network congestion is high. When the congestion returns to a normal mode, a lower-quality bit rate is restored.

III. IP NETWORK MODELING

The simulation of VoIP system was performed where each packet contains one frame. Packet losses are not independent on a frame-by-frame basis, but appear in bursts. The packet loss can be approximated by Markovian loss model such as Gilbert model, as discussed by Bolot in [7]. Thus simulation of IP network was performed by using a 2-state Gilbert-Elliot Model. The model has two states reflecting whether the previous packet is received or lost. The state "0" represents that a packet being correctly received and state "1" represents that a packet being lost. The Gilbert-Elliot model is shown in Fig 1.

Let p be the transition probability for the network model to drop a packet given that the previous packet is delivered i.e. the probability for network model to go from state "0" to state "1". Let q is the probability for the network model to drop a packet given that the previous packet is dropped, i.e. the probability for the network model to stay in state "1". This probability is also known as the conditional loss probability.

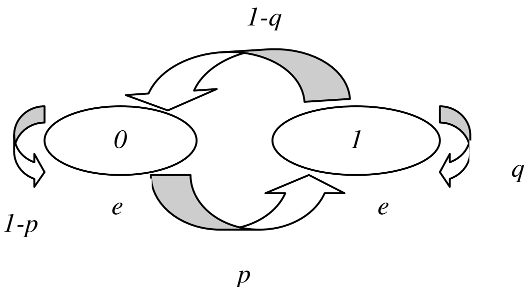


Fig.1 Gilbert-Elliot Model

Let $p0$ and $p1$ denote the probability of the network model to be in state 0 and 1.The probability for a packet to be dropped regardless whether the previous packet is delivered or dropped i.e. the unconditional loss probability is exactly the probability for the network model to be in state 1 ($p1$).

$$p0 = \frac{q}{p+q}\ ; \quad p1 = \frac{p}{p+q} \tag{1}$$

The transition matrix is given as

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \tag{2}$$

A. *VoIP SIMULATIONS*

The VoIP simulations were carried out for making comparison between narrowband speech coding and wideband speech coding algorithms. The speech samples for simulations were taken from [8]. The network impairments were introduced into VoIP frames with the modelling of the IP network through above discussed Gilbert-Elliot model. The speech signal of VoIP system is degraded at different packet loss rates (PLR). The simulations were conducted at different packet loss rates as discussed in Table.1 and then the spectral analysis of the degraded VoIP speech signal is then carried using MATLAB.

IV. SPECTRAL ANALYSIS OF VoIP SIGNAL

The spectrum of the degraded signal is analysed in time and frequency both. Some signal processing algorithms are discussed here [9, 10]:

A. *Energy Spectral Density*

The density of energy of signal $x(t)$ with respect to frequency is represented by,

$$S_{xx}(f) = |X(f)|^2 \tag{3}$$

Table I. Simulated Loss Rates

PLR (%)	p	q	e
3	0.97	0.03	0.005
5	0.95	0.05	0.005
10	0.90	0.10	0.005
15	0.85	0.15	0.005
20	0.80	0.20	0.005

Since finite energy signals posses Fourier transform, spectral analysis performed through spectral energy density.

4.2 Power Density Spectrum

The analysis of power density of the signal is done through periodogram, which is given as,

$$P_{xx}(f) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x(n)e^{-j2\pi fn}\right|^2 \tag{4}$$

B. *Spectrogram*

The spectrogram of the speech signal is an intensity plot of the STFT magnitude. STFT is a sequence of FFTs of windowed data segments where the windows are usually allowed to overlap in time.

C. *Pitch Contour*

While analyzing speech signal, pitch contour plays an important role by tracking the perceived pitch of speech over time. Pitch contour may include multiple sounds utilizing many pitches and can be related to frequency function at one point in time to the frequency function at a later point.

V. SIMULATION RESULTS

The simulations were carried out for making comparison between narrowband VoIP and wideband VoIP. The degraded VoIP signal was analyzed with various digital signal processing algorithms.

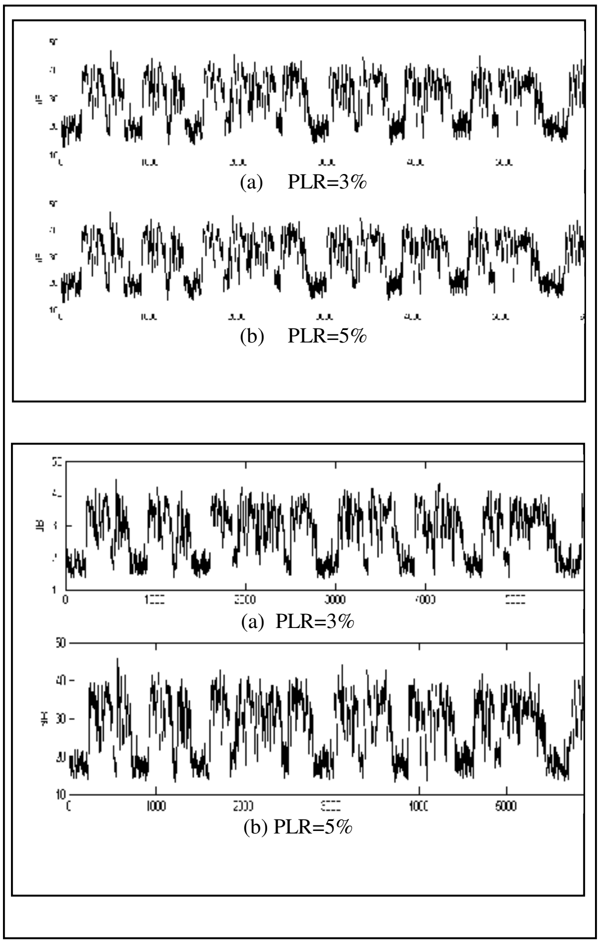


Fig. 2 Spectral Energy Density

The signal spectral quality was tested for narrowband G. 729A coder and wideband AMR-WB coder at 6.6 kbps. The simulation results for various spectral analysis algorithms with packet loss rate of 3% and 20% are shown here:

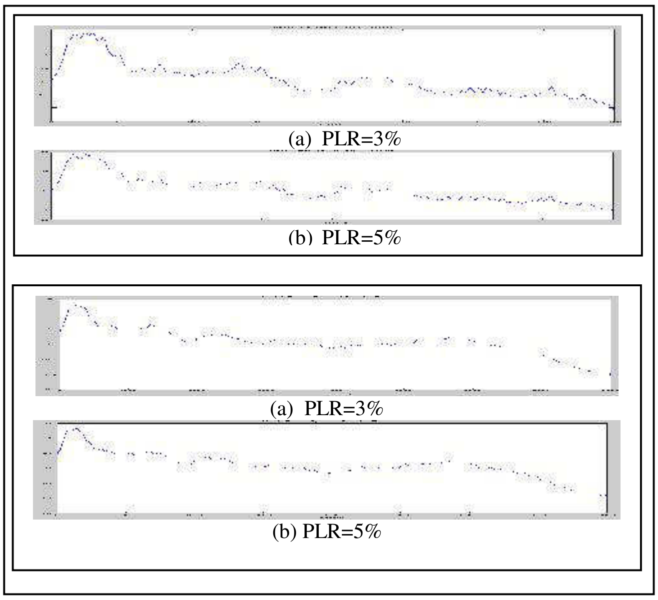


Fig.3 Power Spectral Density

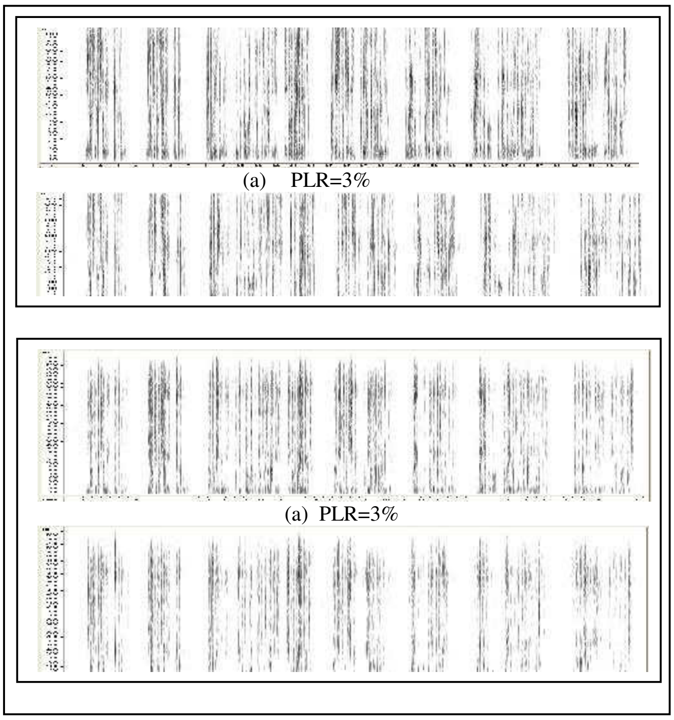


Fig.4 Spectrogram Analysis

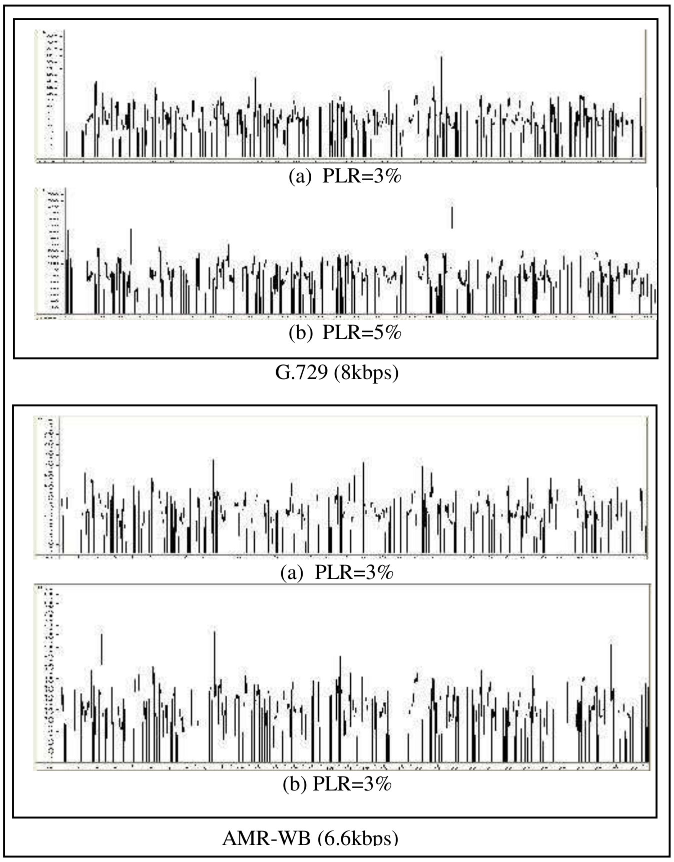


Fig 5 Spectral

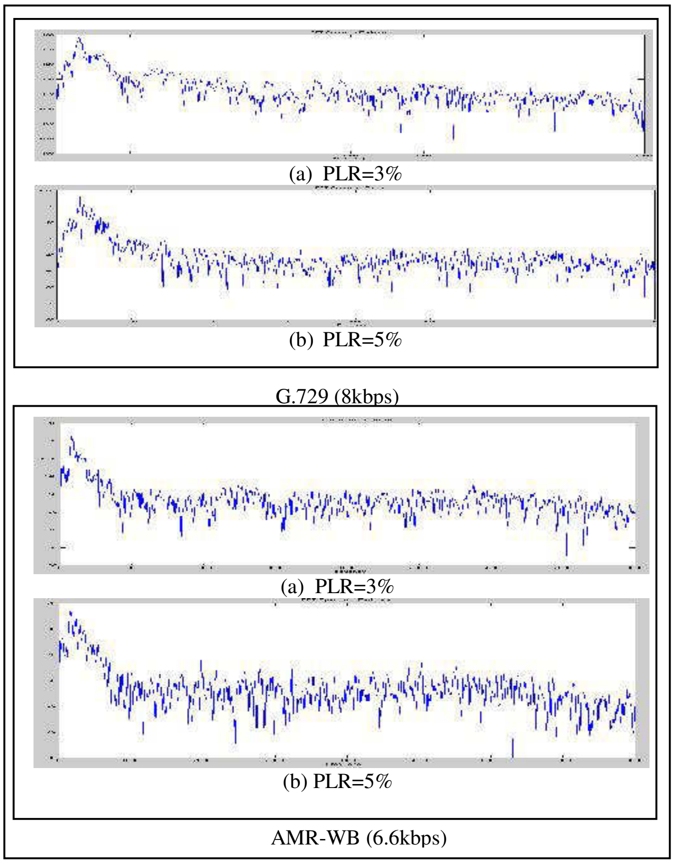


Fig.6 FFT Spectrum

VI. DISCUSSION

The spectral energy density and power spectral density of G.729A and AMR-WB coder is presented in Fig. 2 and Fig. 3.. The results show that the network impairments affects adversely to the speech signal. The additional noise is observed in the waveform with the VoIP speech signal. The spectrogram results in Fig. 4 shows that the wideband coder gives better results than narrowband coders, since the telephony bandwidth is extended from 4 kHz to 7 kHz. It is observed in Fig.5 and Fig.6 that the fundamental frequency is less affected by the wideband VoIP system in packet loss conditions.

VII. CONLUSION

The simulated results of VoIP signal with different speech processing algorithms clearly indicate the significant reduction in noise when VoIP signal is passed through wideband coders. The increased bandwidth of wideband coders enhances the coder's capacity to capture the extra energy above 4 kHz and improves the spectral quality of VoIP speech signal. Today's most demanding VoIP software are also updating their services through wideband coders, to improve the quality and robustness against the transmission networks. In future the study can be used for improving the speech quality using various signal processing algorithms such as wavelet transform performed at higher frequency digital signal processors.

REFERENCES

[1] J.Walter Goralski and C.Matthew Kolon, IP Telephony.McGraw-Hill, 1st Edition (2000).

[2] H P Singh, S Singh, J Singh, S A Khan, "VoIP: State of art for global connectivity—A critical review", Journal of Network and Computer Applications, Vol.37, No.1, 2014, pp:365-379

[3] U.Varshney, A.Snow,M.McGivern and C.Howard," Voice over IP", Communications of the ACM, January 2002.

[4] ITU, ITU-T G.729: CS-ACELP speech coding at 8 kbit/s, ITU 1998.

[5] 3GPP TS 26.171," Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; General description", 2012

[6] B.Bessette, R.Salami, R.Lefebvre, M.Jelinek and co-authors," The adaptive multirate wideband speech codec (AMR-WB)", IEEE transactions on speech and audio processing, vol. 10, No. 8, Nov 2002

[7] J.C.Bolot," End-to-end frame delay and loss behavior in the internet", in Proc. ACM SIGCOMM, Sept. 1993, 99.289-298

Open Speech Repository.

http:// www.voiptroubleshooter.com/open_speech/.

[8] Johan D.Proakis, Dimitris G. Manolakis " Digital Signal Processing :Principles, algorithm and applications, Third edition" june2000.

[9] S Singh, H P Singh, J Singh, "Spectral Analysis of Speech Quality in VoIP for G.729A and AMR-WB Speech Coders", in Proceedings of 2nd IEEE International Conference on Computational Intelligence, Communication Systems and Networks, 2010, pp:182-186

Trends in Multiple Access Techniques in LTE Mobile Wireless Systems

Parmpreet Kaur
Dept. of CSE
GNDU, RC Jalandhar
parmpreet89@gmail.com

Jyoteesh Malhotra
Dept. of CSE
GNDU, RC Jalandhar
jyoteesh@gmail.com

***Abstract*-Long-Term Evolution (LTE) is the project name of a new, high performance air interface for mobile communication systems. It was developed by the Third Generation Partnership Project (3GPP), and it provides a framework for increasing data rates and overall system capacity, reducing latency, improving spectral efficiency and cell-edge performance. In this paper we discuss multiple access techniques for uplink and downlink in LTE. A new, OFDMA-based scheme called single carrier frequency division multiple access (SC-FDMA) was developed for the LTE uplink. For downlink we use orthogonal frequency division multiplexing (OFDMA). Here we discuss these techniques in detail in this paper.**

Keywords Used: LTE, Multiple access Techniques, OFDMA, SC-FDMA, PAPR

I. INTRODUCTION

Long-Term Evolution (LTE) commonly marketed as 4G LTE, is a standard for wireless data communications technology. The goal of LTE was to increase the speed and capacity of wireless data networks using new DSP (digital signal processing) techniques and modulations that were developed around the turn of the millennium. [1][2] The world's first publicly available LTE service was launched by TeliaSonera in Stockholm and Oslo on December 14, 2009. Airtel launched the LTE service in India in April 2012.

Multiple access techniques are schemes that organize the transmission resources among multi-users to avoid interference between the active user's communications. [4]This paper is organised as follows. In section II Uplink and downlink multiple access techniques have been described along with the pros and cons of each of them. In Section III related work done in LTE has been discussed. In section IV various issues and challenges and future scope has been described. Final section gives a conclusion of the paper.

II. LTE Uplink and Downlink access Techniques

For LTE we use OFDMA as the multiple access technique for downlink and OFDMA variant SC-FDMA for uplink.Orthogonal Frequency Division Multiplexing (OFDMA) for Downlink.OFDMA uses OFDM. It differs from OFDM in the fact that OFDMA allows users to share the available bandwidth in same time slot whereas in OFDM only a single user is assigned a particular time slot [5] OFDMA technique is a hybrid technique of FDMA and TDMA techniques. In this technique each user is provided with a unique fraction from the system bandwidth (OFDM subcarriers) for each specific time slot. [4]

In OFDMA a single high rate data-stream is divided into multiple low rate data-streams and is modulated using subcarriers which are orthogonal to each other. The data to be transmitted is spread over a large number of orthogonal carriers, each being modulated at a low rate. But in OFDM there is a big problem, which is high peak to average power ratio (PAPR), due to which distortion is present in the transmitted signal.[7]

A. Advantages of OFDMA

- The primary advantage of OFDM over single-carrier schemes is its ability to cope with severe channel conditions.
- Channel equalization is simplified because OFDMA uses many slowly-modulated narrowband signals rather than one rapidly - modulated wideband signal.
- OFDMA also facilitates the design of single frequency networks (SFNs), where several adjacent transmitters send same signals simultaneously at the same frequency [6]
- Flexibility of deployment across various frequency bands with little modification to the air interface.
- Averaging interferences from neighbouring cells and within the cell are averaged by using allocation with cyclic permutations.
- Offers Frequency diversity by spreading the carriers all over the used spectrum.[12]

B. Drawbacks of OFDMA

- High peak-to-average ratio
- Sensitive to frequency offset, hence to Doppler-shift as well.[6]

The high data rates of the LTE standard do not only need wider bandwidth but also a more advanced modulation technique. While Orthogonal Frequency Division Multiplexing (OFDM) is considered to be optimum to fulfil the downlink transmission requirement, the high Peak-to-Average Power Ratio (PAPR) property of OFDM makes it less favourable for the uplink transmission. Instead, the Single-Carrier FDMA technique is used. This technique is also known as DFT-Spread OFDM (DFTS-OFDM) where DFT is an acronym for Discrete Fourier Transform. [8]

In SC-FDMA (DFT-spread OFDMA) time domain data symbols are transformed to frequency domain by DFT before going through OFDMA modulation as illustrated in Figure 1. The orthogonality of the users comes from the fact that each user occupies different subcarriers in frequency domain, similar to the case of OFDMA. The only different between OFDMA and SC-FDMA is introducing an additional DFT module at the transmitter side and IDFT at the receiver side. OFDMA performs modulation and demodulation operations in the frequency domain and SC-FDMA performs these operations in the time domain.[10]

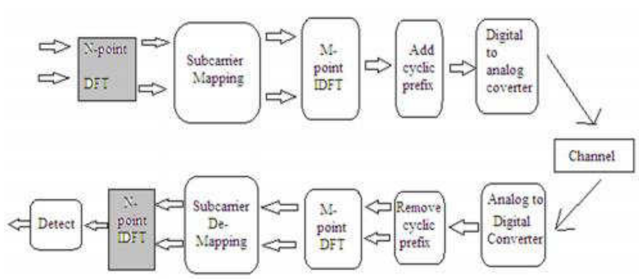


Fig1. OFDMA and SC-FDMA transceiver

The grey boxes shown in the diagram are used specifically for SC-FDMA and are not used in OFDMA.

On the transmitter side, for OFDMA, a user's M information symbols are mapped to allocated subcarriers. An N-point inverse discrete Fourier transform (IDFT) is then performed, and CP is inserted. For SC-FDMA, an M-point DFT is first applied to the user's information modulation symbols. The output of the DFT is mapped to a set of subcarriers which then go through the OFDM transmitter processing of IDFT and CP insertion. The time-domain samples of SC-FDMA waveform are versions of the user's original modulation symbol sequence, thereby resulting in a low PAPR waveform. [11] The discrete Fourier transform (DFT) block, is used for spreading each data symbol over all the used subcarriers. The insertion of this allows the data symbols to be transmitted serially in time domain. [9]

A. Advantages of SC-FDMA

- Reduces battery power consumption due to low PAPR
- Requires a simpler amplifier design
- Improves uplink coverage and cell-edge performance
- Less sensitive to frequency-selective fading as compared to OFDMA

III. Related Work

Burcu Hanta [10] in December 2009 described single-carrier frequency division multiple access (SC-FDMA) scheme which has been selected as the multiple-access scheme of the LTE uplink. SC-FDMA and orthogonal frequency division multiple access (OFDMA) are similar schemes except for a discrete Fourier transform. The data symbols-subcarrier mapping is not one-to-one but instead each subcarrier carries a part from each data symbol transmitted at that time instant.

Mashhur Sattorov [12] in 2009 presented an overview of OFDM spread spectrum scheme's pros and cons issues and to highlight hot solutions for disadvantages.

Ashraf Awad Elkarim Widaa Ahmed [4] in May 2010 presented various multiple access techniques available for data transmission in LTE. Author explains the techniques used for uplink and downlink data transmission and highlights the advantages of each of the techniques.

Motorola [2] in July 2010 gave a journal in which it explained why the need for LTE arose and what where the factor that contributed to its development and evolution. It then explains various stages of development of LTE which led to reduced traffic and better data rates.

Gilberto berardinelli [9] in August 2011 explained DFT-s-OFDM as a suitable modulation and coding. The use of DFT-s-OFDM has been shown to require lower power consuming than OFDM, thus improving the cell coverage or reducing the power consumption of the UE.

Rafal Surgiewicz [8] in 2011 explained that in order to support such a high data rate and avoid the high Peak-to-Average Power Ratio (PAPR) disadvantage in Orthogonal-FDMA, Single Carrier Frequency Division Multiple Access (SC-FDMA) is used in the uplink communication in LTE cellular systems. Here an overview of LTE and the LTE uplink transmissions is given, especially the SC-FDMA technique.

Min Hua [11] in 2013 derived the fundamental performance difference between OFDMA and SC-FDMA, and presents the general performance comparison between them. Analytical results show that

OFDMA performance upper bounds SC-FDMA in the sense of capacity.

Ishan Girdhar [7] in May 2013 explained that LTE has adopted DFT-spread OFDMA technique as the uplink multiple access scheme which use single carrier modulation. Here author shows the PAPR performance of DFT-spreading technique. Also included is the effect of increasing the number of subcarriers assigned to each user on the PAPR performance of DFT-spreading technique.

IV. OPEN ISSUES AND FUTURE SCOPE

In this section important issues extracted from literature survey done in previous section have been described.

- OFDMA is Sensitive to Doppler shift & frequency Selective Fading
- OFDMA has high peak-to-average-power ratio (PAPR), which results in poor efficiency.
- Dealing with co-channel interference from nearby cells is complex in OFDM
- OFDM systems have tight spacing of subcarriers, which results in loss of orthogonality due to frequency errors.
- The fast channel feedback and adaptive sub-carrier assignment is complex in OFDMA.
- PAPR is still an issue in SC-FDMA, when pulse shaping is considered.

LTE is faster, newer and more efficient than other It is being assumed that LTE will provide optimum solution to provide high speed broadband connectivity in high population density areas. Once its practical implementation is realized it will ensure high speed data connectivity to user though reliable links and will enhance the user's experience of browsing.

V. CONCLUSION

Multiple access techniques need a lot of improvement in futuristic development of LTE networks. Work done by researchers in this area has been the main focus of this paper. It was found that LTE aims to provide high speed data access and reduce traffic on the network by ensuring fast communication between user agents. OFDMA is used as downlink multiple access technique in LTE. It is a combination of FDMA and TDMA but it has the problem of having high Peak Average Power Ratio PAPR. This disadvantage of OFDMA is removed by SC-FDMA which is used as Uplink multiple access technique in LTE. It reduces the problem of high PAPR to a large extent. Finally it can be concluded that there is ample room for improvement in access techniques in general and specifically in context of LTE networks.

REFERENCES

[1] LTE Encyclopedia "An Introduction to LTE. 3GPP" (December 3, 2010).

[2] Motorola "Long Term Evolution (LTE): A Technical Overview" (July 3, 2010)

[3] Ashraf AwadElkarim Widaa Ahmed, Ahmed Hamza Ibrahim Makki, Performance "Evaluation of Uplink Multiple Access Techniques in LTE Mobile Communication System" (May 2010)

[4] http://www.anritsu.com/en-GB/Media-Room/Newsletters/files/anritsu_lte_guide.pdf (Anritsu "LTE Resource Guide")

[5] http://www.tutorialspoint.com/lte/lte_ofdm_technology.htm

[6] Ishan Girdhar, Charanjeet Singh, Amit kumar "Performance Analysis of DFT Spread OFDM Systems", Volume 2, No.5, International Journal of Advances in Computer Science and Technology (May 2013)

[7] Rafal Surgiewicz, Niklas Str¨om, Anser Ahmed, Yun Ai "LTE Uplink Transmission Scheme" (2011)

[8] Gilberto berardinelli, troels b.Sørensen, preben mogensen, Kari pajukoski "Transmission over multiple Component carriers in LTE uplink", IEEE Wireless Communications (August 2011)

[9] Burcu Hanta "SC-FDMA and LTE Uplink" (02. December 2009)

[10] Min Hua, Bingying Ren, Mao Wang, Jun Zou, Chunliang Yang, and Tingting Liu "Performance Analysis of OFDMA and SC-FDMA Multiple Access Techniques for Next Generation Wireless Communications" (IEEE 2013)

[11] http://www.academia.edu/217868/Pros_and_cons_of_OFDM (Mashhur Sattorov, Sang-Soo Yeo, Heau Jo-Kang "Pros and Cons of Multi-User Orthogonal Frequency Division Multiplexing" (2009)

[12] Mooray Rumney "3GPP LTE: Introducing Single-Carrier FDMA"

Enhanced S-MAC Protocol for Wireless Sensor Networks

Jasmeet Kaur
CSE Department ,PEC University of Technology
Chandigarh, India
Jasmeet.saini1990@gmail.com

Trilok C. Aseri
CSE Department, PEC University of Technology
Chandigarh, India
a_trilok_chand@yahoo.com

***Abstract*— Wireless Sensor Networks (WSNs) consist of large number of sensor nodes. These nodes are distributed in a particular area for the purpose of collecting data/queries etc. WSNs have played an important role in applications such as industrial automation, smart buildings and intrusion detection so on. The primary concern in WSNs lies in minimizing the energy consumption between sensor nodes while achieving the desired network operation, thus extending the life time of sensor nodes. In this paper, we proposed an Enhanced version of S-MAC. This proposed protocol is better in terms of energy efficiency because it is achieved using dynamic listen time scheme. Performance evaluation is carried out using NS2 simulation platform to verify the protocol improvement.**

Keywords—Medium Access Control (MAC), Wireless Sensor Network (WSN), Energy Efficiency, Sensor-MAC(S-MAC).

I. INTRODUCTION

Wireless Sensor Network (WSN)[1][2]is a type of intelligent wireless networks composed by tens to thousands of sensor with sensing capabilities, communication and ability to self organize. Since the sensor nodes are expected to operate autonomously with small batteries for a number of months or years, energy efficiency is a fundamental criterion in the design of WSN protocols. A major power consuming component of a sensor node is the radio, which is controlled by the MAC protocol. Therefore, an efficient MAC protocol increases the lifetime of a sensor network to a great extent. In addition, the MAC layer controls how nodes share the wireless medium. An efficient MAC protocol can reduce collisions and increase the achievable throughput, providing flexibility for various applications.

We suggest that integrated the idea of dynamic listen time into S-MAC can provide significant performance achieving. The main aim of this scheme is to increase the lifetime of sensor nodes by varying listen time. In basic S-MAC the listen time is fixed and sleep time varies for handling the low traffic loads but it suffers from high energy consumption and less network lifetime issues

The rest of the paper is organized as follows: Section 2 reviews the related work. Section 3 presents the details of our proposed scheme. Performance evaluation was carried out via simulation and results are given in section 4. Finally, section 5 concludes the paper.

II. RELATED WORK

In SMAC (Sensor-MAC) [3] [10] there are periodic sleep and listen schedules of nodes. Each cluster has an independent schedule composed of three periods: SYNC, DATA, and SLEEP. Clocks are synchronized by nodes within the same cluster in SYNC period. Nodes contend for exchange of Request-to-Send (RTS) and Clear-to-Send (CTS) frames in the DATA period when they have packets to send. Nodes that are not involved in the communication go back to sleep at the start of the START period. Other nodes return to sleep after they finish transmission of data packets and acknowledgment (ACK) frame. The main problem of S-MAC is that in each cycle packets can be forwarded by one hop only. To increase the number of hops, SMAC with adaptive listening is introduced which increases the hops to two instead of one.

To improve the efficiency, TMAC (Timeout-MAC) [2] [4] is introduced in which length of the active period is not fixed. Here nodes stay awake for a certain amount of time until no activation event has occurred. Future Request-to-Send (FRTS) packet is used in T-MAC, in which nodes notify the target receiver that it cannot access the medium at the current time. The main advantage of using TMAC is that it handles variable load due to dynamic sleeping schedule but also has a disadvantage of early sleeping problem.

RMAC (Routing Enhanced MAC) [5] uses multi hop forwarding in single operational cycle by shifting the data transmission to the sleep period. In RMAC cycle can be divided in to three stages: SYNC, DATA, and SLEEP. Separate protocol is used in SYN period to synchronize the clock of all the nodes which are in the same cluster. To initiate the communication with the downstream nodes that are multiple hops away, a control frame is send during DATA period. RMAC uses control frames, named PIONs (Pioneer frames) instead of using a pair of RTS (Request to send) and CTS (Clear to send) frames between two nodes. A PION is used as RTS frame as well as CTS frame. A single PION is used to confirm receipt traffic from upstream and downstream nodes.

In DW-MAC (Demand Wakeup-MAC) [6] one-to-one mapping function between the DATA period and the SLEEP period is done to ensure collision free transmission in SLEEP period. Nodes are wake up on demand during the sleep period of a cycle in order to transmit or receive a packet. Effective channel capacity is increased during a cycle as traffic load increases. This allows DW-MAC to achieve low delivery latency under a wide range of traffic loads which includes both broadcast and unicast traffic. Low latency, high power efficiency and high packet delivery are the main factors that contribute in the success of DW-MAC.

Applications where data are delivered from multiple sources to a sink, D-MAC (Dynamic MAC) [7] is used. Staggering of active/sleep schedule of nodes is done based on data gathering tree. Staggering is done so that packets can flow continuously toward the sink. Nodes contend for sending to their receivers which have same depth and same offset. When one node wins channel access its neighboring nodes of the same level lose their chance of transmission. More-to-Send (MTS) frame is introduced in order to increase the number of active slots.

To provide minimum end-to-end latency with energy efficient data transmission Q-MAC (Query Based-MAC) [8] is used. Static sleep schedule is followed in Q-MAC when there is no query and dynamic sleep schedule is followed in Q-MAC when a query is transmitted. Each node acquires readings from sensors based on the type of query and the data flow can be either from single destination or from multiple destinations after a query has been disseminated.

A. S-MAC overview

S-MAC [2] [10], grouping all sensor nodes in flat manner without coordination or control by a cluster head as in other scheduled protocols. The clustering process is done by synchronizing the sleep schedules of neighboring sensor nodes to listen and sleep together at the same time. Using such synchronization scheme, S-MAC forms virtual clusters around common schedules, not strict clusters. This may lead to reducing the control overhead massages. Moreover, no cluster head involves in this clustering process, so the sensor nodes will communicate with neighbors of the same or different clusters in the form of peers based on predefined cycle periods, where the time line of each node divided into cycles of active (listen interval) and inactive (sleep interval) states. As shown in Fig. 1, nodes can only hear their immediate neighbors (e.g., node B can only hear node A and node C). Because node C can overhear the CTS sent by node B, it goes back to sleep at the beginning of the SLEEP period but wakes up at the end of the current transmission. Node B can therefore immediately forward the data packet to node C instead of waiting for the next cycle. For collision problem, S-MAC utilizes a combined scheduling and contention scheme to avoid collision. It avoid overhearing by allowing interfering nodes go to sleep after they hear control packets. The main concern of S-MAC is the energy efficiency while dealing with the other important performance metrics including fairness, latency, throughput, and bandwidth utilization as secondary objectives.

Features of S-MAC:

- Periodic listen and sleep- Periodic sleep accounts major share in the energy saving achieved in S-MAC. In listen period node switches it radio ON to communicate with other nodes and when listen period expire the node switches to sleep state by turning OFF the radio.
- Synchronization- Nodes operating under S-MAC follow synchronized periodic sleep and listen so that all neighboring nodes sleep at the same time and wake up at the same time. The synchronization between nodes is achieved by SYNC packet exchange.
- Schedule- Nodes operating S-MAC has two type of schedule, primary and secondary schedule. The schedule that node receives for the first time is considering as primary schedule and the schedule that node receive from the neighbor after receiving primary schedule is called secondary schedule.
- Overhearing avoidance and collision avoidance- S-MAC puts interfering nodes in sleep after receiving RTS/CTS packet that is not destined to them. RTS/CTS mechanism also reduces the possibility of collision.

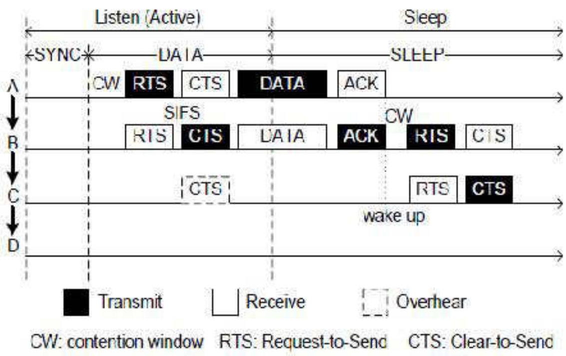


Fig. 1 S-MAC Overview

III. PROPOSED SCHEME

To overcome the energy loss of S-MAC nodes and to make them more energy efficient, we propose a scheme of dynamic listen time. It aims to reduce idle listening of all nodes, considering the fact that sleep time is already dynamic in S-MAC protocol. In dynamic listen time scheme of S-MAC, each cluster has a independent schedule composed of three periods: SYNC, DATA, SLEEP. All nodes of the same cluster wake up at the beginning of the SYNC period to synchronize clocks with each other. Then nodes with packets to send contend for exchange of request to send (RTS) and clear to send (CTS) frames in the DATA period. Nodes that are not involved in communication return to sleep at start of the sleep period, other nodes return to sleep after they finish transmission of data packets and acknowledgement (ACK) frames.

It is noticeable that in each cycle the energy of each node reduces to some extent. When the nodes left with energy less than the threshold energy, the cycle time with dynamic listen time and sleep time of the nodes varies. On reducing listen time and increasing sleep time, S-MAC protocol become efficient in energy consumption and the lifetime of nodes increases.

IV. PERFORMANCE EVALUATION

In this section, we present the simulation environment and the results for the proposed scheme based upon average energy consumption as the main task.

A. Simulation Setup

We have simulated WSN under ns2.34 [9] plat form. In our simulations, each node having fixed amount of energy (100 Joule). The other setting is given in Table-1. We simulated each sensor network once with S-MAC protocol and then with enhanced S-MAC protocol. In these studies we record the remaining energy of each node from the trace file generated by NS-2 and sum of the total energy of all the nodes left is also calculated. From the trace files, total energy consumption was computed.

B. Simulation parameters

• *Network topology*: We consider a grid-based Wireless Sensor Networks consisting of 67 nodes. A single sink can only communicate with a limited number of sensor nodes, so large deployments may include multiple sinks each communicating with virtual clusters of sensor nodes. The function of each sink node is to receive data from its close nodes. All other nodes are normal nodes performing the functions of sensing, receiving, and transmitting data packets.

• *Traffic model:* Every non-sink nodes keep sending a 100 bytes packet to the sink node with varied message inter-arrival time (from 1 second to 10 seconds).

• *Energy model:* The sink nodes as well as all other nodes used in simulation have finite energy (100 joules). The default energy model of wireless node in NS-2 is used in simulations. The different levels of energy consumption for different operations are defined in the following:

• *Routing:* Static routing (AODV protocol) was used in the simulations. Using static routing allowed us to focus on MAC layer energy efficiency and to get rid of the effect of routing protocol.

Table I. Simulation parameters

Parameter	***values***
• Simulator	NS 2.34
• Simulation time	1000 sec
• Number of nodes	67
• Topology	Grid
• Routing protocol	AODV
• Applications	CBR
• Traffic rate	0.1 to 1

V. RESULTS AND DISCUSSION

We analyzed the impact of varying traffic load by changing the packet transmission rate at the source node for each performance metric. We plotted sum of energy left of all the nodes against the time (no of rounds). The simulation runtime is 1000 seconds with packet size of 100 bytes.

• Energy consumption

Energy efficiency is the most important factor for wireless sensor networks. As we expected, we can see from Fig. 2, the Enhanced S-MAC protocol consume less energy than S-MAC. This is because according to new dynamic listen time scheme, listen period decreases when the sum of energy of nodes falls below the threshold energy limit.

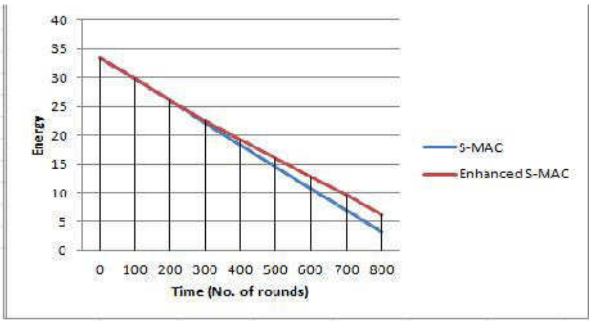


Fig.2 Average energy consumption

VI. CONCLUSION

Due to finite energy resources available in sensor nodes efficient energy consumption in wireless sensor networks remains a challenging problem. Thus it has direct responsibility to the network lifetime.

We have enhanced the S-MAC protocol by introducing dynamic listen time scheme. Performance evaluation is carried out via NS2 simulation experiment shows that the total system capacity is increased. The new scheme achieved significant improvement in terms of energy efficiency with some throughput or latency trade-off.

REFERENCES

[1] J. Yick, B. Mukherjeeand D. Ghosal, "Wireless sensor network survey," Computer Networks, Volume 52, Issue 12, pp. 2292-2330, August 2008.

[2] Pei Huang, Li Xiano, Soltani S, M.W Mutka, Ning Xi, "The Evolution of MAC Protocols in Wireless Sensor Networks : A Survey," IEEE Communication Survey and Tutorials, Volume:15, Issue: 1, pp. 101-120, 2013.

[3] SONG Wen-miao, LIU Yan- ming, ZHANG Shu-e, "Research on SMAC protocol for WSN," Wireless Communications, Networking and Mobile Computing, WiCOM '08, 4th International Conference, pp.1- 4, Oct 2013.

[4] Faisal Nawab, Kamran Jamshaid, Basem Shihada, and Pin-Han Ho, ``TMAC: Timestamp-ordered MAC for CSMA/CA Wireless Mesh Networks," Computer Communications and Networks (ICCCN), 2011 Proceedings of 20th International Conference, pp.1-6, July 2011.

[5] Shu Du, Amit Kumar Saha and David B. Johnson, "RMAC: A Routing-Enhanced Duty-Cycle MAC Protocol for Wireless Sensor Networks," proceedings of INFOCOM, 26th IEEE international conference on computer communication, pp. 1478-1486,2007.

[6] Y. Sun, S. Du, O. Gurewitz, and D. B. Johnson, "DW-MAC: A low latency, energy efficient demand-wakeup MAC protocol for wireless sensor networks," in Proc. MobiHoc, pp. 53–62, 2008.

[7] G. Lu, B. Krishnamachari, and C. S. Raghavendra, "An adaptive energy efficient and low-latency MAC for data gathering in wireless sensor networks," in Proc. IPDPS, 2004.

[8] N. Vasanthi and S. Annadurai, “Energy efficient sleep schedule for achieving minimum latency in query based sensor networks,”in Proc. IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing, vol. 2, pp. 214–219, June 2006.

[9] Network Simulatorns-2, http://www. isi-du/nsnam/ns

[10] Sarika Khatarkar, Rachana Kamble, “Wireless Sensor Network MAC Protocol: SMAC & TMAC,”Indian Journal of Computer Science and Engineering (IJCSE), vol. 4, Issue 4, pp.304-310, 2013.

Sensor Field Optimization Based Enhanced Developed Distributed Energy-Efficient Clustering for Heterogeneous WSNs

Arshdeep Kaur
Dept. of Computer Science & Engineering
Sai Institute of Engineering & Technology,
Manawala, Amritsar

Gagandeep Singh
Dept. of Computer Science & Engineering,
Sai Institute of Engineering & Technology,
Manawala, Amritsar.

***Abstract*--This paper focuses on the various clustering protocols used in heterogeneous wireless sensor networks. Wireless sensor networks (WSNs) are becoming popular in real world applications. Due to the features of the resource-constrained and battery-aware sensors; in WSNs energy utilizationhas found to be a major interesting subject of research. WSNs compose battery-powered nodes which are connected with the base station to for certain action or task. As sensor nodes are battery-powered i.e. will become dead after the consumption of the battery which is also called lifetime of WSNs. So using the energy in well-organized way may result in prolonging the lifetime of the WSNs. Stable election protocol (SEP) has found to be the one of the best clustering protocols for 2-level heterogeneous sensor network. Many enhancements have been done on SEP by optimizing the cluster head selection. The overall objective of this work is to find the short comings of available heterogeneous WSNs.**

I. INTRODUCTION TO WSN

A Wireless sensor network (WSN) is a spread wireless set of connections which is a grouping of independent devices like sensor nodes, routers, gateway where each sensor node having connectivity to sensor. Data is transmitted in a scattered way from beginning to end these nodes and monitoring of the physical or ecological situation is done from remote place. The nodes are deployed in an unplanned way and be in touch wirelessly. In [11] five key features have been told which want to be considered when rising WSN solutions.

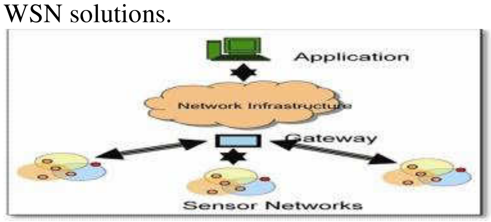


Fig.1. A Wireless Sensor Network

A. *Sensor Node*

Each node have a number of units like, a sensing unit (which can sense physical parameters like heat, noise, vibration and moisture), a processing component (one or more microcontrollers, CPUs where processing of sensed information is done), have a RF transceiver (which consists of a spreader for sending information and a recipient for receiving information from its 1 hop neighbours),a storage space component which may contain many types of memory (program, information and explode memories), have a power Management component (e.g., battery and solar cells) -for maintaining small power utilization.

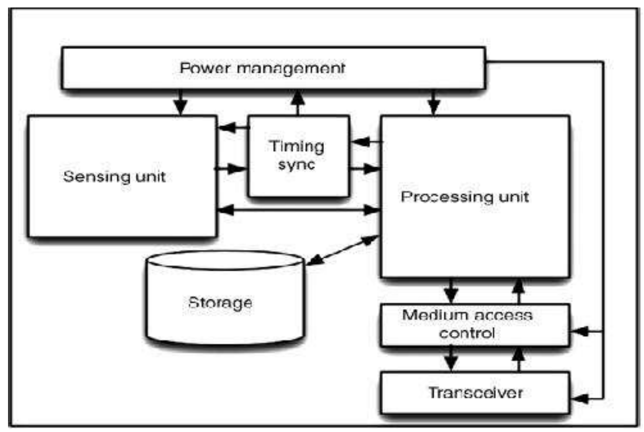


Fig.2. Simple schematic of a sensor node {11]

B. *Architecture of WSN*:

A WSN consists of a huge number of nodes which may be tightly or arbitrarily deployed in an area in which we have interest. There is Base Stations (BS) situated to sensing area. The base station having major function in WSN as sink send queries to nodes while nodes sense the asked queries and send the sensed information in a joint way reverse to Base station. Base station also serves as an entrance for outer surface system i.e. Internet. So the collection of information and send only relevant data to customer via internet is done by Base station.

As we know nodes have little batteries which are hard to modify or recharge. So we have to follow such structural design (having a smaller amount transfer and concentrated communication space) to raise power saving. There are positive structural design like flat-network architecture and hierarchical network architecture

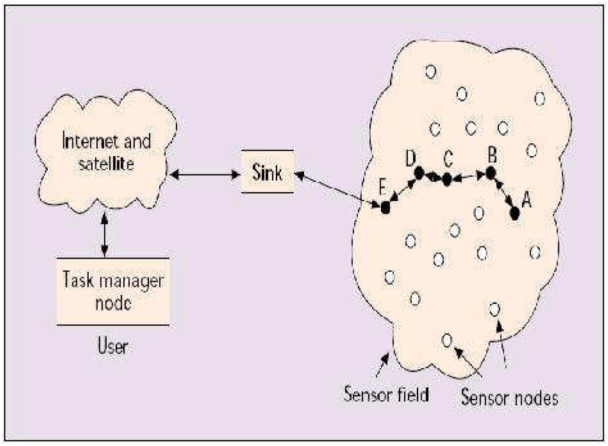


Fig.3. WSN communication architecture [6]

C. *Difference between WSN and ad-hoc network:*

Various people have misunderstanding between WSN and ad-hoc-network. Follow are the main differences between WSN and ad-hoc network.

- Amount of sensor nodes is high in WSN
- Sensor nodes are tightly deployed in WSN
- Nodes are prone to failure.
- Active topology in WSN
- Transmit communication is made in WSN while in ad-hoc end to end.

D. *Operational Challenges in WSN*

Following outfitted challenges of WSN has been described [11].

- Unattended process –No human being contact once nodes have been deployed in an area. So there is a short of reconfiguration in case of any mistake.
- Energy effectiveness: Life span is critical issue in WSN; as nodes are connected with tiny batteries. Energy effectiveness must be measured.
- Limited storage space and computation

E. *Applications of WSN:*

WSN is normally motivated from the armed operation's gives improved outcome to in contrast with conventional wired network. In WSN nodes can be deployed in unmanned region. There are also lots of fields of WSN application like farming, home study, location monitoring, structural health monitoring, heavy manufacturing monitoring, and security monitoring [11].

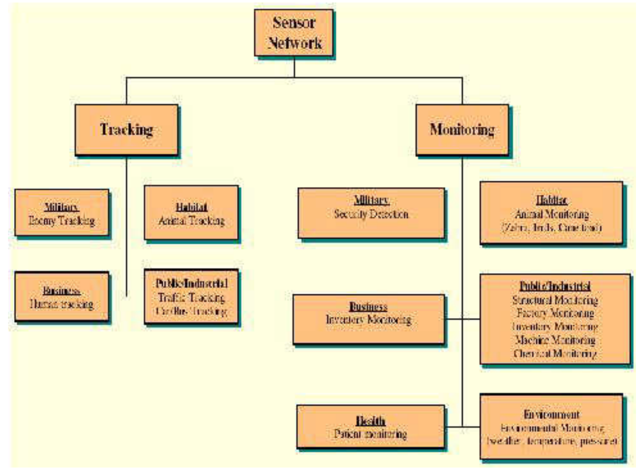


Fig.4. Applications of WSN

F. *Mechanisms in WSN*:

To fulfill the WSN applications, we want some mechanisms to meet there requirements. These mechanisms can be

- Multi-hop wireless communication
- Energy-efficient operation
- Auto-configuration
- Collaboration & in-network processing

In [6] two traditional approaches for above mechanisms have been described.

II. DIRECT TRANSMISSION (DT)

- In DT, communication of information is done to the base station directly nodes, which may cause a condition in which sensor nodes that, are distant away from the base station would expire first.

III. MINIMUM TRANSMISSION ENERGY (MTE)

- In MTE, information is running scared in such a method that communication power price would lowest. Under MTE, sensor nodes that are near the base station having the advanced possibility to be act as relay in contrast with sensor nodes that are distant from the base station. Therefore, the sensor nodes, which are near to base station, can expire first.

Therefore, it is understandable that in both approaches DT and MTE, an area of the sensor field will not be monitored for definite interval of the life span of the WSN and it would upset the sensing method of the field. These approaches do not warranty well-balanced division of the power load between sensor nodes of the sensor network.

Therefore, for improved results, we require to use various additional approaches like routing techniques. There can be numerous routing techniques. Clustering is combination of the sensor based on various criteria.

IV. GAPS IN LITERATURE

The study has exposed that the every WSNs protocol has various limitations; i.e. no one is perfect in all case. In the large amount of the presented literature the majority of the researchers have ignored at least one of the following:

1) Boundary sensor nodes: Nodes which are deployed on the boundaries are known boundary nodes. These sensor nodes have highest space with respect to the location of themselves and the sink. So use extra power than other sensor nodes i.e. sensor nodes closed to the sink. So it will turn into the bottleneck of the sensor network.
2) Region segments: The well-organized deployment of the nodes wants to be required. It is only achievable by using the region based segmentation i.e. separating the sensor field between segments based upon the distant nodes and close to nodes.

A. Problem Statement

This Paper enhances the presentation of the ECRSEP using enhanced cluster head selection and segment based optimization. The sensor field optimization is a serious problem in WSNs and has been ignored by numerous researchers. As in numerous real-time sensing fields the sensor nodes on the corners i.e. on the segment boundaries will become expired before time because no unique protection is exist for them. Therefore, in this research effort has focused on the segment based optimization by separating the sensor field between advance and normal segments. The inspiration at the back this sensor field optimization is to prolong the time when the first sensor node expires. Because in normal sensor nodes which were exist on the boundaries may become expire before time because the space among them and the base station is more so they use more power so ultimately will become expire soon. In ECRSEP thresholding will be used to balance the load between sensor nodes; but in current effort decision tree will be used in its place of thresholding.

The ECRSEP has used dissimilar probability function for cluster head choice for selecting the best cluster head by using the remaining power and also use single hop for transmission. But ECRSEP has ignored the space among sink and cluster head and also use multi hop for transmission. Therefore the planned method will be more capable than the existing techniques. The contrast will also be drawn between planned and existing stable selection protocols based upon network life span, stable time, first sensor node expire time, throughput etc.

B. Objective:

The overall objective of the thesis would be to enhance the performance of the present ECRSEP. Following are the different objectives of this thesis.

1) To calculate the performance of various stable selection protocols hierarchy.
2) To optimize the sensor filed by separating the sensor field between advance and normal segments. Where an advance segment contains the advance sensor nodes and normal segment contains normal sensor nodes.
3) To enhance ECRSEP by using the decision tree in place of thresholding.
4) The multi-hop technique will be used for packets transmission among sink and cluster heads.
5) The contrast will also be drawn between proposed, ECRSEP and other SEP protocols based upon network life span, stable time, first sensor node expire time, throughput etc.

V. PROPOSED ALGORITHM

Following are the various steps required to achieve the objectives of this research work.

Step1: First of the entire initialization of the sensor network will be complete by setting up a variety of constants and variables of the sensor network. Like diameters of network, distance of sink from the sensor network, no of sensor nodes, probability of a sensor node to become cluster head, power supplied to each sensor node, transmitter power per sensor node, receiver power per mode, amplification power, distance between cluster head and sink etc. However as this research work focuses the sensor field optimization so sensor nodes will be placed according to their initial power i.e. advance and normal.

Step2: For all nodes I if any sensor node has 0 or negative power then sensor node will be set as expire and continue for next step.

Step3: If node is not chosen as cluster head in last 1/p rounds then go to step 4 else go to step 2.

Step4:If sensor node (I). Power>avg_residual_power and node (I). Distance<avg_distance then go to step 5 else go to step 2.

Step5: At this time select sensor node as cluster head and Tx and Rx operations will come in action for transfer and receiving the packets among cluster head to sink using multi-hop approach.
Step 7: Evaluate power dissipated and update the remaining power of each sensor node (I) and go to step 2 again.

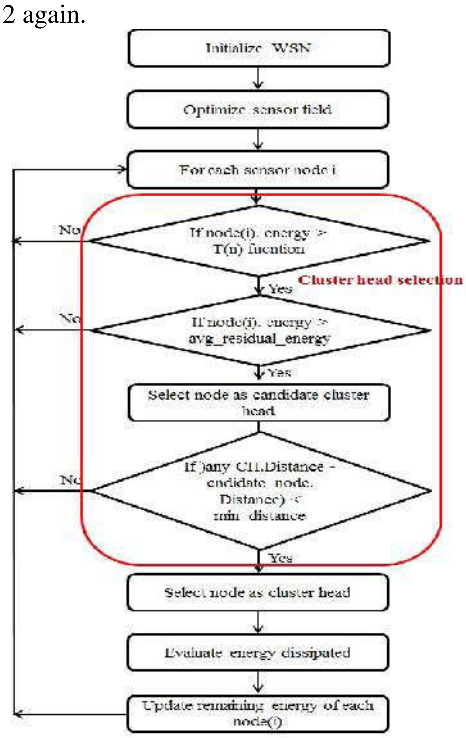


Fig.5. Flowchart

A. Planning of work

The current research work is planned to be finished in the following three stages.

- Literature review
- MATLAB
- Documentation

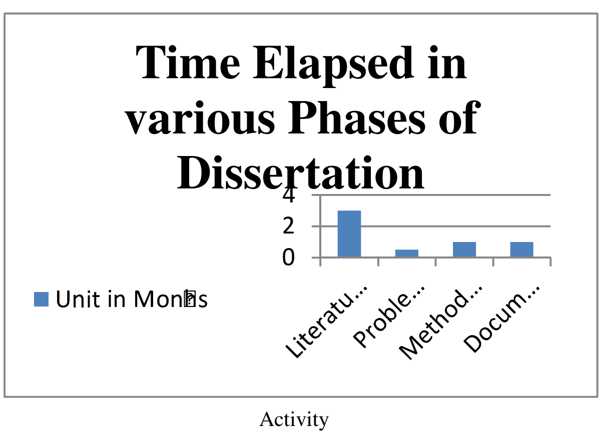


Fig.6. Time Line Graph

B. About MATLAB Simulator:

The instrument or atmosphere in which planned simulation model would be designed is MATLAB (Matrix Laboratory).MATLAB, was developed by Math Works Inc. MATLAB is written in C, C++ and Java. It gives cross platform operating system. Individual can manipulate matrix, implement algorithms, design functions and can create a lot of interfaces and simulation scenarios. It is based on mathematical model. MATLAB provides inbuilt toolbox, which are optional. MATLAB afford drag and drop facility to design scenario. Individual can design simulation scenario as well as make graphs also. It provides analyzer also which helps in examination of scenario. Contrast of different scenario can also be done.

VI. CONCLUSION & FUTURE WORK

The survey has shown that the ECRSEP has used different probability function for cluster head selection for selecting the best cluster head by using the residual energy and also use single hop for communication. But ECRSEP has neglected the distance between base station and cluster head and also use multi hop for communication. Therefore in near future we will propose a modified ECRSEP which will consider the residual energy and also use multi-hop communication to enhance the results further.

REFERENCES

[1] Heinzelman, W., Anantha, P. and Balakrishnan, H. "An Application-Specific ProtocolArchitecture for Wireless Microsensor Networks" IEEE Transactions on Wireless Communications, in IEEE Vol. 1,No 4, October 2002,pp.661-670

[2] Yang, Z., Liu, J., and Chen, X. "An Optimal Mechanism of LEACH Protocol for Wireless Sensor Networks", IEEE ISECS International Colloquium on Computing, Communication, Control, and Management, January 2009, pp.254-256

[3] Melese, D. , Xiong, H. and Gao, Q. "Consumed Energy As A Factor For Cluster Head Selection In Wireless Sensor Networks ,IEEE Wireless Communications Networking and Mobile Computing (WiCOM),6th International Conference, Feb 2010,pp. 1-4

[4] Chen, G.,Zhang, X., Yu, J. and Wang,M. "An Improved LEACH Algorithm Based On Heterogeneous Energy Of Nodes In Wireless Sensor Networks"IEEEInternational Conference on Computing, Measurement, Control and Sensor Network,January 2012,pp.101-104

[5] Khan,A.,Javaid N., Qasim, U., Lu, Z. and Khan, Z. "HSEP: Heterogeneity-AwareHierarchical Stable Election Protocol For WSNs",IEEE Seventh International Conference on Broadband, Wireless Computing, Communication and Applications, in May 2012,pp.373-378

[6] Islam, M., Matin.M. and Mondol,T."Extended Stable Election Protocol (SEP) for Three Level Hierarchical

Clustered Heterogeneous WSN",IEEE Wireless Sensor Systems (WSS 2012), IET Conference,June 2012,pp.1-4

[7] Singhal, S., Gankotiya,A., Kumar, A., Shikha,V. "An Investigation of Wireless Sensor Network: A Distributed Approach in Smart Environment"IEEE Second International Conference on Advanced Computing & Communication Technologies, in July 2012,pp.522-529

[8] Kashaf, A., Javaid, N., Khan, Z. and Khan, I. "TSEP: Threshold-Sensitive Stable Election Protocol For WSNs" IEEE 10th International Conference on Frontiers of Information Technology,Dec 2012,pp.164-168

[9] Beiranvand, Z., Patooghy, A. and Fazeli, M. "I-LEACH: An Efficient Routing Algorithm to Improve Performance & to Reduce Energy Consumption in Wireless Sensor Networks"IEEE5th Conference on Information and Knowledge Technology (IKT),April 2013,pp.13-17

[10] Beiranvand, Z., Patooghy, A. and Fazeli, M. "I-LEACH: An Efficient Routing Algorithm to Improve Performance & to Reduce Energy Consumption in Wireless Sensor Networks"IEEE5th Conference on Information and Knowledge Technology (IKT),April 2013,pp.13-17

[11] Ahlawat, A.and Malik, V. "An Extended Vice-Cluster Selection Approach To Improve V LEACH Protocol In WSN"IEEEThird International Conference on Advanced Computing & Communication Technologies, May2013, pp.236-240.

[12] Ahlawat, A.and Malik, V. "An Extended Vice-Cluster Selection Approach To Improve V LEACH Protocol In WSN"IEEEThird International Conference on Advanced Computing & Communication Technologies, May2013, pp.236-240.

[13] Wang j., Zhang, Z., ShenJ. And Xia F. "An Improved Stable Election Based Routing Protocol With Mobile Sink For Wireless Sensor Networks", IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, June2013,pp.945-950

[14] Zheng, J.and Jamalipour, A. "Wireless Sensor Networks – A Networking Perspective", a John Wiley & Sons, Inc. Publication

A Review of PAPR reduction in OFDM

Vikas Gupta
Dept. of ECE, Adesh Institute of Engineering & Technology, Faridkot, India

Amandeep Kaur
Dept. of ECE, Adesh Institute of Engineering & Technology, Faridkot, India

Rajwant Kaur
Dept. of ECE, Adesh Institute of Engineering & Technology, Faridkot, India

***Abstract--*In recent year, Orthogonal Frequency Division Multiplexing technique plays an important role in wireless digital communication system. OFDM is a multi-carrier modulation technology which enables high capacity of data transmission over a single path. But the major problem in OFDM is the high peak-to-average power ratio due to independent subcarrier. A large PAPR distorts the signal if the transmitter contains non-linear components, which can increase the complexity and reduces the efficiency of power amplifier. Various techniques are surveyed to reduce the PAPR level and complexity. Partial Transmit Sequence is also one of the distortions less technique that improves PAPR performance. However, the high computational complexity is the major disadvantage of PTS due to many IFFT operations. This paper presents various algorithms to reduce the computational complexity.**

Keywords--Orthogonal Frequency Division Multiplexing (OFDM), Peak-to-Average Power Ratio (PAPR), Partial transmits Sequence (PTS).

I. INTRODUCTION

A. *ORTHOGONAL FREQUENCY DIVISION MULTIPLEXING (OFDM)*

OFDM (Orthogonal Frequency Division Multiplexing) is a widely used in wireless & mobile communication system. An OFDM is a part of family of multicarrier modulation technology, which can many signals transmitted at the same time over a single transmission path. The basic concept of OFDM system, a high bit rate is transmitted into a lower bit rate of carriers. Each carriers are orthogonal maintained. Thus, an OFDM signal is produce a complex signal by multiplexing [3]. The OFDM data is generate by taking input data to serial to parallel converter. The Inverse Fast Fourier Transform (IFFT) can bring the required spectrum to time domain and provide the carrier are orthogonal. The Fast Fourier Transform (FFT) is the reverse process of IFFT. The FFT can convert the time domain signal to frequency spectrum and the function of FFT to find the original transmission waveform. The block diagram of OFDM Transmitter and Receiver is shown in Figure 1 [14]. OFDM has a number of attractive features like strengthens to channel fading, flexibility, easy equalization, resistance to impulse interference and capacity to handle strong echoes.

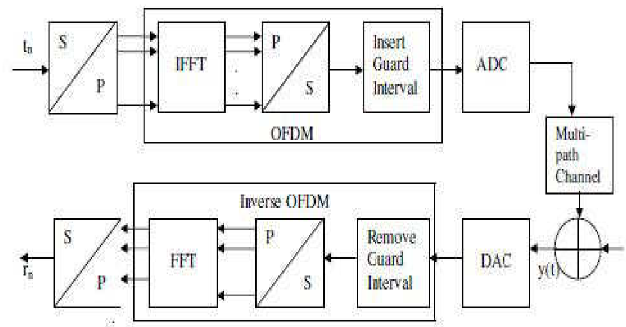


Fig. 1. OFDM transceiver structure [14]

OFDM is one of the very efficient techniques used in high speed digital broadband systems like Digital Television Broadcasting (DTB), Digital Audio Broadcasting (DAB) and Digital Video Broadcasting (DVB). OFDM is a most popular technology of communication system, which has many important applications like Wireless Local Area Networks (WLAN), European Telecommunication Standard Institute (ETSI) and High Performance Radio Local Area Network (HIPERLAN) [14].

II. PEAK TO AVERAGE POWER RATIO (PAPR)

OFDM signal show very high Peak to average power ratio. A high PAPR can cause the complexity increased of the analog to digital converter (A/D) and digital to analog converter (D/A). Therefore, Radio frequency amplifier (RF) can decrease the efficiency and it can operate in non-linear region which damaging the performance of communication system. In OFDM system, an input data block of length N can be written as $X = [X_0, X_1 \ldots\ldots, X_{N-1}]^T$, and each symbol modulating one of a set of subcarrier, $\{ f_n , n = 0,1,\ldots\ldots\ldots, N-1\}$. The N subcarriers are selected to be orthogonal. The data block of the OFDM symbol is given by [1]

$$x(t) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_n \cdot e^{j2\pi n \Delta f t}, \qquad 0 \le t \le NT, \tag{1}$$

PAPR of the OFDM signal is defined as the ratio between the maximum power and the average power during the OFDM signal [12]. Then the Peak to Average Power Ratio is expressed as [1]:

$$PAPR = \frac{\max_{0\le t\le NT} |x(t)|^2}{1/NT \int_0^{NT} |x(t)|^2 dt} \quad \textbf{(2)}$$

The large PAPR can be reduced as the value of *max|x(t)|* decreased [1]. OFDM signal with large Peak to Average Power Ratio is given in figure 2. The PAPR problems are arising by calculation of four sinusoidal signals with different frequency and phase shift logically [12].

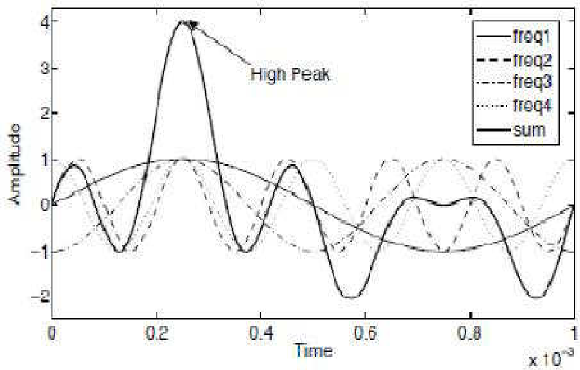


Fig. 2 High Peaks in OFDM signal [12]

Another factor used in PAPR is the Complementary Cumulative Distribution Function (CCDF), which is used to measure efficiency of PAPR technique [1]. The Crest Factor (CF) is defined as the square root of PAPR [14].

$$Crest\ Factor = \sqrt{PAPR} \quad (3)$$

The ccdf expression of the PAPR of OFDM signals can be written as[7]

$$CCDF = max_{0\le t\le NT} \frac{|x(t)|}{E[|x(t)|]} \quad (4)$$

E[|x(t)|] is the average power. In several cases, the large PAPR can be decreased by reducing the value of maximum signal power for the reason that the large value of average power causes interference [7]. There are several techniques to reduced PAPR, which are basically divided into two groups such as signal scrambling techniques and signal distortion techniques. These can be further subdivided into many techniques such as clipping, peak windowing and peak cancellation. Another technique of signal scrambling are block coding, sub- block, selected mapping (SLM) and partial transmit sequence (PTS) [3].

III. PARTIAL TRANSMIT SEQUENCE (PTS)

Partial Transmit Sequence is a distortion less technique based on scrambling rotations to group of subcarriers. PTS is based on the same principle as Selected Mapping (SLM), but gives better performance than SLM. The basic concept of PTS technique is the input data block is portioned into disjoint sub-blocks. The sub-carriers which are transmitted through the sub-blocks are multiplied by weighing value of the phase rotation vector for those sub-blocks. The phase rotation vector is chosen such that the PAPR value is minimized. The block diagram of PTS technique is shown in Fig. 3 [10]. In this, the serial data X is divided into sub sequence by using serial to parallel converter and transmitted in sub blocks and each sub-blocks include N/V non-zero value [10].

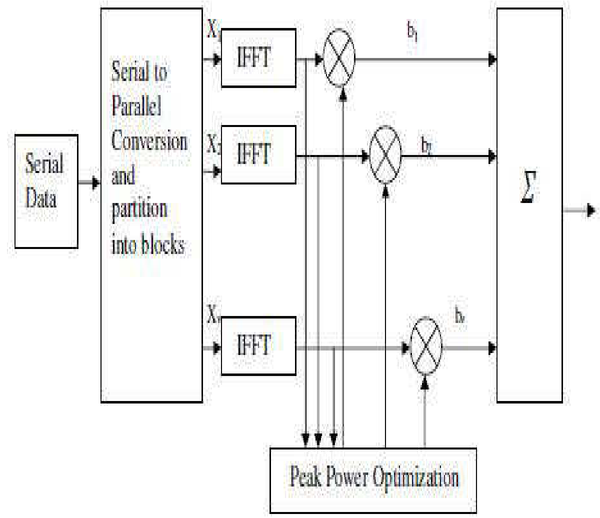


Fig. 3 Conventional OFDM Diagram employing PTS [10]

The Partial Transmit Sequence needed several inverse fast Fourier/wavelet transform (IFFT/IDWT) which produce in large computationally complexity. To decrease its complexity, different algorithms are proposed which are based on PTS [13]:

A. *Particle Swarm Optimization (PSO)*

PSO is a population-based global optimization technique which supported the social manners of bird flocking looking for food. The particle is called the population members which are mass-less and volume-less. All particles represent an explanation of high-dimensional space, its current position and its best position create by its region. The velocity update and position value has two primary operators of PSO technique. In the iteration, each particle repairs its position and velocity as follows [4]:

$$x^i_{k+1} = x^i_k + v^i_{k+1}$$

$$k + c\ r\ (p^i_k - x^i_k) + c_2 r_2 (p^g - x^i_k) \quad (4)$$

Where, x^i_k stand for Particle position
v^i_k correspond to Particle velocity
p^i_k represents Best "remembered" position
c_1, c_2 stand for acceleration constants
r_1, r_2 are random numbers between 0 and 1

The areas of application of PSO are edge detection in noisy images, signature verification, color image segment and QOS adhoc multicast [4].

B. Artificial Bee Colony (ABC)

Artificial Bee Colony is the most successful swarm algorithm based on the behavior of the bees in nature. ABC algorithm is categorized into foraging and mating behavior. The employed bees, onlookers and scouts are three groups in the artificial bee colony to find the optimization problem [4]. In ABC algorithm, a food source position is corresponding to phase vector $b_i = [b_{i1}, b_{i2},\ldots,b_{i(v-1)}]$, i= 1,…..,SN, where S denotes the size of randomly distributed population size. For each employed bee, the new phase vector is expressed by [9]:

$$b_i' = b_i + \phi_i(b_i - b_k) \tag{5}$$

Where ϕ_i is a random number between [-1, 1], b_k is a solution of the region of b_i.

The fitness value of a solution b_i in the population is expressed as [9]:

$$fit(b_i) = \begin{cases} \dfrac{1}{1+fit(b_i)} & if\, fit(b_i) \geq 0 \\ 1 + abs(fit(b_i)) & if\, fit(b_i) < 0 \end{cases}$$

Where, fit(b_i) stand for the PAPR value and is preferred to be at a smallest amount [9].

C. Genetic Algorithm (GA)

GA is an Evolutionary Algorithm which is based on stochastic optimization algorithm.[4] GA initiate with random set of solution called population and which a population of strings can solution of optimization problems. GA involves three principles:

- Selection
- Crossover
- Mutation

GA is valuable and well-organized when the search space is huge multipart, no mathematical investigation is obtainable and conventional investigate method be unsuccessful. But genetic algorithm has some weakness such as it is difficult to working on active data sets and not fit for explaining the restriction optimization problems [4]. GA provides one more solution to reduce the complexity of PTS. It can useful with highly nonlinear problems and non-differentiable function. GA agrees to find the numerical solution to complex problems [8]

D. Differential Evolution (DE)

Differential Evolution is a population-based stochastic parallel swarm evolutionary algorithm. It is used to look for an optimal solution and reducing non linear, non differential and maintain space function. The differential evolution method usually four stages: initialization of the parameter vectors, mutation and difference vectors, crossover and selection [5].

- The differential method starts with an initial solution set, searches for a global optimum point from the feasible region.
- After initialization, the mutation operation of differential evolution creates donor vector to each population member. While doing the mutation, it uses three vectors. The first represents the local best, the second global best which are adaptive in nature and the third selected randomly.
- Once the mutation is complete, the crossover comes into play after generating the donor vectors.
- The function of crossover that it generates the final offspring vector. Crossover be a symbol of a characteristic case of “gene” replace.
- The next step of the algorithm is selection, which determines the population of next generation. It is the best solution of determine the new generation and thus cost function decrease with number of generation [5].

It is similar to Genetic Algorithm, but there will be one difference. The genetic algorithm, mutation is result of small perturbations while differential evolution, mutation is result of arithmetic combination. DE has many benefits like simple to implement, reliable, accurate, robust and high-speed optimization. DE has used todiscover the optimal solution but this process has a time consuming [4]

E. Ant Colony Optimization (ACO)

ACO is based on Swarm Intelligence of meta-heuristic motivated by the foraging manners of ants in the natural. The inspiring source of ACO is the pheromone trail laying which resembles the behavior of ant colony for search shortest path. It means that if the pheromone trail is high then searching the food source also increases. Ant colony optimization has advantage to digital image processing and avoiding the convergence to optimal solution. ACO is prearranged into three major purposes as given [4]:

- Ant Solutions Construct - achieves the solution construction process.
- Pheromone Update – achieves pheromone trail updates
- Daemon Actions –achieves extra updates from a global viewpoint [4].

Ant colony optimization is used to reduce the PAPR. In PTS based ACO method can be implemented by approximately changing the ant location. The modified Ant colony optimization is proposed to discover the optimal angle which helps to reduce the PAPR. The ant colony optimization has compensation of avoid the meeting to a nearby optimal solution [6].

F. *Firefly Algorithm (FF)*

Firefly algorithm is an alternative of swarm-based heuristic algorithm for constrained optimization [4], which is supported of the variation in light intensity. It facilitates the fireflies to travel towards brighter and further attractive position in arrange to achieve optimal solutions [2]. The flashing light is associated with the objective function to be optimized, and formulate new optimization. We can idealize some of the essential flashing properties of suitable fireflies. The firefly algorithm has the following idealized rules [4], [2].

- Every fireflies are unisex and they will travel towards further attractive and brighter.
- The quantity of attractiveness of firefly is comparative to its brightness, thus if there is not a brighter or more attractive firefly then it will travel random ly.
- The light intensity of a firefly is find out by the objective function.

The firefly algorithm has many advantages which make it very efficient for solving the optimization problems [4]. The proposed FF-PTS system supplies approximately the equal PAPR information as thatof the optimal exhaustive PTS, while preserve a low computational load [2].

G. *Bacterial Foraging Optimization (BFO)*

BFO algorithm is a novel evolutionary computation algorithm based on the phenomenon of a bacterial colony. The BFO algorithm is a biologically inspired computing technique which introduced by Passino in 2002. BFO algorithm consist of three most important mechanism namely, chemo-taxis, reproduction and elimination-dispersal [4]. Chemo-taxis are the primary step for a bacterium, which create a cell to cell communication system. The main goal of bacterium is to optimize the best food position in pre-defined iteration. In reproduction, bacteria are set in downward order and divide into two, only the best healthiest bacteria tend to survive and placed in the same food location. In elimination-dispersal, only some bacteria are uninvolved and a few of the bacteria are located in random situation in the environment. BFO based PTS algorithm is a improved arrangement of phase factors for searching OFDM signals. The BFO-PTS algorithm can superior for PAPR reduction, for the reason that the BFO-PTS algorithm has just three control factors which can be effortless used [11].

IV. CONCLUSION

OFDM is a type of multicarrier modulation techniques, which offer high spectral efficiency, low realization complexity and less vulnerability to echoes. OFDM is infinitely used in different communication method. A most important weakness of OFDM is the Peak to Average Power Ratio (PAPR) which distorts the OFDM signal. To recover the PAPR reduction in OFDM systems, different Partial Transmit Sequence supported algorithms are created in this paper. PTS is a non distortion technique which has need of little quantity of redundancy for the recovery of PAPR. But in the PTS technique, the computational complexity raises exponentially with increase in sub-blocks. Partial Transmit Sequence and its based algorithms are analysis in this paper to reduce the complexity of phase factor

V. ACKNOWLEDGMENT

The Author appreciates the help given by the guide Mrs. Mandeep Kaur, Mr. Vikas Gupta and Miss. Amandeep Kaur and Electronics & Communication Department, Adesh Institute of Engineering & Technology Faridkot for the technical Assistance.

REFERENCES

[1]. Abhishek Arun Dash, "OFDM System PAPR Reduction Techniques in OFDM Systems," Bachelor Thesis, Department of Electronics and Communication Engineering National Institute of Technology, Rourkela 2006 – 2010.

[2]. Aman Dhillon, Sonia Goyal, "PAPR Reduction in Multicarrier Modulations Using Firefly Algorithm," International Journal of Innovative Research in Computer and Communication Engineering, vol. 1, Issue 5, July 2013.

[3]. Aman Dhillon, Sonia Goyal, "An Overview of PAPR Reduction Techniques in OFDM system," International Journal of Innovative Research in Computer and Communication Engineering, vol. 1, Issue 5, May 2013.

[4]. Binitha S, S Siva Sathya, "A Survey of Bio inspired Optimization Algorithm," International Journal of Soft Computing and Engineering, vol. 2, Issue 2, May 2012.

[5]. Chien-Erh Weng, Chuan-Wang Chang, Chuang-Hslen Chen, Ho-Lung Hung, "Novel Low-Complexity Partial Transmit Sequences Scheme for PAPR Reduction in OFDM Systems Using Adaptive Differential Evolution Algorithm," Springer Science + Business Media, LLC, 2012.

[6]. Devinder Kumar, Preeti Singh, Jaget Singh, "Complexity Reduction in PTS based OFDM System: A survey," International Journal of Computing Applications, vol. 69, No. 13, May 2013.

[7].Er. Sahibinder Singh, Er. Devesh Mahor, "Hybrid Evolutionary Algorithm Based on PSO to Reduce Non linear effect for 802.11a High Speed Network," International Journal of Scientific & Engineering Research Volume 3, Issue 4, April-2012.

[8].Marco Lixia, Maurizio Murroni, Vlad Popescu, "PAPR Reduction in Multicarrier Modulations using Genetic Algorithm," International Conference on Optimization of Electrical and Electronic Equipment, OPTIM 2010.

[9]. Neemi Taspinar, Dervis Karaboga, Mahmut Yildirim, Bahriye Akay, "PAPR Reduction using Artificial Bee Colony Algorithm in OFDM Systems," Turk J Elec. Eng. & Comp Sci., vol. 19, No. 1, 2011.

[10]. R.Gayathri, V.Sangeetha, S Prabha, D.Meenakshi, N.R. Raajan, "PAPR Reduction in OFDM Using Partial Transmit Sequence (PTS)," International Journal of Engineering and Technology, vol. 5, No. 2, Apr-May 2013.

[11]. R.Manjith, M.Suganthi, "A Sub-optimal PTS Algorithm Based on Bacterial Foraging Optimization Technique for PAPR Reduction in MIMO-OFDM System," Journal of Theoretical and Applied Information

Technology, vol. 57, No. 2, Nov. 2013.

[12]. Radheshyam Patel, Chirag Prajapati, Vishal Naik, Prof. T.P.Dave, “PAPR Reduction Technique in OFDM System,” International Journal of Computer Science Trends and Technology, vol. 2, Issue 1, Jan- Feb 2014.

[13]. Vikas Preet Tiwana, Sonia Singla, “Study of Various PTS Based Algorithms for PAPR Reduction in OFDM Systems,” American International Journal of Research in Science, Technology, Engineering & Mathematics, March-May, 2013.

[14]. V.Vijayarangan, Dr.(Mrs.)R.Sukanesh, “An Overview of Techniques for Reducing Peak to Average Power Ratio and its Selection Criteria for Orthogonal Frequency Division Multiplexing Radio Systems,” Journal of Theoretical and Applied Information Technology, vol. 5, No. 5, 2005-2009.

Track-1

Technical Session 2

OPTICAL COMMUNICATION

Investigation of Dispersion Penalty for 10 Gb/s Link with Different Values of Dispersion

Gurpreet Kaur
Department of Electronics & Communication Engg.,
Punjabi University,
Patiala, India.
dhammu.gurpreet836@gmail.com

Gurmeet Kaur
Department of Electronics & Communication Engg.,
Punjabi University,
Patiala, India.
farishta02@yahoo.co.in

***Abstract*— Fiber dispersion degrades the performance of optical communication systems by broadening optical pulses as they propagate inside the fiber. This paper demonstrates the effect of dispersion on dispersion penalty at different lengths of fiber for pseudorandom data at the OC-192 rate. This paper shows that the large chirp from directly modulated lasers operating at the OC-192 rate (9.953 Gb/s) has prevented their use in such high data rate systems. Simulation results show power required to achieve the constant bit error rate with the increase in length.**

Keywords—Directly modulated laser (DML), Bit Error Rate (BER), Single Mode Fiber (SMF).

I. INTRODUCTION

Optical transmission over optical fibers was first proposed in 1966. Since 1980 such systems have been deployed worldwide and indeed revolutionize the technology behind telecommunication [1]. A new method to extract the line-width enhancement factor has been purposed, which introduce a new way to extract rate equation parameters for laser lasing at the wave length for zero dispersion in optical fiber (1310nm) [2]. Directly modulated distributed feedback lasers have attracted much attention recently for application in metropolitan areas systems operating at the OC-48 rate (2.488 GB/s) and below [3]. Because they provide suitable optical output at the lowest cost, footprint and power dissipation [4]. However the frequency chirp characteristics of directly modulated lasers significantly limit the maximum achievable transmission distance over standard single mode fiber [5]. Dispersion is one of the major factors that affect the performance of optical link. As a pulse propagates along the optical fiber it spreads in time resulting in a decrease in peak power and total energy in its own bit period. Spreading of power into adjacent bit periods resulting high bit error rate (BER) and subsequently decrease the transmission distance. In this paper the effect of dispersion on dispersion penalty for different values of fiber length has been demonstrated. The organization of this paper is as follows: introduction in section I, theory of chromatic dispersion in section II, experimental setup in section III, results and discussions in section IV, and a short conclusion has been presented in section V.

II. CHROMATIC DISPERSION

In optical fiber communications systems, pulses of light representing digital ones are launched into the fiber. Each bit (a digital one or zero) is allocated to its own time slot or bit period (T), which is equal to the reciprocal of the bit rate (BR). Dispersion in optical fibers is caused by the refractive index of glass being a function of wavelength, which results in the spectral components of a pulse traveling at different group velocities along the fiber. Hence, chromatic dispersion broadens optical pulses beyond their time slot, leading to inter symbol interference (ISI) as shown in Fig.1. It is found that there are linear effects of chromatic dispersion and attenuation increase with increasing the distance along the fiber optic length[6] and eventually this will lead to errors at the receiver output with ones being recorded in bit periods into which zeroes were transmitted and zeroes into which ones were transmitted. The type of pulses that used in the simulation is Gaussian pulse [7]. The pulse envelope in time t at the spatial position z, propagating from transmitting to the receiving end of an optical fiber communication system is A (z, t). This pulse is described by the non-linear Schrodinger equation given in "equation 1" [8]:

$$\frac{\partial A}{\partial Z} + \beta_1 \frac{\partial A}{\partial t} + \frac{i}{2}\frac{\beta_2 \partial^2 A}{\partial t^2} + \frac{\alpha}{2} A = i\Delta\beta A \qquad (1)$$

Where A is the pulse envelope, β_1 is the first order dispersion, [9] β_2 is the second order dispersion parameter causes pulse broadening due to chromatic dispersion. α is the attenuation coefficient of the fiber. The term with includes the effect of fiber loss and nonlinearity

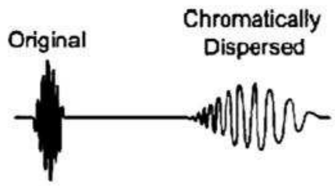


Fig.1. Optical Pulse Broadening caused by chromatic dispersion [10]

There are two distinct sources of pulse spreading in single mode fiber namely; material dispersion and waveguide dispersion and they are commonly referred to Chromatic dispersion [10]. Chromatic dispersion causes pulse broadening and it limits the transmission distance, bit rate and number of channel on an optical communication [11]. Dispersion becomes more problematic when the optical pulses in transport fiber begin to overlap. Pulse interference depends on dispersion value, and data bit rates, optical source spectral width and fiber length. The dispersion parameter of fiber is given by "equation (2)" [12];

$$D= d\,(1/Vg)/d \qquad (2)$$

Where Vg is group velocity and D is called the dispersion parameter. The typical value of dispersion parameter for single mode fiber (SMF) is 16ps/nm.km at 1550nm. This paper discusses the effect of different values of dispersion parameter on power penalty. The group velocity dispersion limits the bit rate distance product BL by "equation (3)" [1];

$$BL \quad (4|D| \quad)^{-1} \qquad (3)$$

Where is the root mean square (RMS) source spectral width in wavelength units, B is the bit rate, L is the fiber length. Dispersion induced pulse broadening can also decrease the receiver sensitivity. Here power penalty $_d$ is defined as the required increase (in dB) in the received power that would compensate the peak–power reduction, and $_d$ is given by "equation (4)" [1];

$$_d=10\log_{10}(f_b) \qquad (4)$$

Where f_b is the pulse broadening factor and this is the ratio of RMS width of broadened optical pulse and RMS width of the optical pulse at the fiber input. Another relation also exists for dispersion penalty to examine the performance of optical link. An approximate relation to calculate the dispersion penalty as a function of fiber length L is given by "equation (5)" [13];

$$_d=5*(\log(1+2\pi(B\,D \quad)^2 L^2)) \qquad (5)$$

The tolerable range of this penalty is 0.5 dB-1.0 dB. The system sensitivity degrades rapidly with the dispersion. It is possible to design a system such that the pulse spreads outside the bit slot but ISI is reduced through pulse shaping at the receiver.

III. OPTICAL FIBER RECEIVER

The transmitter comprised a directly modulated laser (DML) operating at 193.6 THz (1548.51 nm) and an extinction ratio of 6.6 dB. The modulation format is non return to-zero, the data load is a pseudo random bit sequence (PRBS) of 2^{23} -1 length.

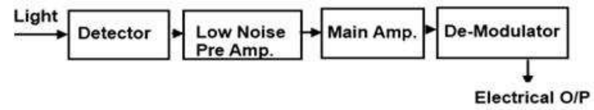


Fig.2. Receiver setup with two stage amplification [14]

The data rate is 9.953 Gb/s and 100nA value of thermal noise is fixed. As shown in Fig.2 the first stage of optical fiber receiver is detector, which converts the received signal into an electrical form. In the amplification stages there are two stages of amplifier are used i.e.; pre amplifier and main amplifier. These amplification stages amplify the converted signal, for further processing. Demodulator or decision circuit reproduces the original electrical signal from modulated incoming signals. All spans are standard dispersion fiber with different values of dispersion such as 11ps/nm.km to 19ps/nm.km, and no optical dispersion compensation is used.

IV. RESULTS AND DISCUSSIONS

Results of the work done have been summarized in Fig.3 to Fig.7.

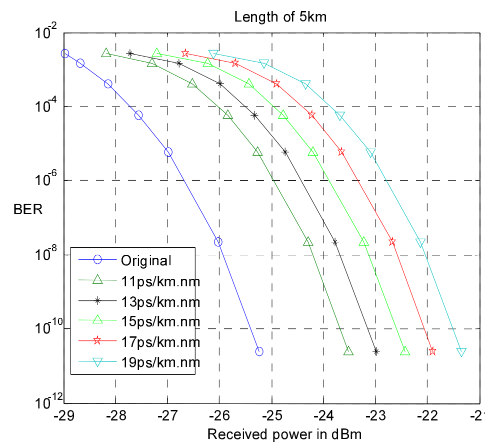


Fig.3. BERs for different values of dispersion at fiber length of 5km

Fig.3 shows the bit error rates (BERs) obtained with uncompensated receiver at different values of fiber dispersion at length 5km. The maximum value of dispersion penalty is 3.8887 dB at 19ps/nm.km. At BER of 2.78×10^{-3}, the percentage increase in received power is 2.80, 4.51, 6.55, 8.71, and 10.95 for dispersion value of 11ps/nm.km, 13ps/nm.km, 15ps/nm.km, 17ps/nm.km and 19ps/nm.km respectively. And at BER of 2.51×10^{-11}, the percentage increase in received power is 7.27, 9.75, 12.43, 15.21, and 18.18 for dispersion value of 11ps/nm.km, 13ps/nm.km, 15ps/nm.km, 17ps/nm.km and 19ps/nm.km respectively.

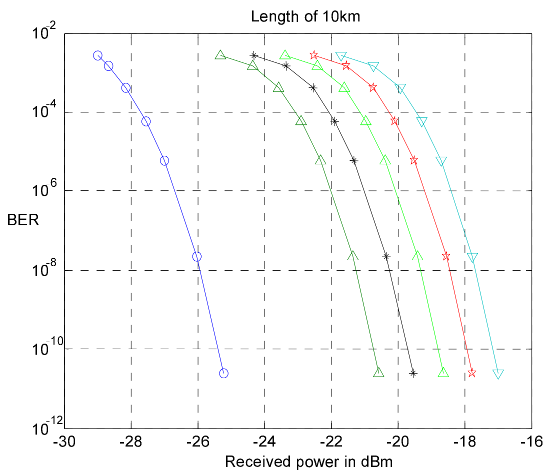


Fig.4. BERs for different values of dispersion at fiber length of 10km

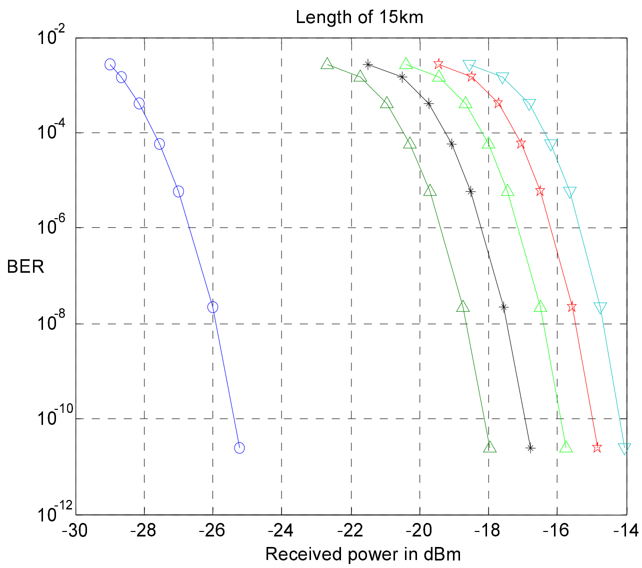


Fig.5. BERs for different values of dispersion at fiber length of 15km

Fig.4 shows the bit error rates (BERs) obtained with different values of fiber dispersion at fiber length of 10km. The maximum value of dispersion penalty is 8.3208dB at 19ps/nm.km. At BER of 2.78x10^{-3}, the percentage increase in received power is 14.46, 19.22, 24.02, 28.81, and 33.62 for dispersion value of 11ps/nm.km, 13ps/nm.km, 15ps/nm.km, 17ps/nm.km and 19ps/nm.km respectively. And at BER of 2.51x10^{-11}, the percentage increase in received power is 22.72, 29, 35.44, 42, and 48.62 for dispersion value of 11ps/nm.km, 13ps/nm.km, 15ps/nm.km, 17ps/nm.km and 19ps/nm.km respectively.

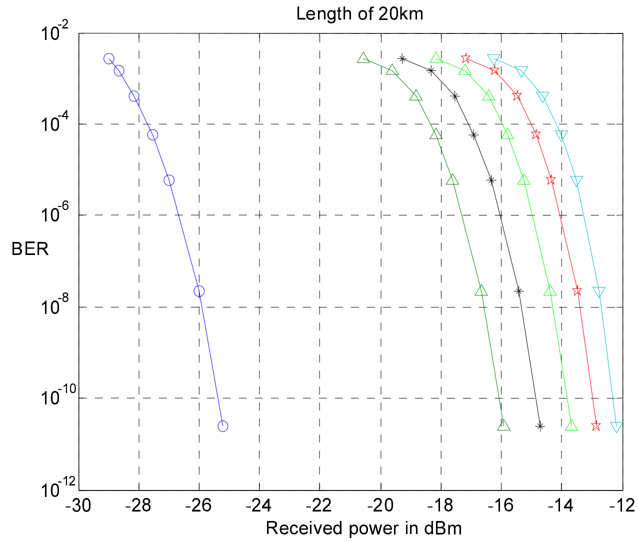


Fig.6 BERs for different values of dispersion at fiber length of 20km

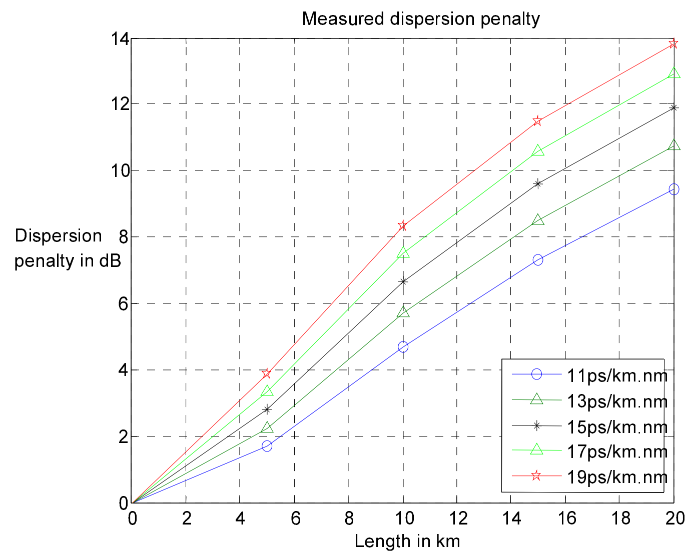


Fig.7. Measured dispersion penalty, at BER= 10^{-9}, with different values of fiber dispersion.

Fig.5 shows the bit error rates (BERs) at different values of fiber dispersion at length 15km; here the maximum dispersion penalty is 11.4721dB. At BER of 2.78x10^{-3}, the percentage increase in received power is 27.62, 34.81, 41.94, 49.1, and 56.1 for dispersion value of 11ps/nm.km, 13ps/nm.km, 15ps/nm.km, 17ps/nm.km and 19ps/nm.km respectively. And at BER of 2.51x10^{-11}, the percentage increase in received power is 40.34, 50.21, 60.12, 69.94, and 79.62 for dispersion value of 11ps/nm.km, 13ps/nm.km, 15ps/nm.km, 17ps/nm.km and 19ps/nm.km respectively.

Fig. 6 shows the bit error rates (BERs) versus received power at length of 15km, here the maximum dispersion penalty are 13.8334 dB. At BER of 2.78x10^{-3}, the percentage increase in received power is 40.76, 50.26, 59.61, 68.92, and 78.16 for dispersion value of 11ps/nm.km, 13ps/nm.km, 15ps/nm.km, 17ps/nm.km and 19ps/nm.km respectively. And at BER of 2.51x10^{-11}, the percentage increase in received power is 58.51, 71.56, 84.22, 96.11, and 106.72 for dispersion value of 11ps/nm.km, 13ps/nm.km, 15ps/nm.km, 17ps/nm.km and 19ps/nm.km respectively.

Fig.7 plots the dispersion penalty for different values of fiber dispersion as a function of length. With the percentage increase in dispersion of 72, the dispersion penalty has increased by 126.23%, 77.57%, 57.14%, and 46.67% at fiber length of 5km, 10km, 15km and 20km respectively. And with 300% increase in fiber length, the dispersion penalty has increased by 448.5%, 377.2%, 325.5%, 286.18% and 255.62% for dispersion value of 11ps/nm.km, 13ps/nm.km, 15ps/nm.km,17ps/nm.km and 19ps/nm.km respectively.

V. CONCLUSION

The design of an optical system involves many interrelated variables among the fiber, source and photo detector operating characteristics. In carrying out an optical fiber link analysis, several iterations with different device characteristics may be required before it is satisfactorily completed. Dispersion is one of the major factors that affect the performance of an optical link. It is clear from the comparison between the results at deferent distances, that the effects of chromatic dispersion increases linearly with increasing the distance along the fiber optic communication systems. In this paper, affect of dispersion on dispersion penalty with the change in fiber length has been studied. It has been concluded from this study that with the increase in fiber length dispersion penalty has also increased. With the percentage increase in dispersion of 72.72, the dispersion penalty has increased by 46.67% at fiber length of 20km. And with 300% increase in fiber length, the dispersion penalty has increased by 286.18% for typical value of dispersion 17ps/nm.km.

REFERENCES

[1] Govind, P. A., "Fiber-Optics Communication Systems", John Wiley and Sons, Inc, Third Edition, pp 16, 2002.

[2] Cartledge, J. C., et al., "Extraction of DFB laser rate equation parameters for system simulation purposes," Journal of light wave technology, vol. 15, pp. 852–860, May 1997.

[3] Antoniades N., et al., "Measurements of laser rate equation parameters for simulating the performance of directly modulated 2.5 Gb/s Metro area transmission systems and networks," Proc. LEOS, vol. 2, pp. 692–693,2000.

[4] Papagiannakis I., et al., " Transmission Performance Improvement Studies for Low-Cost 2.5 Gb/s Rated DML Sources Operated at 10 Gb/s," ECOC 2008, Brussels, Belgium, 21-25 September 2008.

[5] Hinton, K., et al., "Modeling high-speed optical transmission systems", IEEE Journal of Selected Areas Communication, vol. 11, pp. 380–392, 1993.

[6] Sabah H. S., "International Journal of Modern Communication Technologies & Research, Vol. 2, Issue-3, March 2014.

[7] Pearce, M. B. et al.," Optimal input Gaussian pulse width for transmission in dispersive nonlinear fibers", JOSA B, Optical Society of America, Vol. 16, Issue 8, pp. 1189-1196, 1999.

[8] Saeid, S. H., "Performance Evaluation of Single Mode Fiber Optics for Long Distance Optical Communication", International Journal of Computer Applications, Harvard, June, 2013.

[9] Saeid, S. H., "Simulation of Single Mode Fiber Optics and Optical Communication Components Using VC++", International Journal of Computer Science and Network Security, Vol. 8, No. 2, pp. 300-308, February 2008.

[10] Danziger, Y., et al., "High-order Mode Fiber: An Innovative approach to chromatic dispersion management that enables optical networking in long-haul high-speed transmission systems", Optical Networks Magazine, vol. 2, Issue 1, 2001.

[11] Ball, P., "Optical and broadband Networks" Lecture Note, Oxford Brookes University, Oxford, UK, 2009.

[12] Papannareddy, R., "Introduction to Light wave Communication Systems", Artec House, Norwood, MA, 1997.

[13] Casimer, M. D., et al., "Fiber optics Essential", Journal of fiber optic technology, Elsevier, ISBN: 13: 978-0 12-208431-7.

[14] http://www.daenotes.com/electronics/communication-system/optical-fiber.

All-Optical Logic Design Techniques Based On Sagnac Interferometers

Karamdeep Singh
Dept. of ECE
Punjabi University
Patiala, India
karamdeep1989@gmail.com

Gurmeet Kaur
Dept. of ECE
Punjabi University
Patiala, India

Abstract—**Development of All-Optical Logic functions for All-Optical Signal Processing, have attracted considerable interest in recent years. Efforts have been made to realize All-Optical Logic devices utilizing several non-linear effects and Interferometric configurations such as Mach-Zehnder Interferometer (MZI), Ultrafast Non-Linear Interferometer (UNI) & Sagnac Interferometers. Out of all Interferometric configurations, All-Optical logic devices based on Sagnac Interferometer are attractive, due to their ability to operate at high bit rates & low latency. This paper is an overview of Sagnac Interferometer fundamentals & its variants. All-Optical Binary as well as Multi-Valued (MV) Logic devices based on Sagnac Interferometers are also illustrated.**

Keywords—All-Optical Signal Processin, All-Optical Logic Design; Sagnac Interferometers, Nonlinear Optical Loop Mirror (NOLM), Terahertz Optical Asymmetric Demultiplexer (TOAD).

I. INTRODUCTION

Over the past few years, interest in design of ultrafast non-linear, All-Optical switching devices is increasing. In order to cope with growing demand of computational speed due to excessive use of internet and multimedia technologies, All-Optical signal processing devices have been developed over the course of time [1], [2]. All-Optical Signal Processing devices avoid costly Optical-Electronic-Optical (O/E/O) conversions usually required in traditional electronic signal processing systems, by implementing all operations in optical domain itself [3], [4], [5]. All-Optical Logic functions and modules serve as basic building blocks of complex All-Optical Signal Processing circuits for carrying out specialized operations such as TTL Decrementing, Routing etc. Design and analysis of All-Optical Logic functions and modules for All-Optical Signal Processing is witnessing widespread research activity at present time. Approximately, in past two decades, various researchers have presented a number of approaches to realize All-Optical devices, such as exploiting 3^{rd} order ($^{(3)}$) non-linear processes in Semiconductor Optical Amplifiers (SOA) [6], [7] & Highly Nonlinear Fiber (HNLF) [8] , utilizing 2^{nd} order ($^{(2)}$) non-linear processes in Periodically Poled Lithium Niobate (PPLN) waveguides [9]. Further, Interferometric structures such as Semiconductor Optical Amplifier Mach-Zehnder Interferometer (SOA-MZI) [10], [11], Ultra fast Nonlinear Interferometer (UNI) [12] and Sagnac Interferometers [1], [2], [3], [4], [5] have been used in number of instances for implementing All-Optical logic functions and modules. Among all approaches used for design of All-Optical Logic devices, Sagnac Interferometer based devices are attractive as compared to its counterparts due to their promising features such as fast switching time, low latency, high repetition rate and low power consumption [1], [2], [3], [4], [5].

In this paper, advances in the field of design and analysis of All-Optical Logic devices based on Sagnac Interferometers have been reviewed. In section I, an introduction was provided to importance of All-Optical Signal Processing along with various approaches used to realize All-Optical logic devices. Section II, illustrates the basics of Sagnac Interferometers & its various topological variations used for All-Optical Logic implementation. Literature survey concerned with Sagnac Interferometer based schemes to implement All-Optical logic functions, is presented in Section III. Section IV, concludes this paper.

II. SAGNAC INTERFEROMETERS: FUNDAMENTALS & TYPES

Interferometers are the devices which determine certain properties by employing concept of superimposed waves. Two waves are combined in a special way inside an Interferometric device, so that some meaningful property is resulted. Sagnac Interferometers were introduced by Georges Sagnac in 1913. Sagnac Interferometers consist of a ring cavity made from fibers in which, two optical beams are launched to propagate in opposite directions. One beam traverses the loop distance in Clock Wise (CW) directions, whereas other one in Counter-Clock Wise (CCW) fashion. The two beams after travelling equal distance interfere at beam splitter with same phase. However, if one of the two paths in the ring cavity is made shorter as compared to other such as, CW path is made shorter than CCW path using some arrangement, then phase of signal traversing shorter distance will be different from the other signal travelling longer distance. Thus, interference at beam splitter will be different. This configuration is called Sagnac Interferometer [1], [2], [13], [14], [15], [16], [17].

Several variations of Sagnac interferometer have been suggested over the years such as: Non-Linear Optical Loop Mirror (NOLM), Terahertz Optical Asymmetric Demultiplexer (TOAD) and Semiconductor Laser Amplifier in a Loop Mirror (SLALOM) which are illustrated as under:

A. Non-Linear Optical Loop Mirror (NOLM)

Non-Linear Optical Loop Mirrors (NOLM) utilize Kerr-non-linearity effects such as Self-Phase Modulation (SPM) & Cross-Phase Modulation (XPM) in optical fibers to induce phase shift on counter propagating signals in Sagnac loop to achieve desired processing aim. NOLM is a configuration consisting of an asymmetric coupler with a Non-Linear Element (NLE) such as Dispersion Shifted Fiber (DSF) or Highly Nonlinear Fiber (HNLF) arranged in a ring structure to perform phase shift unbalancing as shown below [13], [15], [17]:

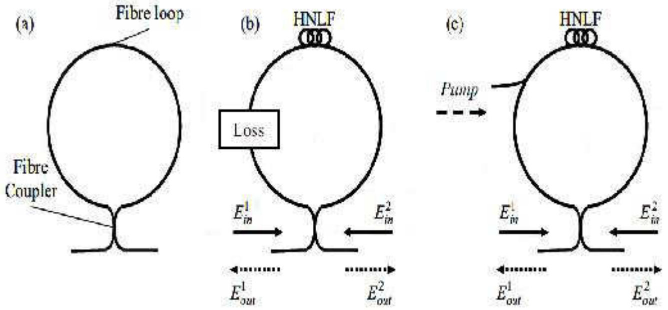


Fig. 1. Non-Linear Optical Loop Mirror (NOLM) based on Sagnac Effect (a) Basic Strcuture, (b) Exploiting Self-Phase Modulation (SPM), (c) Utilizing Cross-Phase Modulation (XPM) [13].

The three variants of NOLM are shown in Fig. 1. (a-c). As shown in Fig. 1 (b) & (c), Self-Phase Modulation (SPM) and Cross-Phase Modulation (XPM) phenomena's inside DSF or HNLF, respectively have been used to generate phase shift to induce path difference between two beams propagating in opposite direction inside NOLM structure. The only difference between SPM & XPM based NOLM variants is use of additional pump signal in the latter case, which helps to reduce dependence on optical power of probe signal to induce path difference & thus improving dynamic performance of the realized All-Optical devices [13], [15], [17], [18], [19].

B. Terahertz Optical Asymmetric Demultiplexer (TOAD)

Terahertz Optical Asymmetric Demultiplexer (TOAD) exploits slow non-linearities present in Semiconductor Optical Amplifier (SOA), which is used as NLE in this Sagnac Interferometer variant. The schematic diagram of TOAD configuration is given as under [20], [21], [22]:

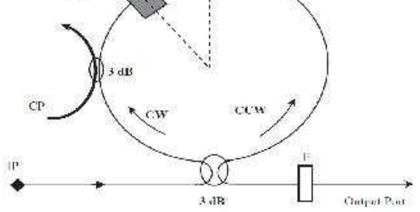


Fig. 2. Terahertz Optical Asymmetrical Demultiplexer (TOAD) arrangement [22].

As shown in Fig. 2, TOAD switch consist of a 2×2 (50:50) coupler, which is used to split signal into two parts and a SOA placed asymmetrically at a offset of x from the midpoint of the Sagnac loop. The loop also contains another 3 dB coupler, through which Control Pulse (CP) is injected in SOA along with Clock-Wise (CW) Input Pulse (IP). A path difference is induced by XPM effect in SOA between Clock-Wise (CW) & Counter-Clock-Wise (CCW) Input Pulses (IP). Both input pulses after traversing whole Sagnac loop recombine at input coupler. Depending on the phase difference between CW and CCW signals, data will be either transmitted to output port or reflected back to input port [3], [4], [5], [20], [21], [22].

C. Semiconductor Laser Amplifier in a Loop Mirror (SLALOM)

Semiconductor Laser Amplifier in Loop Mirror (SLALOM) configuration is similar to TOAD Sagnac arrangement. The only difference lies in the use of Polarization Controllers (PC) inside the Sagnac loop in SLALOM configuration as illustrated in Fig. 3 below [1], [2], [15]:

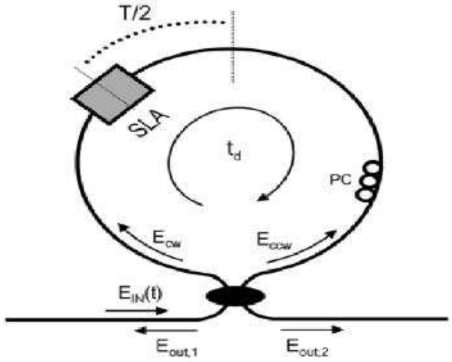


Fig. 3. Semiconductor Laser Amplifier in Loop Mirror (SLALOM) Configuration [15].

As depicted in Fig. 3, SLALOM configuration consists of an input coupler, Polarization Controller (PC) along with asymmetrically placed NLE, which is Semiconductor Laser Amplifier (SLA) [1], [2], [15].

III. ALL-OPTICAL LOGIC DEVICES BASED ON SAGNAC INTERFEROMETERS

A number of All-Optical logic devices have been implemented based on Sagnac Interferometer techniques discussed earlier such as NOLM, TOAD and SLALOM over the years. They are presented in the form of year wise literature survey as under:

T. Houbavlis, et al., [1] (1999) proposed a SOA assisted Sagnac Interferometer (SLALOM configuration) All-Optical XOR gate with 5 GHz repetition rate, exhibiting Contrast Ratios (CR) up to 13 dB. The operation of the proposed XOR gate was based on the differential phase change between two counter propagating clock pulses. It has been said that proposed scheme was input data pattern independent.

T. Houbavlis, et al., [2] (1999) demonstrated a Boolean All-Optical XOR Gate operating at 10 Gbps using a Semiconductor Optical Amplifier (SOA) based three-terminal Sagnac Interferometer (SLALOM configuration). A 1000 μm long SOA, having 100 ps gain recovery time was used in the Sagnac Interferometer Loop. Contrast Ratios (CR) as high as up to 14:1 and 12:1 have been achieved using the proposed architecture. It has been said that the above said technique is input data pattern independent.

A. J. Poustie, et al., [16] (1996) used the Semiconductor Optical Amplifier (SOA) in a Fiber Loop Mirror (Terahertz Optical Asymmetrical Demultiplexer (TOAD) configuration) for implementing an All-Optical circulating shift register. A Distributed Feedback (DFB) Laser operating at 1.55 μm centre wavelength, generating 20 ps wide pulses at 1.5 repetition rate was used as source. The reflected pulses from loop mirror were amplified and rotated in polarization domain before feeding back into loop mirror as switching pulses to realize circular shift register operation.

K. L. Hall and K. A. Rauschenbach, [17] (1996) demonstrated an All-optical bit pattern generation and pattern matching scheme. The proposed scheme utilized Non-Linear Optical Loop Mirror (NOLM) based XOR gate. The noise generated form fibre loop memory has been used as bit patterns and XOR gate was used for matching data patterns. NOLM was used for implementing the XOR gate. It has been said that the proposed scheme can work at 10 Gbps data rate.

A. Bogoni, et al., [18] (2005) studied the use of Non-Linear Loop Mirror (NOLM) for the implementation of All-Optical regenerative and reconfigurable logic gates. All-Optical AND/OR/XOR and NOR/XNOR functions have been realized utilizing Self-Phase Modulation (SPM) and Cross-Phase Modulation (XPM) process in Dispersion Shifted Fiber (DSF) placed in NOLM for 10 GHz NRZ and 40 GHz RZ signals. It has been shown that Four-Wave Mixing (FWM) process was suppressed by choosing both signals centred at same wavelength. Contrast Ratios (CR) of 20 dB and 6.5 dB have been recorded for AND/OR/XOR/NOR functions and XNOR function, respectively.

Y. Miyoshi, et al., [19] (2008) proposed a Highly Non-Linear Fiber (HNLF) based Non-Linear Optical Mirror (NOLM) approach to generate multiple logic gates (AND, NAND, OR, NOR, XOR and XNOR) using single device structure at 40 Gbps. It has been shown that by simply adjusting the Polarization Controllers (PC) and Variable Optical Attenuators (VOA), different logic functions can be realized. Investigations have been carried out in order to determine the effect of dispersion and Self-Phase Modulation (SPM) on processing speed of the proposed device.

T. Chattopadhyay, et al., [23] (2009) proposed an All-Optical R-S Flip-Flop for Multi-Valued Logic (MVL) signals utilizing polarization encoding in Semiconductor Optical Amplifier (SOA) based Terahertz Optical Asymmetrical Demultiplexer (TOAD) switch. It has been illustrated that proposed R-S Flip-Flop consisted of structure containing TOAD based binary latch and binary to quaternary Encoder/Decoder. Binary latch has been used as All-Optical memory element.

D. K. Gayen, et al., [24] (2010) described an All-Optical Incrementer/Decrementer based on Semiconductor Optical Amplifier Terahertz Optical Asymmetrical Demultiplexer (SOA-TOAD) switches. Half-Subtracter and Half-Adder have been used as building blocks to realize incrementer/decrementer. Numerical simulations have been performed to investigate the Contrast ratio (CR) variations by using different Gain recovery time SOAs and by varying control pulse. CR of 38.06 dB and Bit-Error Rate (BER) of 10^{-7} have been reported.

T. Chattopadhyay, [22] (2010) utilized Semiconductor Optical Amplifier (SOA) based Terahertz Optical Asymmetrical Demultiplexer (TOAD) to realize tri-state logic gates. It has been said that different polarization states of light were used to represent ternary logic states. In order to verify operation of proposed approach numerical simulations have been done.

T. Chattopadhyay, et al., [20] (2011) theoretically analyzed an All-Optical 3-bit binary to Quaternary Signed Digit (QSD) converter using optical tree architecture based on Semiconductor Optical Amplifier (SOA) Terahertz Optical Asymmetrical Demultiplexer (TOAD) configuration. It has been illustrated that different polarization states have been used to represent different numbers. Numerical simulations have been done to verify the operation of the proposed converter.

T. Chattopadhyay and J. N. Roy, [21] (2011) exploited polarization properties of light to demonstrate a Semiconductor Optical Amplifier (SOA) assisted Sagnac Gate based All-Optical Quaternary Successor (QSUC) circuit. Extinction Ratio (ER) of 11.76 dB with 0.895 dB Amplitude Modulation (AM) has been shown using numerical simulations. Modulo-arithmetic and universal inverter implementations have also been discussed.

C. Taraphdar, et al., [25] (2011) developed an All-Optical ternary logic module based on Terahertz Optical Asymmetrical Demultiplexer (TOAD) using Semiconductor Optical Amplifiers (SOA). It has been said that proposed single logic module can be used to generate 27 logic functions alone. It has been illustrated that polarization state of light is used to represent a particular logical state, in this scheme, 3 polarization states were used to represent 3 logic states (ternary logic).

D. K. Gayen, et al., [3] (2011) numerically examined a Terahertz Optical Asymmetrical Demultiplexer (TOAD) based All-Optical reconfigurable logic module. It has been shown that proposed logic module is able to be reconfigured to perform XOR, NOR & AND gates. Numerical simulations have been conducted to verify the operation of the above said approach. Contrast Ratio (CR) and Extinction Ratio (ER) of the order of 35.34 dB and 26.03 dB, respectively have been observed at 10^{-9} Bit-Error Rate (BER) using proposed method.

T. Chattopadhyay and J. N. Roy, [4] (2011) designed tri-state logic gates (OR, AND & NOT) based on Semiconductor Optical Amplifier (SOA) arranged in Terahertz Optical Asymmetrical Demultiplexer (TOAD) configuration. It has been illustrated that three polarization states have been used to represent three logic states. Numerical simulations have been performed to determine Contrast Ratio (CR), Extinction Ratio (ER) and Q-Factor of the above said approach based logic gates.

T. Chattopadhyay, [5] (2011) described All-Optical XOR gate based on Semiconductor Optical Amplifier placed in Terahertz Optical Asymmetrical Demultiplexer (TOAD) arrangement with no additional input assisting beams. Numerical simulations have been conducted to analyze the performance of proposed device in terms of Contrast Ratio (CR), Extinction Ratio (ER), Amplitude Modulation (AM) and Signal to Noise Ratio (SNR) parameters at 11.11 Gbps bit rate.

D. K. Gayen, et al., [26] (2012) theoretically analyzed Semiconductor Optical Amplifier (SOA) based Sagnac structure (TOAD configuration) to realize All-Optical Carry Lookahead Adder. Analysis has been carried out by changing switching pulse energy, linewidth enhancement factor, SOA small signal gain. Contrast Ratios (CR) > 38.30 dB have been reported using proposed approach.

The study in the form of literature review illustrated a handful of binary as well as multi-valued All-Optical logic devices based on Sagnac Interferometer configurations. The Non-Linear Optical Loop Mirror (NOLM) was the first Sagnac configuration utilized to implement All-Optical logic functions such as XOR gate at 10 Gbps [17] and reconfigurable logic gates at 10 GHz & 40 GHz [18] & at 40 Gbps [19]. It has been observed that NOLM configuration based All-Optical logic devices operate at high bit rates, as switching is based on passive elements such as Highly Non-Linear Fibers (HNLF) and Dispersion Shifted Fibers (DSF). However, in order to efficiently utilize non-linear effects in HNLF and DSF, higher control powers and longer fibre lengths are required, which in turn make integrating the NOLM based systems on single chip very difficult. The second approach discussed was Semiconductor Laser Amplifier in Loop Mirror (SLALOM), which can solve this integration problem, as it requires Semiconductor Laser Amplifier (SLA) as active element, which can be easily integrated on a single chip. SLALOM Sagnac arrangement has been used to realize All-Optical XOR Gate at 5 GHz [1] & at 10 Gbps [2]. However, the operating bit rate of SLALOM based devices is lesser as compared to NOLM based devices. In order to overcome this lower operating bit rate problem, Terahertz Optical Asymmetric Demultiplexer (TOAD), which can provide high bit rate operation, has been widely deployed. Various All-Optical logic devices such as Circulating Shift Register [16], Incrementer/Decrementer [24], Tri-State Logic Gates [22], [4], 3-Bit Binary to QSD Converter [20], Quaternary Successor (QSUC) Circuit [21], Ternary Logic Module [25], Reconfigurable Logic Module [3], XOR Gate [5] & Carry Look Ahead Adder [26] have been obtained using TOAD Sagnac Arrangement.

It has been inferred from the study of literature, concerning with Sagnac Interferometer configurations that most of the All-Optical Logic Devices realized were based on TOAD arrangement. Enhancement of some of these All-Optical binary and multi-valued logic devices using other two configurations i.e. SLALOM and NOLM can be subject of future research.

IV. CONCLUSION

Sagnac Interferometer based All-Optical Logic devices are faster, consume lesser power and have high Signal to Noise Ratio (SNR) as compared to other Interferometric based devices such as MZI & UNI. In this paper, a basic introduction to Sagnac Interferometers was given along with its variations such as Non-Linear Optical Loop Mirrors (NOLM), Terahertz Optical Asymmetric Demultiplexer (TOAD) & Semiconductor Laser Amplifier in Loop Mirror (SLALOM). All-Optical logic devices developed using above said Sagnac Interferometer variants were discussed in the form of literature review. It has been summarized that there is further scope of research in improvement of All-Optical Logic devices using NOLM and SLALOM Sagnac configurations which have been already implemented using TOAD configuration.

REFERENCES

[1] T. Houbavlis, K. Zoiros, K. Vlachos, T. Papakyriakopoulos, H. Avramopoulos, F. Girardin, G. Guekos, R. Dall'Ara, S. Hansmann and H. Burkhard, "All-Optical XOR in a Semiconductor Optical Amplifier-Assisted Fiber Sagnac Gate," IEEE Photonics Technology Letters, Vol. 11, No. 3, March 1999, pp. 334-336.

[2] T. Houbavlis, K. Zoiros, A. Hatziefremidis, H. Avramopoulos, L. Occhi, G. Guekos, S. Hansmann, H. Burkhard and R. Dall'Ara, "10 Gbit/s All-Optical Boolean XOR with SOA fibre Sagnac Gate," Electronics Letters, Vol. 35, No. 19, September 1999, pp. 1650-1652.

[3] D. K. Gayen, J. N. Roy, C. Taraphdar and R. K. Pal, "All-Optical Reconfigurable Logic Operations with the Help of Terahertz Optical Asymmetric Demultiplexer," Optik, Vol. 122, No. 8, April 2011, pp. 711-718.

[4] T. Chattopadhyay and J. N. Roy, "Semiconductor Optical Amplifier (SOA)-Assisted Sagnac Switch for Designing of All-Optical Tri-State Logic Gates," Optik, Vol. 122, No. 12, June 2011, pp. 1073-1078.

[5] T. Chattopadhyay, "Eliminating the Additional Input Beam in All-Optical XOR Gate using Terahertz Optical Asymmetric Demultiplexer (TOAD) based Interferometer: A Theoretical Analysis," Optik, Vol. 122, No. 16, August 2011, pp. 1486-1491.

[6] D. Nesset, M. C. Tatham and D. Cotter, "All-Optical AND Gate Operating on 10 Gbit/s Signals at the Same Wavelength using Four-Wave Mixing in a Semiconductor Laser Amplifier," Electronics Letters, Vol. 31, No. 11, May 1995, pp. 896-897.

[7] A. Sharaiha, H. W. Li, F. Marchese and J. L. Bihan, "All-Optical Logic NOR-Gate using a Semiconductor Laser Amplifier," Electronics Letters, Vol. 33, No. 4, February 1997, pp. 323-325.

[8] J. Wang, Q. Sun and J. Sun, "All-Optical 40 Gbit/s CSRZ-DPSK logic XOR Gate and Format Conversion using Four-Wave Mixing," Optics Express, Vol. 17, No. 15, July 2009, pp. 12555-12563.

[9] K. R. Parameswaran, M. Fujimura, M. H. Chou and M. M. Fejer, "Low-Power All-Optical Gate based on Sum Frequency Mixing in APE

Waveguides in PPLN," IEEE Photonics Technology Letters, Vol. 12, No. 6, June 2000, pp. 654-656.

[10] T. Fjelde, D. Wolfson, A. Kloch, C. Janz, A. Coquelin, I. Guillemot, F. Gaborit, F. Poingt, B. Dagens and M. Renaud, "10 Gbit/s All-Optical Logic OR in Monolithically Integrated Interferometric Wavelength Converter," Electronics Letters, Vol. 36, No. 9, April 2000, pp. 813-815.

[11] J. -Y. Kim, J. -M., Kang, T. -Y., Kim and S. -K., Han, "10 Gbit/s All-Optical Composite Logic gates with XOR, NOR, OR and NAND Functions using SOA-MZI Structures," Electronics Letters, Vol. 42, No. 5, March 2006, pp. 303-304.

[12] S. A. Hamilton and B. S. Robinson, "40-Gb/s All-Optical Packet Synchronization and Address Comparison for OTDM Networks," IEEE Photonics Technology Letters, Vol. 14, No. 2, February 2002, pp. 209-211.

[13] A. Bogoni, F. Fresi, P. Ghelfi, E. Lazzeri, L. Poti and M. Scaffardi, "Fiber based Schemes for Ultrafast Subsystems: Nonlinear Optical Loop Mirrors Traditional Design and Novel Applications," in Optical Fiber New Developments, C. Lethien, Rijeka, Croatia: InTech, pp. 515-538.

[14] R. Anderson, H. R. Bilger and G. E. Stedman, ""Sagnac" Effect: A Centuary of Earth Rotated Interferometers," American Journal of Physics, Vol. 62, No. 11, November 1994, pp. 975-985.

[15] J. M. M. Canet, "Photonic Logic Gates: Boosting All-Optical Header Processing in Future Packet Switched Networks," Ph. D. Dissertation, Valencia Nanophotonics Technology Centre, Spain, December 2006.

[16] A. J. Poustie, R. J. Manning and K. J. Blow, "All-Optical Circulating Shift Register using a Semiconductor Optical Amplifier in a Fiber Loop Mirror," Electronics Letters, Vol. 32, No. 13, June 1996, pp. 1215-1216.

[17] K. L. Hall and K. A. Rauschenbach, "All-Optical Bit Pattern Generation and Matching," Electronics Letters, Vol. 32, No. 13, June 1996, pp. 1214-1215.

[18] A. Bogoni, L. Poti, R. Proietti, G. Meloni, F. Ponzini and P. Ghelfi, "Regenerative and Reconfigurable All-Optical Logic Gates for Ultra-Fast Applications," Electronics Letters, Vol. 41, No. 7, March 2005, pp. 435-436.

[19] Y. Miyoshi, K. Ikeda, H. Tobioka, T. Inoue, S. Namiki and K. -I. Kitayama, "Ultrafast All-Optical Logic Gates using a Non-Linear Optical Loop Mirror based Multi-Periodic Transfer Function," Optics Express, Vol. 16, No. 4, February 2008, pp. 2570-2577.

[20] T. Chattopadhyay, M. K. Das, J. N. Roy, A. K. Chakraborty and D. K. Gayen, "Interferometric Switch based All Optical Scheme for Conversion of Binary Number to its Quaternary-Signed Digit Form," IET Circuits, Devices & Systems, Vol. 5, No. 2, March 2011, pp. 132-142.

[21] T. Chattopadhyay and J. N. Roy, "Polarization Encoded All-Optical Quaternary Successor with the Help of SOA Assisted Sagnac Switch," Optics Communication, Vol. 284, No. 12, June 2011, pp. 2755-2762.

[22] T. Chattopadhyay, "All-Optical Symmetric Ternary Logic Gates," Optics & Laser Technology, Vol. 42, No. 6, September 2010, pp. 1014-1021.

[23] T. Chattopadhyay, J. N. Roy and A. K. Chakraborty, "Polarization Encoded All-Optical Quaternary R-S Flip-Flop using Binary Latch," Optics Communication, Vol. 282, No. 7, April 2009, pp. 1287-1293.

[24] D. K. Gayen, J. N. Roy, R. K. Pal and S. P. Ghosh, "All-Optical Incrementer/Decrementer with the Help of Semiconductor Optical Amplifier-Assisted Sagnac Switch," International Journal of Optomechatronics, Vol. 4, No. 2, June 2010, pp. 177-194.

[25] C. Taraphdar, T. Chattopadhyay and J. N. Roy, "Designing of an All-Optical Scheme for Single Input Ternary Logical Operations," Optik, Vol. 122, No. 1, January 2011, pp. 33-36.

[26] D. K. Gayen, J. N. Roy and R. K. Pal, "All-Optical Carry Lookahead Adder with the Help of Terahertz Optical Asymmetric Demultiplexer," Optik, Vol. 123, No. 1, January 2012, pp. 40-45.

Analysis and Augmentation of IMD in Multitone Analog CATV Transmission Systems

Prabhpreet Kaur Bhatia
CT Institute of Technology & Research,
Maqsudan, Jalandhar
prabhpreet.bhatia91@gmail.com

Vineet Anand
CT Institute of Technology & Research,
Maqsudan, Jalandhar
anand.vineet56@gmail.com

Anurag Sharma
CT Institute of Technology & Research,
Maqsudan, Jalandhar
er.anurags@gmail.com

Manoj Kumar
CT Institutions
Jalandhar

Abstract–**In this paper, we analyze analog CATV transmission system where only one wavelength is used for transmission on which multiple analog RF channels are modulated. We analyze different modulation distortions such as composite second order (CSO), composite triple beat (CTB) and intermodulation distortions at different frequencies using different lasers. Good performance of CSO > −4.5dBc and CTB > −8.2 dBc are obtained at 62.5MHz, 125MHz and 187.5MHz in three tones CATV transmission system using FP etalon at receiver site.**

Keywords*–*CATV, EDFA, CSO, CTB

I. INTRODUCTION

During recent years, a Community Access Television (CATV) system has become widespread throughout the cable industry. CATV transmission systems have promoted from coaxial cable based one way broadcast analog video transmission to modern two ways hybrid fiber transmission. In traditional one way transmission several RF amplifiers were used to overcome high losses of the coaxial cables that affect the quality of received signal. The use of optical fiber reduces the RF amplifiers in the CATV transmission. To increase the system link budget, higher transmission power or lower fiber transmission loss are required. The maximum transmission distance of CATV system is limited by the RF parameters such as Composite second order (CSO), and Composite triple beat (CTB). These distortions parameters are degraded by using Stimulated Brillouin scattering (SBS) [1]. In general, CSO and CTB distortions induced by SBS are main limitations in the system [2]. When several numbers of channels in CATV system are amplified simultaneously then there is chance of having intermodulation products which produce distortion in the network. This is mainly due to non-linearity effects in the amplification in the CATV system. To improve the performance of CATV system these parameters should be degraded. Several methods have been proposed to overcome these problems induced by distortions in the CATV system. However, Sophisticated SBS suppression, CTB pre-distortion, differential detection techniques and expensive externally modulated transmitter are required [4]. It is difficult to obtain good performance of CSO and CTB. Hai han Lu.[3,4] proposed a directly modulated NTSC 80-channel Eribium Doped Fiber Amplifier (EDFA) repeated system using half-split-band (40 channels per transmitter) and wave division multiplexing (WDM) techniques to improve the performance of CSO and CTB in fiber optical CATV system. J.Yu [5] proposed an Electro absorption Modulated Laser (EML). EML can be linearized by emulation and reversal of the nonlinear distortion using a companion EML inside an optoelectronic interferometer. Suppression of composite second order and triple beat signals <−61 and <63dBc, respectively, have been achieved with a 77-channel loading of CATV carriers. Low distortion lasers are extremely important element in AM-VSB CATV transmission systems.

II. RELATED WORK

Hai-Han-Lu[6] compared the conventional externally modulated optical fiber CATV system with or without distortion compensation fiber (DCF).Excellent performance of composite second order (CSO) > −78 dBc and Composite triple beat (CTB) > −65dBc are obtained by using externally modulated AM-VSB CATV 77-channel erbium doped fiber amplifier (EDFA) repeated systems, which use single mode fiber (SMF) and reverse dispersion fiber (RDF) as a compensation device. H.-H. Lu [10] proposed and demonstrated an externally modulated NTSC 77-channel erbium-doped fiber amplifier (EDFA) repeated system employing FP etalon at the receiving site to improve system performance. The function of FP etalon is to suppress the spectral width and to ameliorate the fiber dispersion, in which resulting in system with better performance. For a successful deployment of long-haul fiber optical CATV transport systems, the cost and complexity are obviously considerable concern. Fabry-Perot (FP) etalon, with an optical characteristic of narrow optical band-pass filter, can reshape the laser spectrum and reduce the spectral width [10]. It is expected to have good performance in fiber optical CATV transport systems. In addition, FP etalon can be very compact and low cost. Germanov.[9] described a new method to calculate the CSO and CTB intermodulation distortions spectrum for fully frequency range of coaxial cable CATV amplifier for any input frequency plan. The calculation takes into account the output level ripple and pre-emphasis, using frequency dependent intermodulation coefficient. Hai-Han Lu[11] proposed a four wavelength bi-directional Dense Wavelength Division Multiplexing (DWDM) CATV system that uses Chirped Fiber Grating (CFG) as the dispersion compensation device to reduce the fiber dispersion and Cross Phase Modulation (XPM) induced crosstalk simultaneously. Excellent performance of CSO (>72dB) and CTB (>69dB) were obtained over a 50km Single Mode Fiber (SMF) transport. Chia-Hsiung Chang[12] use a repeater less bi-directional wavelength division multiplexing(WDM) transmiss- ion system for delivery of AM-VSB CATV analog video signals over conventional single mode fiber (SMF). He used and compared two kinds of multiplexers, the optical circulator and the WDM multiplexer configurations, for supporting bi-directional operation. Optical circulator offers satisfactory system performance of CSO > 71.2dBc, and CTB > 63.5dBc. Weissleder et al. [13] analyzed and modified the noise level by using Advanced design Software (ADS) in analog CATV systems. Po Yi Wuetal. [14] used unconverted technique and a phase modulation technique to reduce noise and distortion and improve the performance of carrier- to-noise ratio, composite sec-ond order and composite triple beat (>73/72dBc). Hai-Han Lu [15,16] proposed and demonstrated light wave transport systems to improve the CSO and CTB performance by using electro absorption modulator and optical single sideband modulation technique. Here CATV and microwave signals are transmitted simultaneously. The lower frequency side mode injection locked technique used in directly modulated fiber optical CATV transport system largely increases the resonance frequency of laser and improves the and CTB over 100 km single-mode-fiber(SMF) transmission. We have already achieved good results within MIMO-OFDM system with OADM recently for optical-OFDM system and Monitoring and Com-pensation of Optical Telecommunication Channels [17–22]. Amandeep Kaur [23] described the concept of Intermediation distortion for the same. Then experimental cum simulative analysis was made for different tones varies from two to twenty tone systems were analyzed based on their power spectrum.

In this paper, after having comprehensive literature review of System under consideration, we firstly describe the concept of Intermediation distortion. Then experimental cum simulative analysis is made for different input sources in the systems are analyzed. Also the results with external modulators are considered as the decrease in the CSO and CTB distortions could be seen with its use. Based on their power spectrums results are concluded in the last.

III. MEASUREMENT OF INTERMODULATION DISTORTION

The distortion of the signal in CATV transmission is due to the deviation of the signal from the linearity which is known as intermodulation distortion such as Composite second order distortion (CSO) and Composite triple beat distortion (CTB). Non-linearity in the propagation characteristics of the fiber transmission produces the new frequencies of the fi +fj and fi −fj. These frequencies lie in the transmission bandwidth and distorted the signal. The power level of the second and third order distortion known as Composite second order and Composite triple beat distortion, for specific channel are normalized to the carrier power of the channel and measured in dBc units.

If f1 and f2 are frequencies of two tones, then third order distortion products are occurring in both sides of these tones at 2f2 −f1 and 2f1 −f2. Assuming the power levels of two tones is equal. IMD3 is the difference of power of fundamental signals and third order products .it is defined as following:

$$IMD3 = Po - Po3 \tag{1}$$

Po3 is the power level of output third order product. Po is the power level of one the fundamental tone.

Calculate the intercept point of output third order product.

$$\frac{O}{P} = \frac{IMD_3}{2} + P_o = \frac{1}{2}(3P_o - P_{o3})\ \mathrm{db} \tag{2}$$

The improved results seen are due to the use of FP etalon to decrease the line width of the optical signal, in which leading to the reduction of the fiber dispersion. These parameters can also be decreased when we use external modulator in the system. CSO and CTB distortions can be stated as :

$$CSO = 10\log\left[\frac{mD\lambda^2Lf}{4c} + \sqrt{16(\Delta\tau)^2} + \sqrt{\frac{4\lambda_c^2L^2\pi^2f^8}{c^2}}\right] + 10\log N_{CSO} + 6 \tag{3}$$

$$CTB = 10\,log\left[\frac{4m^2D^2\lambda_c^4L^2f^2}{4c} + (4(\Delta\tau)^2 + 4\pi^2 f)\right] + 10logN_{CTB} + 6 \tag{4}$$

where D is the dispersion coefficient, λ_C is the optical carrier wavelength, L is the fiber length, f is the RF frequency, (   ) is the fiber dispersion, is the spectral width, N_{CSO} and N_{CTB} are the product counts of CSO and CTB. The use of FP etalon lets the spectral width to change from a broad width into a narrow width, leading in lower fiber dispersion. Then there would be significant reductions in the CSO/CTB distortions.

IV. EXPERIMENTAL SETUP AND RESULTS IN FIBER OPTIC BASED CATV SYSTEM.

In fiber based CATV transmission system use single wavelength for transmission on which multiple RF analog channels are modulated. In digital communication system BER is the most common performance matric while a distortions measurement is critical matric in analog transmission systems. We can produce these results through simulations in OptSim.

Three tones direct modulated CATV transmission system is considered in fig1. In this transmission system single Sine-wave frequency generator is used to generate a comb of three frequencies at 62.5 MHz, 125 MHz and 187.5 MHz.

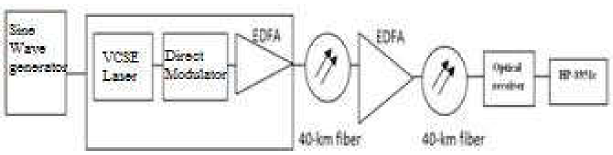


Fig.1. Multitone directly modulated with VCSEL CATV transmission system.

These three frequencies are then modulated onto 1550 nm wavelength by Vertical Cavity Surface Emitting Laser (VCSEL and then propagated onto 80 km single mode optical fiber cable to PIN based opti-

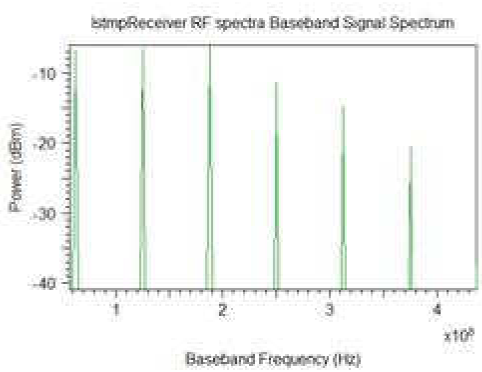


Fig.2. RF spectra for VCSEL directly modulated CATV transmission system

-cal receiver. Spectrum analyzer is used to view the RF spectra,fig.2, to measure the CSO and CTB distortions and other distortions in the system. Fig.2 shows the CSO = −11.5 dBc (at f = f1+ f3= 250 MHz and f = f2+ f3= 312.5 MHz), and CTB = −15.2dBc (f = f1+ f2+ f3= 375 MHz).

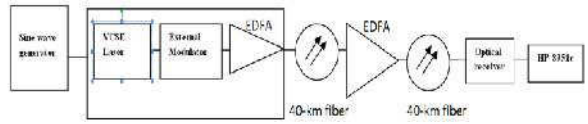


Fig .3. Multitone externally modulated with VCSEL CATV transmission system

Then three tones direct modulated CATV transmission system is considered, fig.3. In this transmission system single Sine-wave frequency generator is used to generate a combination of three frequencies at 62.5 MHz, 125 MHz and 187.5 MHz. These three frequencies are then modulated onto

1550 nm wavelength. The system uses VCSEL with external modulator and then signal is propagated onto 80 km single mode optical fiber cable to PIN based

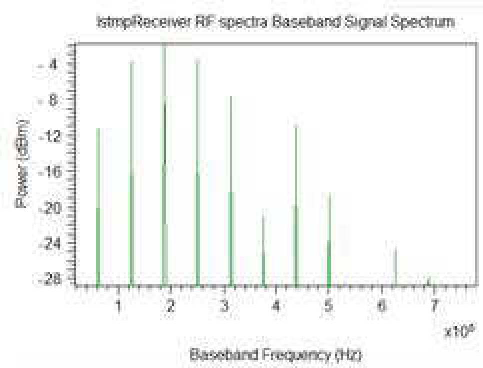


Fig. 4. RF spectra for VCSEL externally modulated CATV transmission system

optical receiver. With the use of external modulator we can see the decrease in values of CSO and CTB distortions as shown in RF spectra,fig.4. The values are CSO = −4.5 dBc (at f = f1+ f3= 250 MHz and f = f2+ f3= 312.5 MHz), and CTB = −8.2 dBc (f = f1+ f2+ f3= 375 MHz) .

Further, three tones direct modulated CATV transmission system is considered, fig.5. In this transmission system single Sine-wave frequency

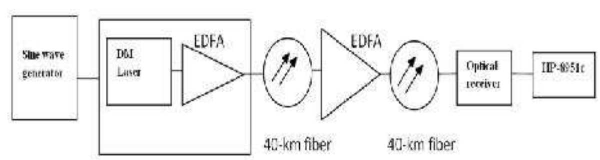


Fig.5. Multitone directly modulated CATV transmission system using DM Laser

generator is used to generate a combination of three frequencies at 62.5MHz, 125MHz and 187.5MHz. These three frequencies are then modulated onto 1550nm wavelength by DFB laser diode modulator and then propagated onto 80km single mode optical fiber cable to PIN based optical receiver. Spectrum

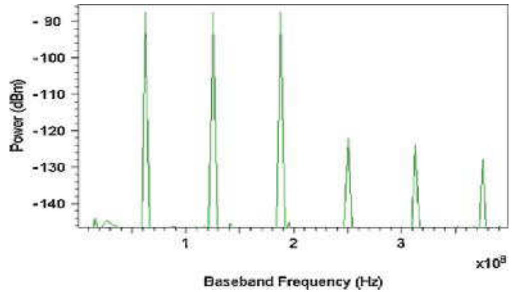


Fig.6. RF spectra for directly modulated three tone CATV transmission system using DM laser

analyzer is used to the view the RF spectra to measure the CSO and CTB distortions and other distortions in the system. Fig. 6 shows the CSO=−36dBc (at f=f1+f3 =250MHz and f=f2 +f3 =312.5MHz), and CTB=−41dBc (f=f1 +f2 +f3 =375MHz).

Finally, figure 7 shows the CATV transmission system DM Laser with FP filter. In this transmission system single Sine-wave frequency generator is used to generate a comb of three frequencies at 62.5 MHz, 125 MHz and 187.5 MHz. These three frequencies

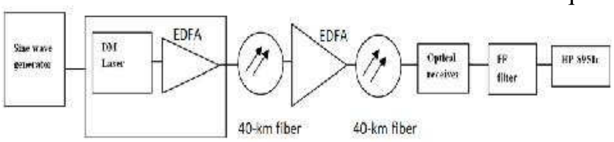


Fig .7. Multitone directly modulated with DM laser with FP filter CATV transmission system

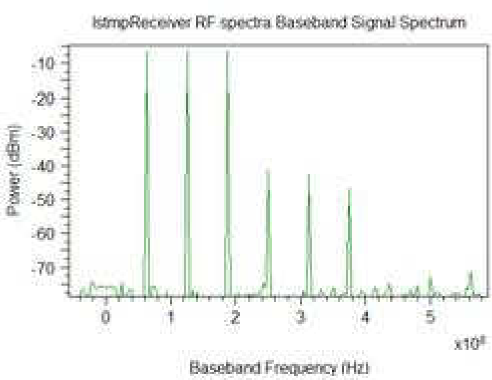


Fig 8. RF spectra for directly modulated CATV transmission system using DM laser with FP filter

are then modulated onto 1550 nm wavelength and then propagated onto 80 km single mode optical fiber cable to PIN based optical receiver. With the use of FP filter the CSO and CTB distortions are further decreased. The RF spectra to measure the CSO and CTB distortions and other distortions in the system. Fig.8 shows the CSO = −17 dBc (at f = f1+ f3= 250 MHz and f = f2+ f3= 312.5 MHz), and CTB = −22 dBc (f = f1+ f2+ f3= 375 MHz).

V. CONCLUSION

Intermodulation distortion in optical fiber based CATV transmission system is recognized by the second and third order harmonic sidebands in power spectra. In optical fiber based CATV transmission system distortion is much less than Coaxial based CATV transmission systems. Optical attenuator and FP filter is very effective in optical fiber based CATV transmission systems in order to decrease the intermodulation distortion such as CSO and CTB. Here we find the results of CSO and CTB distortions through simulations in Opt-Sim. We find that the Three Tone direct modulated transmission system using VCSEL with external modulator have better performance of CSO > −4.5dBc and CTB > −8.2 dBc at three frequencies 62.5MHz, 125MHz and 187.5MHz. We presented an externally modulated NTSC 77-channel EDFA-repeated system employing FP etalon at the receiving site to overcome fiber dispersion. In comparison with the conventional externally modulated fiber optical CATV transport systems, good performance of CSO/CTB were achieved over a 40-km SMF transport. Employing a simple FP etalon in fiber optical CATV transport systems is useful in real networks, as it is simple, passive, and potentially low cost. This systems are suitable for the long-haul fiber optical CATV transport system.

REFERENCES

[1]. M.R.Philips, Nonlinear distortion generated by dispersive transmission of chirped intensity modulated signal, IEEE Photon. Technol. Lett. 3 (May) 1991, pp: 481–483.

[2]. B. Clesca, Second order distortion improvements or degradation brought by erbium Doped fiber amplifier in analog links using directly modulated lasers.

[3]. E.E. Bergmann, Dispersion induced composite second order distortion at 1.55_m, IEEE Photon. Technol. Lett. 3 (January) (1993) pp: 59–61.

[4]. Hai-Han Lu, Ching-Ting Lee, Directly modulated CATV transmission system using half split band and wavelength division multiplexing techniques, IEEE Photon. Technol. Lett. 10 (November (11)) 1998.

[5]. J. Yu, Linearization of 1.55_m Electro absorption Modulated laser by distortion emulation and reversal for 77channel CATV Transmission, IEEE Photon. Technol. Lett. 10 (March (3),1998.

[6]. Hai-Han Lu, Performance comparison between DCF and RDF disper-sion compensation in fiber optical CATV systems, IEEE Trans. Broadcast. 48 (December (4)),2002.

[7]. Wanjiun Liao, Adaptive slot allocation in DOCSIS based CATV networks, IEEE Trans. Multimedia 6 (June (3)) 2004.

[8]. Raffaella Vinzio, Analytical formulation of distortion and chirp in CATV DFB lasers including spatial hole burning, IEEE J. Quant. Electron. 34 (February (2)),1998.

[9]. V. Germanov, Calculating the CTB/CSO spectrum of CATV amplifiers and optical receivers, IEEE Trans. Broadcast. 44 (September (3)) ,1998.

[10]. Hai-Han Lu, CSO/CTB performances improvement in a bi-directional DWDM CATV system, IEEE Trans. Broadcast. 50 (December (4)) (2004).

[11]. Rostami A. and A.Salmanogli, "Investigation of light amplification in Si-nano crystal-Er doped optical fiber," Progress in Electromagnetics Research B, Vol. 9,27-51,2008.

[12]. Chia-Hsiung Chang, Demonstration of repeterless bidirectional transmission of AM-VSB CATV signals over conventional single mode fiber, IEEE Photon. Technol. Lett. 12 (June (6)),2000.

[13]. H. Weissleder, Initial lessons learned in level planning for mixed signal transmission in CATV networks, IEEE Trans. Consumer Electron. 46 (May (2)),2000.

[14]. Po-Yi Wu, An unconverted phase modulated fiber optical CATV Transport system, J. Lightwave echnol. 29 (August (16)) (2011).

[15]. Hai-Han Lu, CATV radio on fiber transport system based on EAM and SSB mod-ulation techniques, IEEE photon. Technol. Lett. 16 (November (11)),2004.

[16]. Hai-Han Lu, Improvement of fiber optic CATV transport system performance based on low frequency side mode injection locked frequency, IEEE Photon. Technol. Lett. 20 (March (5)),2008.

[17]. K.S. Bhatia, T.S. Kamal, R.S. Kaler, Peak-to-average power ratio reduction using coded signal in optical-orthogonal frequency division multiplexing systems, IET Optoelectron. 6 (5),2012,pp: 250–254.

[18]. K.S. Bhatia, R.S. Kaler, T.S. Kamal, Design and simulation of optical-ofdm sys-tems, J. Russ. Laser Res. 33 (5),2012,pp:202–208.

[19]. K.S. Bhatia, T.S. Kamal, R.S. Kaler, An adaptive compensation scheme-based coded direct detection optical–orthogonal frequency division multiplex (OFDM) system, Comp. Electrical Eng. 38,2012pp: 1573–1578.

[20]. K.S. Bhatia, R.S. Kaler, T.S. Kamal, R. Randhawa, Simulative analysis of inte-grated DWDM and MIMO-OFDM system with OADM, Optik 124 2013,pp: 117–121.

[21]. K.S. Bhatia, T.S. Kamal, Modeling and simulative performance analysis of OADM for hybrid multiplexed Optical-OFDM system, Optik 2013.

[22]. K.S. Bhatia, R.S. Kaler, T.S. Kamal, R. Randhawa, Monitoring and compensation of optical telecommunication channels by using optical add drop multiplexers for optical OFDM system, J. Opt. Commun. 33 (1) 2012.

[23]. A.Kaur, K.S.Bhatia, A.Kaur, Kulwinder Singh, Simulation analysis of IMD in multitone analog CATV transmission systems, Optik 2013.

Performance Investigation of GE-PON Fiber To The Home Network Under Varying Data Rates and Users

Sooraj Parkash
CT Institute of Technology & Research, Maqsudan, Jalandhar.
soorajparkash879@gmail.com

Dinesh Kumar
CT Institute of Technology & Research, Maqsudan, Jalandhar.
dinesh14313@yahoo.co.in

Anurag Sharma
CT Institute of Technology & Research, Maqsudan, Jalandhar.
er.anurags@gmail.com

Rahul Malhotra
CT Institute of Technology & Research, Maqsudan, Jalandhar.
rahulmalhotra@ctgroup.in

Abstract-- **In this paper we analyze the effect on bit error rate (BER) by varying the fiber length from 100km to 200km.Also we have observed the effect on BER by increasing the number of users from 32 to 64 and data rate from 1.25Gbps to 5Gbps. An optical line termination (OLT) device is installed in the central office (CO),and an optical network termination (ONT) device is installed on the other end,in or near each home or business site. Fiber distribution is done using a tree-and-branch architecture.**

Keywords—FTTH, PON, EDFA, GEPON, SMF, CO, DM.

I. Introduction

The deployment of PONs might not be the best option depending on different geographical considerations. To overcome these PON limitations, optical–wireless convergence is introduced to provide cost-effective, high speed ubiquitous access to end users [1,2]. Moreover, PON is also identified as a potential backhaul solution for the NG wireless broadband access network due to its cost effectiveness and high bandwidth capacity[8].Passive optical network (PON) access architecture is the accepted choice of triple-play (voice, video, and data) service delivery from service providers to the end users in FTTH (fiber-to-the-home) access networks. Due to certain characteristics and requirements unique to the LTE network, improved performance from LTE–PON converged networks can be achieved by using a proper architecture for the integration and also by using an efficient resource allocation mechanism[3,4,5]. A few studies have been reported on the LTE–PON integration recently [9, 10].

However, improving the performance of the directly integrated Gigabit Ethernet PON (GEPON)–LTE network, by using improved resource handling mechanisms, has so far received minimal attention.

Three major PON technologies are currently accepted as the basis for the FTTH deployments: Broadband PON (BPON), Gigabit PON (GPON), and Ethernet PON (EPON or GEPON).GEPON architecture is specified by the IEEE 802.3ah standard and supports Gigabit Ethernet 1.25 Gbps transmission rate for data component in both down-and up-link configuration. First, we describe in details the configurations for Central Office OLT and single end-user ONT and then will generalize the treatment to all 32,48,64 users. In this example we consider downstream configuration of GEPON with bit rate 1.25, 2.5, 5 Gbps and support for triple-play. The triple-play service is realized as a combined results of data, voice and video signals. The high speed internet component is represented by a data link with 1.25, 2.5, 5 Gbps downstream bandwidth. The voice component can be represented as VOIP service (voice over IP, packet-switched protocol) and can be combined with data component in physical layer simulations. In the centralized mechanism, a central controller placed at the optical line terminal (OLT) governs the bandwidth allocation for the wireless subscriber stations where the bandwidth requests from the subscribers are directly sent to this central controller [6, 7]. Finally, the video component can be represented as a RF video signal (traditional CATV) or as IPTV signal that also can be combined with data .In our case we consider the former case with RF video link. To optimize the bandwidth in PON the transmission through the optical fiber path employs the CWDM technique with data/voice component transmitted at wavelengths in the range of 1480-1500nm ,and video within the 1550-1560nm range. This example demonstrates an OptSim design for FTTH GEPON link. We consider the typical GEPON FTTH design for downlink with 32,48,64 subscribers and 100,150,200 km reach. Passive optical network (PON) access architecture is the accepted choice of triple-play (voice, video, and data) service delivery from service providers to the end users in FTTH (fiber-to-the-home) access networks. There are three major PON technologies are currently accepted as the basis for FTTH deployments: Broadband PON (BPON), Gigabit PON (GPON), and Ethernet PON (EPON or GEPON). The GEPON architecture is specified by IEEE802.3ah standard and supports Gigabit Ethernet 1.25 to 5 Gbps transmission rate for data component in both down- and up-link configuration.

II. SYSTEM MODEL AND SIMULATION

Fig.1 consists of sine generator, summer, electrical generator, DM laser, pre-amplifier and optical MUX. Sine generator produces rectangular pulses. Electrical generator produces short electrical pulses, discrete mode laser used to produce narrow line width optical pulses, pre amplifier amplifies the small electrical pulses for further amplification process.In fiber-optic communication, a single-mode optical fiber (SMF) is an optical fiber designed to carry light only directly down the fiber - the transverse mode. Modes are obtained by combining Maxwell's equations and the boundary conditions, these modes define the way in which the wave travels through space that is how the wave is distributed in space. Waves can have same modes but have different frequencies. This is that case in single-mode fibers where we can have waves with different frequencies, but having the same mode, which means they are distributed in space in the same way, and that gives a single ray of light. As we know the ray travels parallel to the length of the fiber, called as transverse mode since its electromagnetic vibrations occur perpendicular (transverse) to the length of the fiber. In this setup, no booster amplifier has been used.

An Erbium Doped Fiber Amplifier (EDFA) consists of a piece of fiber of length “L”, whose core is doped with Erbium ions. Such ions can be called as simple two-level systems, i.e. they have only two energy states: 1) A fundamental state 2) An excited state. A strong “pump” laser light at the proper wavelength (usually 980 nm or 1480 nm) is propagated into the core of the fiber in order to excite its ions. Excited ions, when “hit” by an input “signal” photon, have a certain probability (depending on the wavelength of the input photon) of releasing by stimulated emission a photon identical to the hitting one. The releasing of energy of the stimulated photon brings the excited ion to its fundamental state.The Fiber Optic Splitter, also named as “Beam splitter”, is based on the quartz substrate of integrated waveguide optical power distribution device, same as the coaxial-cable transmission system, the optical network system (ONS) also needs to be an optical signal coupled to the branch distribution, which requires the fiber optic splitter, which is one of the most important device (passive) in the optical fiber link, it is an optical fiber tandem device with many input and output terminals, especially applicable for a optical network (passive)

CENTRAL OFFICE (TRANSMITTER)

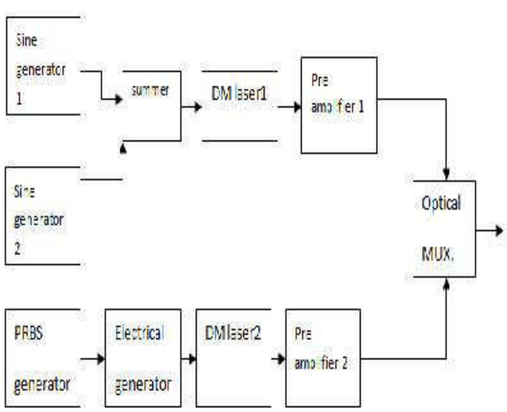


Fig. 1. Transmitter design for GEPON network

(EPON,GPON,BPON,FTTX,FTTH etc.) to connect MDF and terminal equipment for achieving the branching of the optical signal.

TRANSMISSION MEDIUM (FOR GEPON)

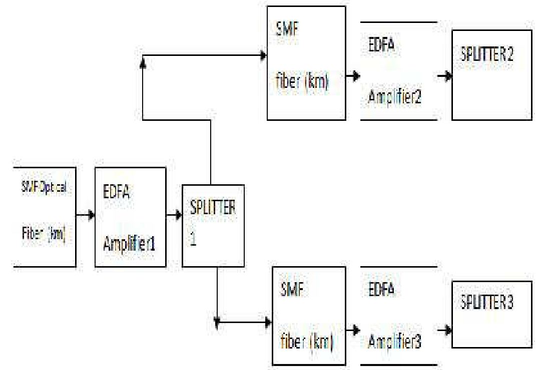


Fig. 2. Transmission medium design for GPON network.

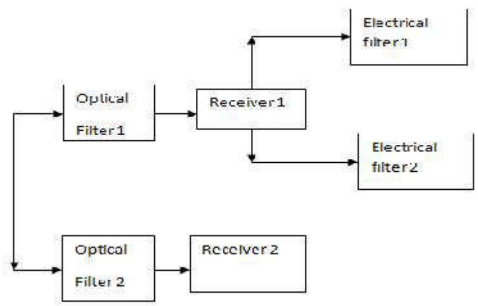


Fig.3. Receiver design for GPON network for user 1 only.

Fig. 2. shows the components used are single mode fiber ,EDFA and optical splitters. This diagram shows the transmission medium for GEPON. The output is amplified by using EDFA and the output of EDFA splitted by using optical splitter.The output of the splitter goes to receiver of GEPON. In the receiver various components are used optical filter , receiver, electrical filter. Fig3.shows an optical receiver in which optical filters and electrical filters are used.

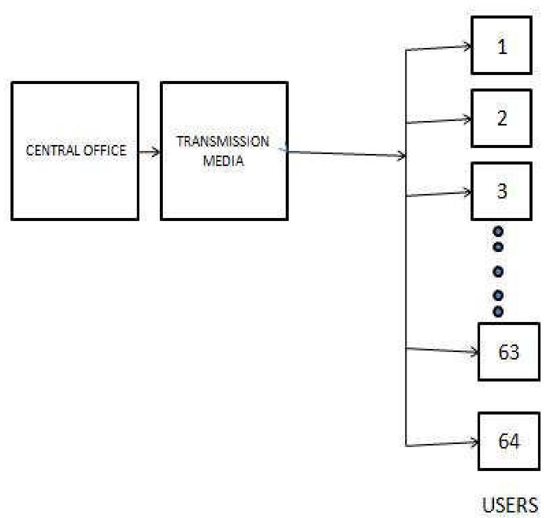


Fig. 4. Complete simulation setup OPTSIM

In figures 5,6,7 the graphs show the relation between bit error rate(BER) and fiber length(fiberL in Km).When the fiber length increases BER decreases and vice versa. So graph shows that when fiber length increases the signal strength degrade.

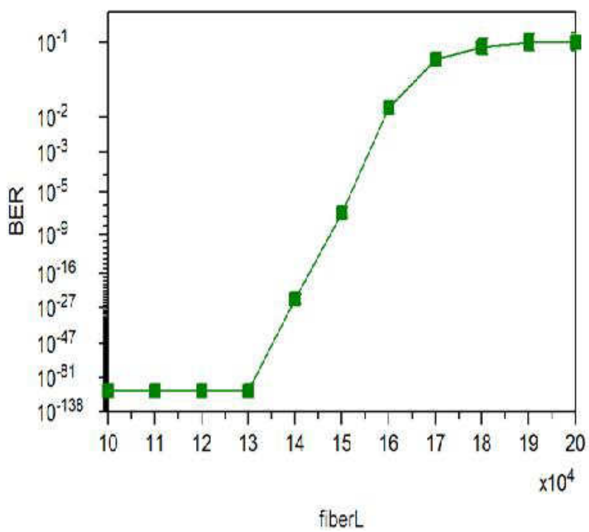


Fig .5. BER vs Fiber length for 32 users

Table I. Effect on bit error rate by using variable data rates, fiber length and users.

S No.	*Data rate Gbps.*	*Fiber length(km)*	*Users*	*BER*
1	1.25	100	64	1.4473E-59
2	1.25	100	48	6.2724E-62
3	1.25	100	32	1.0927E-63
4	1.25	150	64	4.2911E-48
5	1.25	150	48	5.846E-50
6	1.25	150	32	1.0458E-51
7	1.25	200	64	1.1971E-55
8	1.25	200	48	1.9328E-58
9	1.25	200	32	6.3689E-61
10	2.5	100	64	5.0805E-07
11	2.5	100	48	4.8426E-07
12	2.5	100	32	4.7022E-07
13	2.5	150	64	3.0488E-10
14	2.5	150	48	2.6239E-10
15	2.5	150	32	2.3669E-10
16	2.5	200	64	3.0738E-18
17	2.5	200	48	9.0512E-19
18	2.5	200	32	3.4096E-19
19	5	100	64	8.35E-02
20	5	100	48	6.91E-02
21	5	100	32	7.00E-02
22	5	150	64	5.56E-02
23	5	150	48	9.25E-02
24	5	150	32	9.23E-02
25	5	200	64	3.58E-02
26	5	200	48	3.61E-02
27	5	200	32	3.58E-02

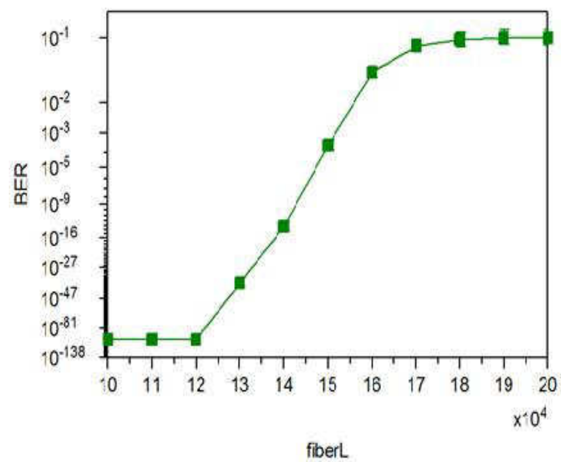


Fig .6.BER vs Fiber length for 48 users

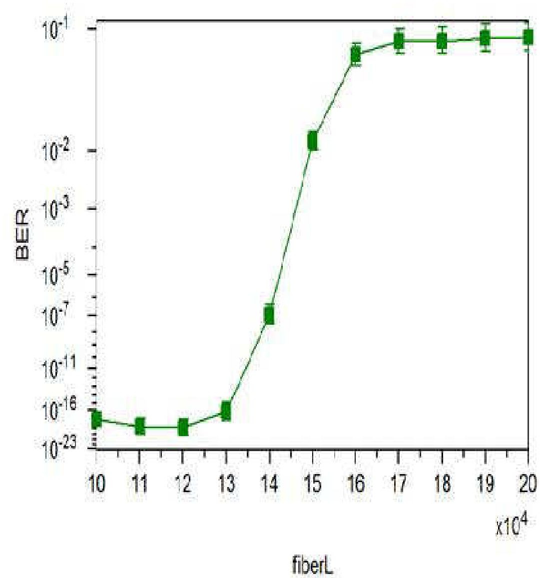


Fig .7. BER vs Fiber length for 64 users

III. CONCLUSION

GEPON is next generation optical convergence is most cost effective and providing high bandwidth for data transmission. The GEPON network is triple layer network which is suitable for transmission of voice, video and data over a single line. In this paper we observe that if the length of fiber increases then

BER decreases. And if the fiber length is decreases then BER increases. This setup is capable for handling 200 users with fiber length 160km and providing BER 1.7953e-009 without booster amplifier with data rate 2.5 Gbps. The main purpose of using GEPON network is to provide better quality of service (QOS) and groove baseband recording (GBR) traffic in convergence areas.

REFERENCES

[1] Gangxiang ,S. Tucker,R.S ,Chang-Joon, C., "Fixed mobile Convergance architecture for broadband access: integration of EPON and WiMax" IEEE Communication Mag. 45(8), pp.44–50 (2007).

[2] Sarkar, S.Dixit, S. Mukherjee, B. Hybrid wireless optical broadband-access network (WOBAN) " a review of relevant challenges, J. Lightwave Technol". 25(11),3329–3340 (2007).

[3] Ranaweera, C., Wong, E., Lim, C. " Nirmalathas, A.: Next generation optical–wireless converged network architectures. IEEE Netw". 26(2), 22–27 (2012).

[4] Ali, M.A, Ellinas, G. Erkan, H. Hadjiantonis, A. Dorsinville, R."On the vision of complete fixedmobile convergence. J. Lightwave Technol" 28(16), 2343–2357 (2010).

[5] Madamopoulos, N., Peiris, S., Antoniades, N.Richards, D. Pathak,B.Ellinas, G.Dorsinville, R.,Ali,M."A fully distributed 10 GEPON- based converged fixed-mobile networking Transport infrastructure for next generation broadband access.IEEE/OSA J. Opt.Communication. Network."4(5), 366–377(2012).

[6] Jung , B . Choi, Han, Y. Kim, M. Kang, M. "Centralized Scheduling mechanism for enhanced end-to-end delay and QoS support in integrated architecture of EPON and WiMAX. J.Lightwave Technologies" 28(16),2277–2288 (2010).

[7] Luo, Y. Ansari, N.Wang, T. Cvijetic, M. Nakamura, S. "A QoS architecture of integring GEPON and WiMAX in the access network". In. " Sarnoff Symposium, 2006 IEEE, pp.1–4 (2006).

[8] Ranaweera,C.S.Iannone,P.P., Oikonomou, K.N.,Reichmann, K.C., Sinha, R.K.: "Cost optimization of fiber deployment for small cell backhaul".In."Optical Fiber Conference and National Fiber Optic Engineers Conference (OFC/NFOEC)". NTH3F.2 (2013).

[9] Ranaweera, C.W ong, E.Lim, C.Nirmalathas , A " Next Generation optical–wireless converged network architectures". IEEE Networking. 26(2), 22–27 (2012).

[10] Ali, M.A., Ellinas, G., Erkan, H., Hadjiantonis, A., Dorsinville, R " On the vision of complete fixedmobile Convergence J. Lightwave Technol".28(16),2010,pp:2343–235

Effect of Crosstalk on DWDM Ring Configuration

Amandeep Kaur
Department of ECE
CT Institute of Technology & Research
Jalandhar, Punjab, INDIA

Sabba Amin
Department of ECE
CT Institute of Technology & Research
Jalandhar, Punjab, INDIA

Amanjot Singh
Department of ECE
CT Institute of Technology & Research
Jalandhar, Punjab, INDIA

Ashima
Department of ECE
GNDU
Amritsar, Punjab, INDIA

***Abstract*--In this paper, We analyzed the performance of a DWDM ring at the different values of crosstalk BER and eye diagram at different nodes. We have observed that at crosstalk value -20 the corresponding BER is e^{-26}.**

Keywords-- DWDM,SONET,BER,OADM

I. INTRODUCTION

Dense Wavelength Division Multiplexing (DWDM) is an optical networking technique. DWDM devices combine the output from several optical transmitters for transmission across a single optical fiber [4].In this each fiber have a set of parallel optical channels each using slightly different wavelengths. It hires light wavelengths to transmit data parallel-by-bit or serial-by-character. Currently, the dense wavelength-division multiplexing (DWDM) technology consummate multiplexing of 150–300 wavelengths in one fiber with 10–40 Gbs transmission rate per wavelength[1]. The wavelength channels in one fiber can be used for both directions or two fibers are used with each for one direction [2]. A media-access protocol is required to control the transmissions of the various network nodes to avoid collisions and to manage contention for the network bandwidth [3]. A key advantage to DWDM is that it's protocol and it is bit-rate-independent. DWDM based networks are capable to transmit data in IP, ATM, SONET /SDH, and Ethernet. Therefore, DWDM based networks can carry different types of traffic at different speeds over an optical channel[4].Typically,DWDM is capable to transport up to 80 wavelengths in the Conventional band or C band spectrum, with all 80 channels in the 1.55 µm region. DWDM referoriginally to optical signals multiplexed within the 1550 nm band so as to leverage the capabilities (and cost) of erbium doped fiber amplifiers (EDFAs), which are effective for wavelengths between approximately 1525-1565 nm (C band), or 1570-1610 nm (L band) [5]. The channel spacing for DWDM is 0.8/0.4 nm (100 GHz/50 GHz grid), this low channel spacing allows transmitting simultaneously much more information. DWDM can increase capacity by 10-100 times without the need for a new optical fiber[6].DWDM fibers can transmit at speed of up to 400Gbps[7].DWDM is extremely adaptable and versatile as well, in that it can vary the kind of data as well as the wavelength at which that data travels[8]. In this paper we worked on DWDM ring configuration. A DWDM ring network comprises of a fiber in a ring configuration that completely links nodes; few systems utilize two fiber rings for network protection against any unwanted disturbances [9].

II. SIMULATION SETUP

Figure 1 demonstrates an equivalent design of the OC-192 ring configuration in OPTSIM. In this case, the DWDM ring consists of six nodes and six fiber spans. Eight wavelengths are used with 3 head-end nodes adding/dropping 5 channels at the time and 3 serial OADM nodes adding/dropping two channels.

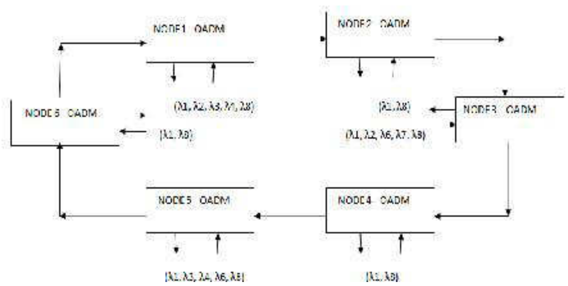


Fig.1. DWDM ring configuration

We used here unidirectional fiber link with signals propagating in clock-wise direction. In this, we use Corning Leaf fiber. We used 8 wavelengths at 50 GHz (0.4 nm) spacing starting at 1550 nm wavelengthi.e 1= 1.550µm, 2=1.5504µm& so on. After each node we put also a plotter block Multi-Plot to observe an optical spectrum evolution along the ring. The Fig. 2.shows head end nodes 1 OADM with 5 wavelengths adding/dropping at (1, 2, 3, 4, 8)& null signal is given at the rest of ports. Similarly, at OADM 1&2 adding/dropping wavelengths at (1, 2, 6, 7, 8) and (1, 3, 4, 6, 8) respectively. At serial nodes 2,4and 3 OADM with 2 wavelengths adding/dropping at 1 and 8. The OADM block can specify parameters like crosstalk level, switching configuration, first channel wavelength and channel spacing, optical filter bandwidth used in demultiplexing.

The OADM block is modeled as a compound component (CC) or hierarchy block & it is composed inside from other OPTSIM library blocks. Each OADM node has one input and one output for line signal, and also 8 inputs/outputs for added/dropped wavelengths. At headend node 1 five single-channel NRZ-modulated transmitters modulated at 10Gbps bitrate (OC-192 rate) are connected to add-ports 1,2,3,4 and 8.

Since the OPTSIM requires that all input ports for CC models to be connected we use NullSignal model and connect it to the other 3 ports. At the output of CC we can connect only the ports of interest – in this case ports 1, 2, 3, 4 and 8. Output from these ports are connected to optical receiver blocks and then to plotter model Multi-Plot which will provide plots for electrical signal output waveform, spectrum, and eye diagram. We analyzed the BER corresponding to its Crosstalk at three nodes 1,3 and 6.The figure 2 demonstrate the configuration of headend node 1.

III. RESULTS AND DISCUSSIONS

The BER performance against different values of crosstalk is analyzed and the graph is plotted using a plotter model multi-plot in OPTSIM.Fig. 3.shows graph between BER corresponding to its crosstalk.

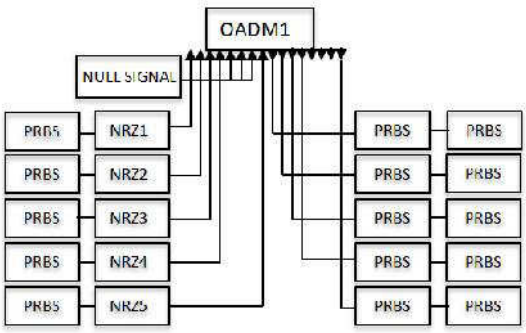


Fig. 2.Headend Node 1

From Fig. 3.it is observed that BER increases with increase in Crosstalk at node 1. The Fig. 4. Shows graph between crosstalk & BER at node 3, it is observed from Fig. 4. that BER decreases with increase in crosstalk in this DWDM ring. The Fig. 5. Shows eye diagram for dropping channels at node 3. The Fig. 7. shows eye diagram for dropping channels at node 6. It is clear from observations, in Fig. 5. & Fig. 7. that eye is clearly open, this means BER decreases.

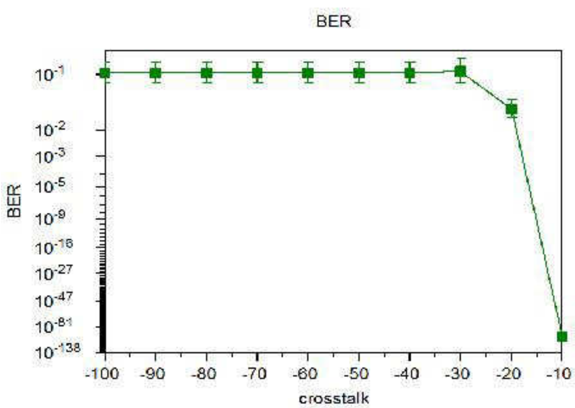


Fig. 3. Plot of BER and Crosstalk at node 1

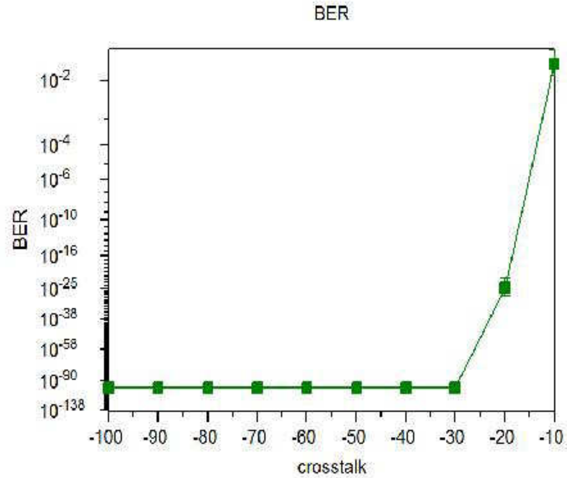


Fig. 4. Plot of crosstalk and BER at node

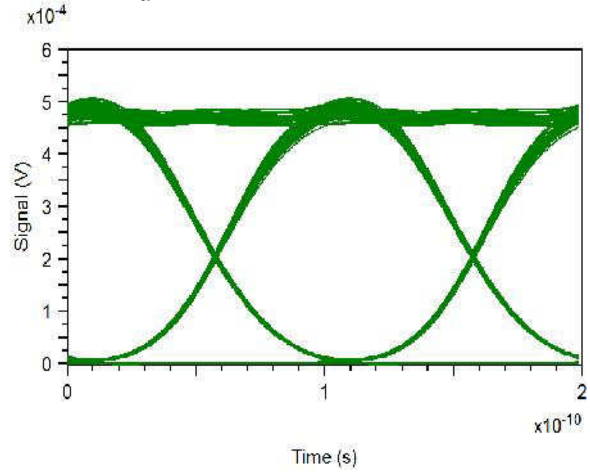


Fig. 5. Eye diagram for dropping channels at node 3

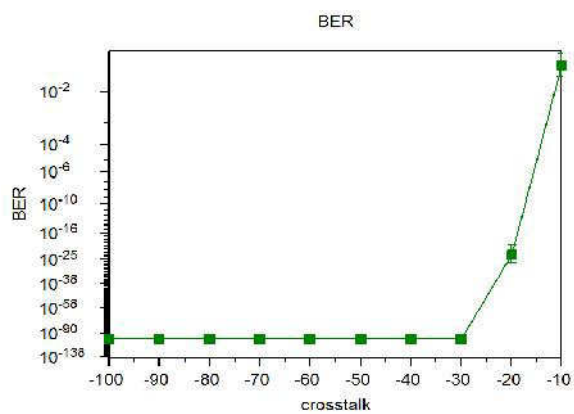


Fig. 6.Plot of crosstalk and BER at node 6

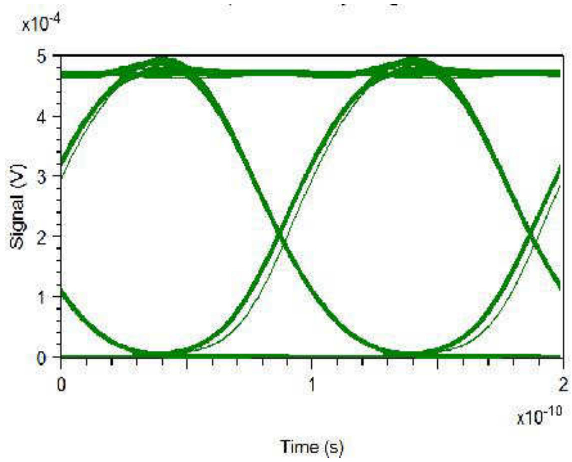


Fig. 7. Eye diagram for dropping channels at node 6

It is observed from Fig. 6. that the BER decreases with increase in Crosstalk at node 6.

Table 1 shows values of BER corresponding to change in Crosstalk at different wavelength and at different nodes

TABLE 1

Crosstalk	Wavelength	Node 1	Node 6
-10	1	7.1742e-036	7.0035e-005
	8	3.3880e-029	3.2953e-002
-20	1	3.1989e-002	3.2265e-093
	8	5.9346e-002	6.6046e-024
-30	1	1.0707e-001	1.5248e-028
	8	1.0242e-001	1.3081e-039
-40	1	1.0299e-001	2.3753e-028
	8	1.0550e-001	1.8975e-289
-50	1	1.0263e-001	9.1987e-026
	8	8.1358e-002	0.0000e+000
-60	1	1.0263e-001	1.7098e-025
	8	1.0631e-001	0.0000e+000
-70	1	1.0263e-001	2.1258e-025
	8	1.0631e-001	0.0000e+000
-80	1	1.0263e-001	2.2834e-025
	8	1.0631e-001	0.0000e+000
-90	1	1.0263e-001	1.4051e-013
	8	1.0631e-001	5.3807e-016
-100	1	1.0263e-001	2.0168e-029
	8	1.0631e-001	0.0000e+000

IV. CONCLUSION

We have observed the effect of Crosstalk in DWDM ring performance. We also found that the BER increases with increase in crosstalk in node 1 OADM and at node 3and node 6 the BER decreases with increase in crosstalk. This makes signal too much noisy.

REFERENCES

[1] Bhupesh Bhatia,VijayrajShokeen & R. K. Singh, "Evaluation of Optical Ring Network to Measure Inter-channel Crosstalk using "Eye Diagrams" ", International Journal of Computer Applications (0975 – 8887), Volume 40, No.15, February 2012,pp- 35-40.

[2] V.Khanaand Krishna Mohanta, "10Gbps Data Transmission and Implementation of DWDM Link",International Journal of Advanced Research (2013), Volume 1, Issue 7, 670-674.

[3] Shaowen Song, "DWDM And The Future Integrated Services Networks",IEEE Canadian Review - Spring / Printemps 2000, pp- 5-7.

[4] ReenaAntil, Pinki, Mrs. SonalBeniwal, "An Overview of DWDM Technology & Network", International journal of Science & Technology Research Volume 1.Issue 11,pp- 43-46, December 2012.

[5] Er. Karamjeet Kaur, Er. ChakshuGoel, "Threshold based Bit Error Rate Optimization in Four Wave Mixing Optical WDM Systems", American Journal of Engineering Research (AJER) e-ISSN : 2320-0847 p-ISSN : 2320-0936 Volume-03, Issue-03, pp-225-229,2014.

[6] Masahiro Ojima, D. Eng.Hiroyuki Nakano, D. Eng. Shinya Sasaki, D. Eng. ShoichiHanatani,"A Dense-wavelength-division-multiplexingOptical Network System",Hitachi Review Vol. 48 (1999), No. 4 pp-219-224.

[7] BishalPoudel," BER Analysis of an Optical Cross-Connect Due to Cross-talk in DWDM System",International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-3, Issue-3, pp- 84-86,February 2014.

[8] Vikrant Sharma and DalveerKaur, "Dense Wavelength Division Multiplexing - A Review",International Journal of Current Engineering and Technology ISSN 2277–4106, Vol.3, No.5 (December 2013), 2115-2117.

[9] S.S.Riaz Ahamed, "PERFORMANCE ANALYSIS OF DWDM",Journal of Theoretical and Applied Information Technology, 2005 - 2008 JATIT, pp- 590-594.

Investigate the Effect of Laser Source & Line Width on FWM Power in WDM System

Veerpal Kaur
Dept. of ECE.
Punjabi University
Patiala, Punjab
kaur.veerpal617@gmail.com

Kamaljit Singh Bhatia
Dept. of ECE
S.G.G.S.W.U,
Fatehgarh Sahib, Punjab
kamalbhatia.er@yahoo.com

Amrit Kaur
Dept. of ECE
Punjabi University
Patiala, Punjab
amrit_aks@yahoo.com

Harsimrat Kaur
Dept. of ECE
CTIEMT
Jalandhar, Punjab

***Abstract*-In long-haul WDM communication system, transmission characteristics are degraded by four wave mixing. In this paper we analysis the FWM probe signal with laser source & line width of laser source at various laser source power & chromatics dispersion in wave length division multiplexed system. Similarly Power of FWM signal is minimum,Qfactor maximum &bit error rate is minimized at optimized value of line width, laser source power and specific value of dispersion.**

Keywords--Four wave mixing, Wavelength Division Multiplexing, BER, Dispersion.

I. INTRODUCTION

The latest trend in light wave communication system is towards huge information transmission capacity by explosive growth in data rate communication by using wave length division multiplexing. [1-2] When signal power is high, then fiber refractive index become a function of power and linear response of optical system is lost.[1] FWM due to changes in refractive index with signal power is called Kerr effect. In non linear FWM process in which three frequencies f_i,f_j,f_k (k i,j)interact through third order nonlinearity of optical fiber to generate new frequency[1]

$$f_{ijk} = f_i + f_j - f_k \quad (1)$$

These new frequency components falls on the same transmission window and original signal get degraded. Thus three wave give rise nine new frequency component. The number of side bands use to FWM increase is given as[2].

$$M = \frac{N^3 - N^2}{2} (2)$$

Where N is number of channels and M number of newly generated side bands .Performance of WDM system degraded by low chromatic dispersion and non-linear effect of fiber link. [2-3] Conventional WDM system dispersion used to suppress the FWM. The cause of FWM analyzed in and suppression methods introduced.[7] Reduction of FWM investigated

in [6] optical phase conjugator as wave front reversal, which is used to suppress the FWM in OFDM system. [6] OPC support all the data type format.OPC simultaneously used to cancel the fiber loss, dispersion loss, Kerr effect in long communication. [8] By placing OPC in mid way in optical transmission system was designed.FWM is cancelled in WDM system [8].We have already achieved good results within Monitoring and Compensation of Optical Telecommunication MIMO-OFDM system with OADM [11-14].

In this paper, section II presents theoretical analysis of FWM performance degradation in intensity–modulated direct detection (IM-DD) WDM system with equal channel spacing and with varying chromatic dispersion, section III describe the experiment simulation setup for analyse the laser power and line width, result and discussion summarized in section IV and conclusion presented in section V.

II. FUNDAMENTAL OF ANALYSIS

FWM is one of major limiting factor in WDM optical communication system due low dispersion, equal channel spacing .As a result, Estimating power of FWM is very important for design and evaluation of WDM system. A well know formula used to estimate the Power of FWM derived by Hill et al[3] and later reformulated by efficiency depends upon phase mismatch factor by Shibata et al.[4].This new formula used for estimating FWM crosstalk is recently used in WDM system.[3].Power of FWM can be written as

$$P_{ijk} = \frac{\eta}{9} D^2 \gamma^2 P_i P_j P_k exp(-\alpha L) \left\{ \frac{[1-exp(-\alpha L)]^2}{\alpha^2} \right\} \quad (3)$$

Where $P_iP_jP_k$ are signal input power at frequency f_i ,f_j,f_k,(k i,j) respectively is the length of fiber, is the fiber attenuation coefficient the degeneracy factor D takes value equal to three or six for degenerate and non degenerate FWM, re

spectively. The nonlinear coefficient is given[1] by

$$\gamma = 2\pi n_2 / \lambda A_{eff} \quad (4)$$

Where A_{eff} effect area of fiber , vacuum wave length andn_2 non linear is refractive index. Also is the FWM efficiency which can be expressed as[4]

$$= \frac{\alpha^2}{^2+\Delta k^2}\left\{ + \frac{4\exp(-\alpha L)\sin^2(\Delta k L/2)}{1-\exp(-\alpha L)^2} \right\} \quad (5)$$

In this expression k is phase match factor which depends upon dispersion of fiber and channel spacing and writtenas [4]

$$\Delta k = \frac{2\pi\lambda_k^2}{c}\Delta f_{ij}\Delta f_{jk}\left[D_C + \frac{\lambda_k^2}{2c}(\Delta f_{ij} + \Delta f_{jk})\frac{dD_c(\lambda_k)}{d\gamma}\right] (6)$$

Where $f_{ij} = f_i - f_j$ channel spacing, D_c chromatic dispersion, $dD_c/d\lambda$ dispersion slope, and $_k$ is the wavelength at frequency f_k. Where k mismatch factor not a function of signal powers, so it calllinear phase matching factor. FWM efficiency depends on the fiber dispersion, the channel separation, and the fiber length, but not on the transmitted power [4].So as above expression dispersion increase phase match factor increase and efficiency decreases. Power of FWM directly proportional to the efficiency so power of FWM decrease by increasing dispersion. [10]So similarly power of FWM reduces directly by reducing laser power at expense of equal channel spacing. Similarly investigate the effect of line width at various laser source powersat 3 channel which are equally spaced at 50 GHz.

III. SIMULATION SETUP AND DESCRIPTION

Work is carried out using simulation to observe the effect of FWM in WDM system in presence of laser source power and line width at various laser source power. Then keeping into mind dependence of FWM power on laser source power as per equation (3) Three channels at frequencies (in THz) 193.025, 193.075, 193.125 with NRZ modulation format at 10Gbps launched into fiber at dispersion 6(ps/nm-km), attenuation constant 0.2(dB/km) with fixed amplifier gain as shown in Fig.1. Simulation parameters used were as followed in Table1 & Table2&Table3 to obverse the effect of FWM at probe channel .We analysis the system performance by QFactor, eye-diagram, output spectrum and bit error rate.

TABLE I. Simulation Parameter And Their Values (Pump Power)

Parameters	Values
Pump frequency (THz)	193.025to193.125
Probe frequency(THz)	193.075
Channel spacing (GHz)	50
Pump power(dBm)	0 to -100
Reference bit rate(Gbit/s)	10
Attenuation (dB/km)	.2
Fiber length(km)	100
Dispersion (ps/nm-km)	6
Preamplifier gain (dB)	25

TABLE II. Simulation parameter and their values

(Pump & Line Width)

Parameter	Values
Pump frequency (THz)	193.025to193.125
Probe frequency(THz)	193.075
Channel spacing (GHz)	50
Pump power(dBm)	0 to -40
Reference bit rate(Gbit/s)	10
Attenuation (dB/km)	.2
Fiber length(km)	100
Dispersion (ps/nm-km)	2
Line width(MHz)	10-40

TABLE III.Simulation parameter and their values (dispersion)

Parameters	Values
Pump frequency (THz)	193.025to193.125
Probe frequency(THz)	193.075
Channel spacing (GHz)	50
Pump power(dBm)	10
Reference bit rate(Gbit/s)	10
Attenuation (dB/km)	.2
Fiber length(km)	100
Dispersion (ps/nm-km)	0-10
Preamplifier gain (dB)	25

The schematic model for an optical communication system implementing the four wave mixing effect of different transmitter component.In this model N number of user with modulator driver, modulator and non-linear fiber & at receiver side demultiplexer, filter with PIN sensitity receiver is used to receive the signal.

IV. RESULTS & DISCUSSIONS

During simulation, we observed the system performance with different transmitter component like laser power,modulator driver and line width. We analyse in term of Q factor, FWM signal power, eye diagram, input& output spectrum.

A. *Effect of laser power*

We make to analyse the performance of WDM system based upon laser source power. It obverse that Qfactor increase with increasing the lasersignal power. As shown in fig 2.

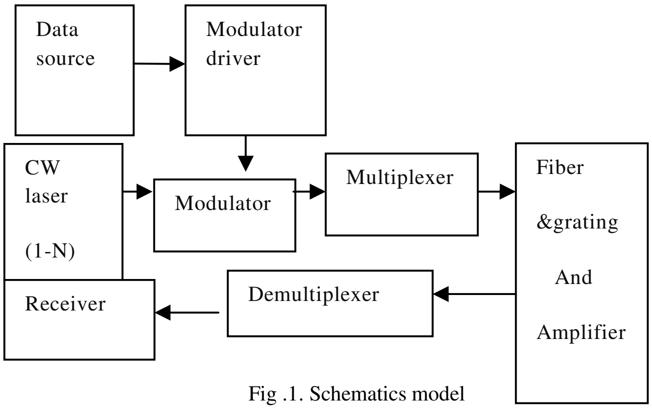


Fig .1. Schematics model

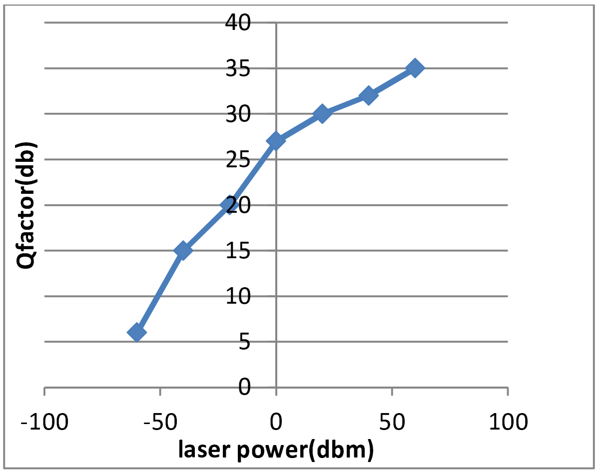


Fig. 2 Qfactor verse laser power

Similarly we observed that FWM power increase with increase laser source power.

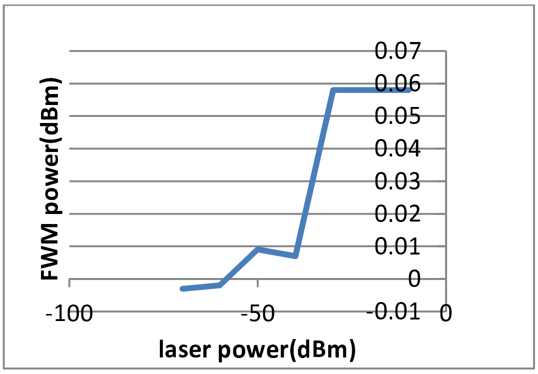


Fig. 3 FWM power verse source power

B. Effect of changing modulator driver

Fig.4 shows the effect of changing modulator driver on four wave mixing at output. Different modulator driver like are NRZ&RZ raised cosine, NRZ&RZ rectangle, RZ supergaussian, RZ soliton. The output optical spectrum for NRZ raised cosine modulator driver is shown in fig 4.(a).The output spectrum of other modulator driver are present as follow

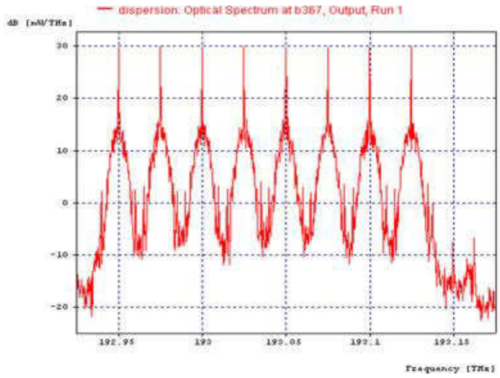


Fig 4 (a) NRZ raised cosine modulator driver

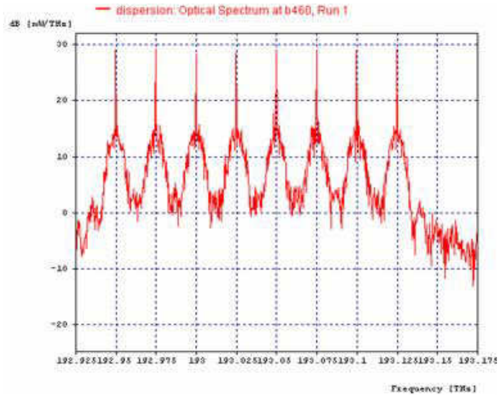


Fig.4 (b) NRZ rectangle modulator driver

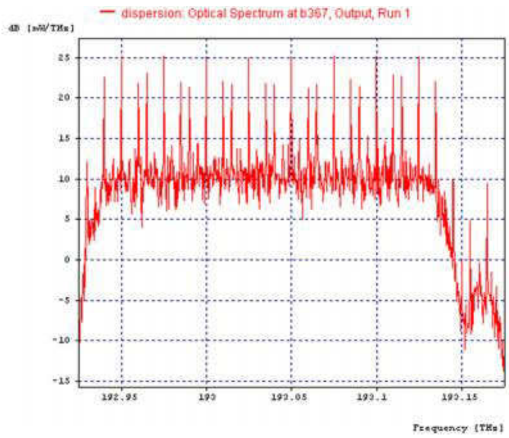


Fig.4. (c) RZ raised cosine modulator driver

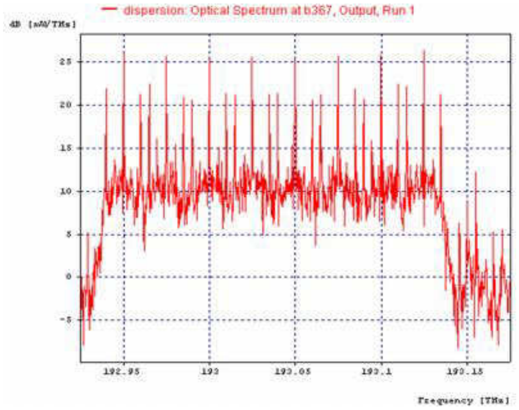


Fig.4 (d) RZ rectangle modulator driver

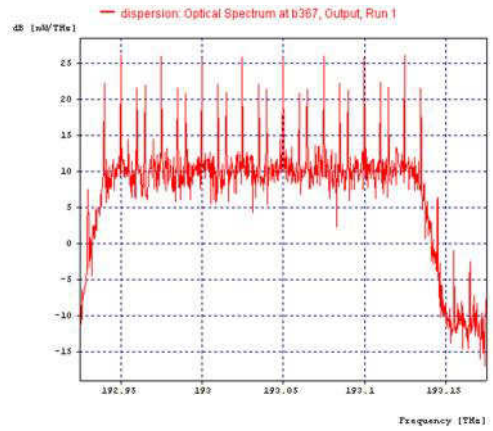


Fig.4. (e) RZ super Gaussian

NRZ raised cosine modulator driver gives the output much clear than other modulator driver. Hence, it clear that four wave mixing is minimum when raised cosine modulator is used.

C. *Effect of chromatics dispersion*

We make to analyse the performance of WDM system based upon chromatic dispersion of fiber.

It obverse that by increase thechromaticsdispersion of fiber power of FWM signal is decreasing, also analysis the new peaks appear inspectrum also decreases. It also obverse that when dispersion is zero than then FWM effect is maximum. Power of new generated component shown in fig .5

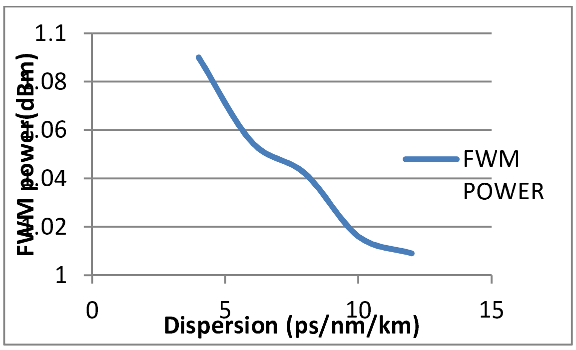


Fig. 5 Power of FWM verse dispersion

D. *Effect of line width*

We make to analyse the performance of WDM system based upon laser source power with different line width of source. It obverse that line width increase from 0 to 20 MHz at different source power, FWM signal power decrease and minimum value of FWM power at 20MHz.Then further increase the line width 20 to 40 MHz power of FWM signal increase. It also obverse that nature of FWM signal change from destructive to constructive as line width changes from 0 to 40 MHz .As shown in fig.6

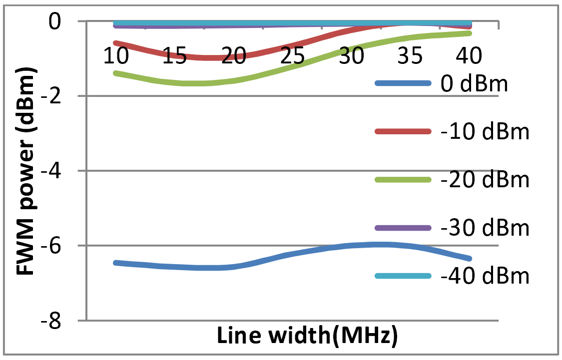


Fig. 6 FWM power vs. line width

We also obverse that optimized value of line width around 20 MHz value of Q Factor is maximum at 0 to -20dBm and Q Factor remains constant at -30 to -50dBm. Similarly bit error rate is minimum at optimized value of line width and more variation above this optimized value of line width.

V. CONCLUSION

In this paper we demonstrated the impact of laser source power, line width at various laser source and dispersion coefficient on FWM in WDM for 3 equally spaced channel. We obverse that as laser source power decrease, FWM signal power and Q factor decrease and similarly demonstrated the impact of line width of various laser source power on FWM in WDM, Then We obverse that as optimized value of line width 20MHz at which FWM signal power is minimize, Q factor maximum and bit error rate are minimize. Similarly chromatics dispersion effect on FWM, power of FWM decreases and new peaks appeared in spectrum also reduces

REFERENCES

[1] G. P. Agrawal,Fiber-Optic communication system, 3rdEd.Wiley, New York, 2002.

[2] G.P.Agrawal, Nonlinear fiber optics,3rd Ed,Academic press,San Diego,CA,2001.

[3] R.W. Tkach, A.R. Chraplyvy,F.Forghieri,A.H.Gnack and R.M.Derosier, “Four-photon mixing and high speed WDM system,” J. Lightwave Technol.,vol13,pp.841-8849,May 1995.

[4] K.O.Hill,D.C.Joshon, B.S.Kawasakiand R.I.MacDonald, “CW three-wave mixing in single mode fiber,”J. Appl. Phys., vol.49, pp. 5098-5106, Oct. 1978.

[5] N. Shibata, R. P. Braun, and R. *G.* Waarts, “Phase mismatch dependence of efficiency of wave generation through four-wave mixing in a single mode optical fiber,” IEEE J. Quantum Electron.vol. 7, pp. 1205-1210,July 1987.

[6] Xuefeng Tang,Zongyan Wu, “Reduction of Intra Channel non linearity using OPC,”Journal of Photonics Technology ,vol 17,no.9 ,September 2005.

[7] Kaewplung and P.Kikuci, “simultaneous cancellation offiber loss, dispersion and Kerr effect by midway OPC and Incorporated with distributed Raman amplifier,” Journal Of Lightwave Technology,vol24,no.1,pp5464, January 2006.

[8] Vladimirpechenkin, jvan j.fair, “On FWM suppression in DM-OFDM system with an optical phase conjugation(OPC) module,” J. Light wave Tech., Vol. 29, No.11, june1, 2011.

[9] Mr. T. Sabapathi, A. Selvamani, “Suppression of FWM by OPC conjugation in DWDM fiber linkInternational journal on recent advancements in electrical,electronics and control engineering.,2011.

[10] S.Sugumaran, Arulmozhivarman, “Effect of chromaticdispersion on FWM in WDM system and its suppression,” Journal of international conference and workshop on emerging Trend in technology 2011

[11] B.K Mishra, Loch ,An. Jolly , KarishmaMhatre, “Analysis of four wave mixing effect in WDM communication system for different channel spacing,” International journal of computer application., Dec.2013

[12] Kamaljit Singh Bhatia, R.S. Kaler, T.S. Kamal and Rajneesh Kaler, “Monitoring and Compensation of Optical Telecommunication Channels by using Optical Add Drop Multiplexers for Optical OFDM System”, 2012 De Gruyter.

[13] E.K. Singh Bhatia, T.S. Kamal, R.S. Kaler, “Peak-to average power ratio reduction using coded signal in optical-orthogonal frequency division multiplexing systems”, IET Optoelectronic., Elsevier Science. (2012), Vol. 6, Iss. 5, pp. 250–254,

[14] Kamaljit Singh Bhatia, R.S. Kaler, T.S. Kamal, “Design and simulation of optical-OFDM systems,” Journal of Russian Laser Research, Springer Science Business Media (2012), Volume 33, Number 5, pp. 202–208, September, 2012

Performance Comparison of WDM-OFDM System

Harmandeep Kaur
Dept. of ECE
Sri Guru Granth Sahib World University, Sirhind
bainsharman94@gmail.com

Kamaljit Singh Bhatia
Dept. of ECE
Sri Guru Granth Sahib World University, Sirhind
kamalbhatia.er@gmail.com

Harsimrat Kaur
Dept. of ECE
CTIEMT, Jalandhar City
simrat6@yahoo.com

***Abstract*—In this paper WDM transmission of 5 OFDM signals with QAM modulation is demonstrated. The results showed an improvement in system performance with an increase in laser power and gives an efficient results at around 70-130 mW. In this paper we combine OFDM and WDM PON's in order to encapsulate the advantages of both to have the maximum benefit.**

***Keywords*—WDM-PON, BER, Q-Factor**

I. INTRODUCTION

In recent the wavelength division multiplexed passive optical networks (WDM-PONs) are considered as a promising candidate to provide broadband access for the next generation networks, since WDM-PONs have a number of excellent features including wide bandwidth etc(3). WDM-PON is based on wavelength division multiplexing (WDM). This kind of PON achieved one point to multi-point communication, which uses the wavelength as identification for each optical network unit (ONU). WDM–PON achieved multiple access, while make full use of optical fiber transmission bandwidth (4).

With the increment of personal communication demand, high-speed passive optical network (PON) techniques have attracted more and more attention; among which, orthogonal frequency division multiplexing (OFDM) PON is considered to be a promising candidate for the next-generation high speed PON for its high spectral efficiency (SE), chromatic dispersion (CD) and polarization mode dispersion (PMD) tolerance, natural compatibility with digital signal processing (DSP)-based implementation and advantages in bandwidth allocation and control [1]. Recently many systems are being studied using OFDM. High-speed optical transmission with per-channel data rates beyond 100 Gb/s is being actively researched for future optical transport systems. It is important to achieve high spectral efficiency (SE) when scaling up the per-channel bit rate in order to increase the overall system capacity. It is also important to achieve a reasonably long transmission distance in order to reduce the overall system cost. Higher-order modulation formats such as 16-QAM, 36-QAM, and 64-QAM were recently used to increase SE [2].

In this paper we combine OFDM and WDM PON's in order to encapsulate the advantages of both to have the maximum benefit.This paper is organized as follows. Section II discusses the system model and experimental setup followed results and discussions in section III. Section IV concludes the paper.

II. EXPERIMENTAL SETUP

In order to evaluate the system BER performance and Q factor, simulations using commercial fiber optics system simulation tool, OPTSIM were carried out based on the experimental setup as shown in fig.2. It has been used by many researchers to evaluate the system performances depending on various aspects. OFDM modulates baseband signals in the signal transmitter. The signals are conversed in frequency, and transmitted to outer modulation, turned into light signals (optical OFDM signals). Then OFDM signals are sent through the transmitters in optical fiber to the receiver part. Direct detection receiver using photodiode detector can receive these optical signals. After OFDM signals are converted into electricity ones. They will be conversed in frequency for OFDM baseband signal demodulation. OFDM signal transmitter and the receiver block diagram is shown in figure 1 and figure 2 respectively. OFDM modulation of optical signal with high frequency spectrum utilization, anti-multipath interference capability is strong, excellent tolerance to the degree of dispersion and polarization mode dispersion, etc. OFDM can be easily combined with a variety of multiple access techniques. Provide access to multiple users at the same time. Therefore it can be combined with WDM - PON named OFDM-WDM-PON (4).

In the transmitter part, signals (bit streams) are converted from serial to parallel. QAM modulator maps these signals and generate the baseband I/Q components of the QAM signal. It uses the IFFT transformation to calculate IFFT on QAM symbol and produces the OFDM signal. RF modulator (QUADMIXIQ) is used to quadrature mix up the OFDM signal from baseband to the carrier frequency. Then again we convert the signal from parallel to serial form and at the OFDM transmitter output we get RF signal. Then the OFDM signal is passed through optical media. In this 5 such signals are passed using WDM technology to further increase the capacity.

In receiver part, the OFDM signal if passed through serial to parallel converter for detection of signal. RF demodulator (QUADMIXIX_DOWN) is used to quadrature mix down the RF modulated OFDM signal to baseband. Two Bessel filters are used to filter out the replica of the signal centered at twice the center frequency. Then the QAM signal is recovered by calculating FFT on OFDM signal. The QAM signal is passed through QAM demodulator to get the original signal in the form of bit stream. After passing through parallel to serial converter we get original streams in serial form. This whole process is performed at transmitter and receiver side on each of the 5 signals of WDM and thus 5 OFDM signals are recovered at the receiver side.

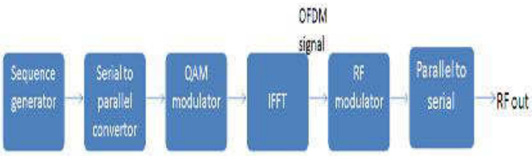


Fig.1. Block diagram of OFDM transmitter

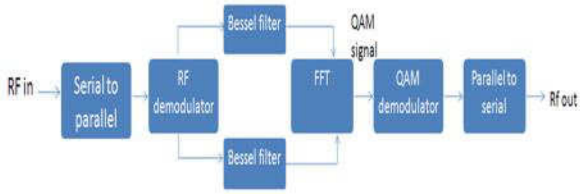


Fig.2.Block diagram of OFDM receiver

In the main schematic, shown in figure 3, we use 5 OFDM signals using OFDM transmitter as compound component and modulate those using CW laser to transmit it through fiber. Using combiner the signals have been multiplexed and transmitted through the fiber. After transmission the signals are demultiplexed using splitter and fed to the different photodiode which converts those to an electrical OFDM signals. RF demodulator compound component is used to recover the bit stream.

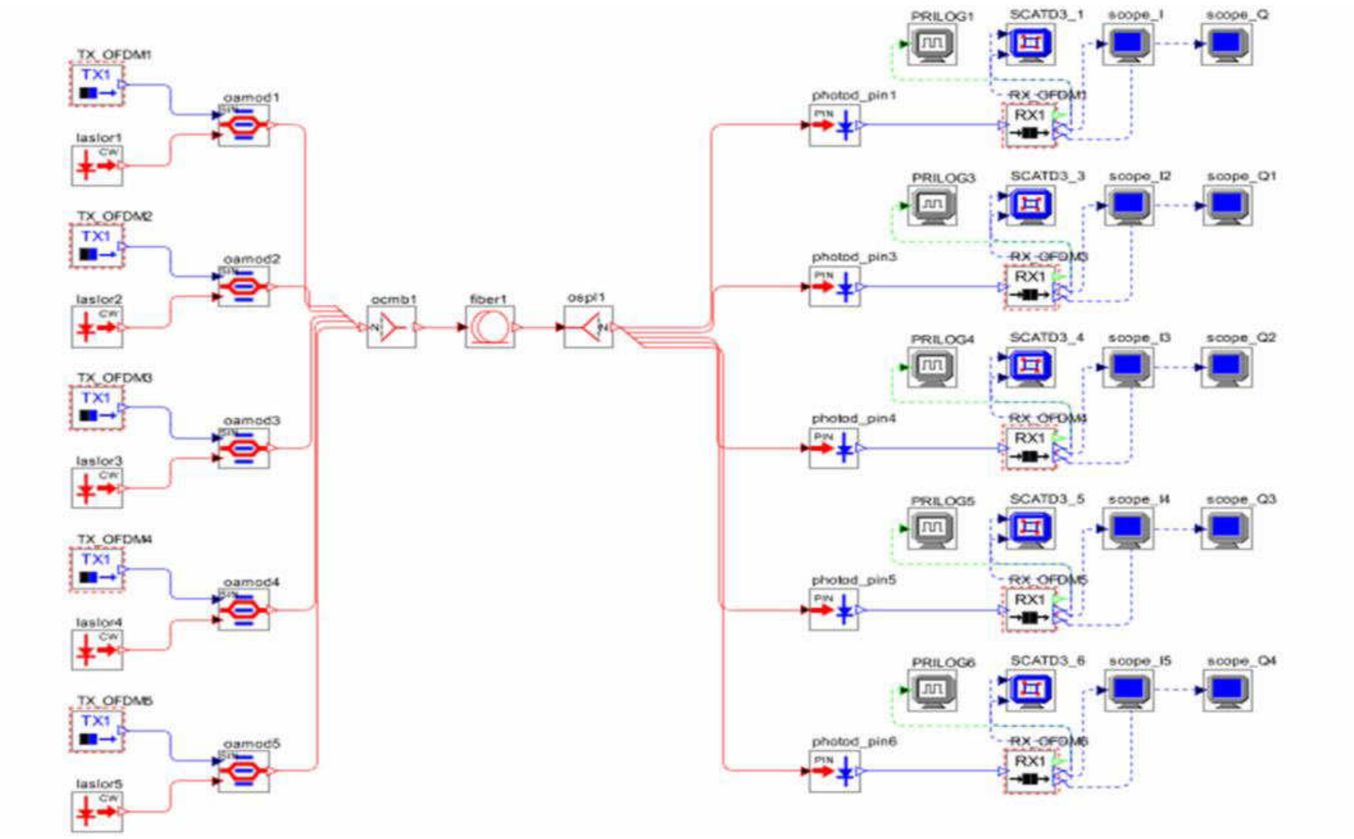


Fig.3.Setup using OPTSIM

III. Experimental Results and Discussion

WDM system with 5 OFDM signals is simulated in this paper to compare the performance of system. BER and Q factors are analyzed with respect to the variation in laser power by viewing the eye diagram on simulation tool. We have observed the results by varying the laser power. With an increase in laser power of all the 5 lasers, an increase in Q-factor and corresponding decrease in BER is seen as observed by eye diagram of each. After a certain limit of laser power the Q-factor starts decreasing and BER increases. The results are efficient at around 70-130 mW of laser power giving 14.48 Q-factor and 9.61*e-006 (0.0023) BER at 100 mW. The eye diagram and scattering diagram under these conditions is shown in figure 5 and figure 6

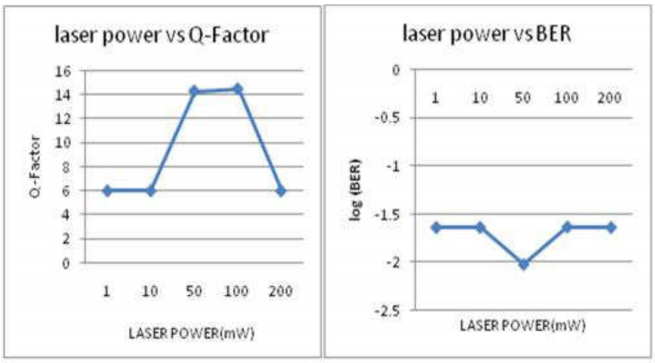


Fig.4. Performance comparison of the system

Table I. Comparison of Results

Laser Power Used	BER	Q-Factor
1 mW	0.02275	6.02
10 mW	0.02275	6.02
50 mW	1.40*e-005	14.31
100 mW	9.61*e-006	14.48
200 Mw	0.02275	6.02

IV. CONCLUSION

Transmission of 5 OFDM signals is performed with WDM technique. The approach of this paper is to compare the BER performance and Q-Factor by increasing laser power in OFDM-WDM system. The BER of the system increases with increase in laser power and then again decreases after certain limit. The results are efficient at around 70-130 mW of laser power giving 14.48 Q-factor and 9.61*e-006 (0.0023) BER at 100 mW.

REFERENCES

[1] Cheng Lei, Minghua Chen, Hongwei Chen, Sigang Yang, Ying Yu Shizhong Xie, "1-Tb/s WDM-OFDM-PON System with Subband Access Scheme and Flexible Subcarrier-Level Bandwidth Allocation", IEEE Photonics, vol.5, no.1, Feb. 2013.

[2] Xiang Liu, S. Chandrasekhar, P. J. Winzer, B. Zhu, D. W. Peckham, S. Draving, J. Evangelista, N. Hoffman, C. J. Youn, Y. H. Kwon, and E. S. Nam, "3× 485-Gb/s WDM Transmission over 4800 km of ULAF and

respectively. Scattering diagram represents the in-phase and quadrature components of these signals. Table-1 compares the results of BER and Q-Factor at varying powers of lasers. Figure 4 shows the performance comparison of the system.

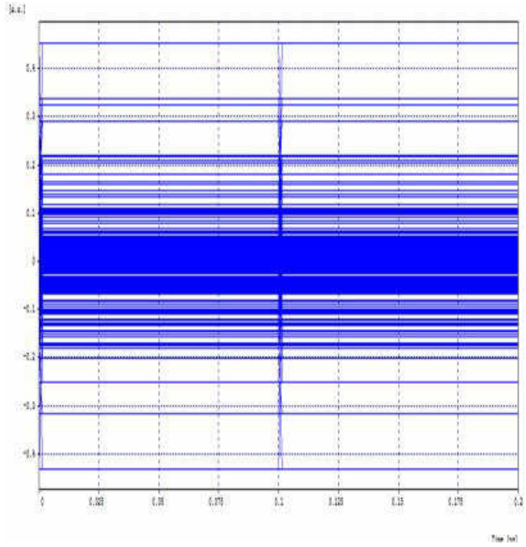


Fig.5. Eye Diagram representation

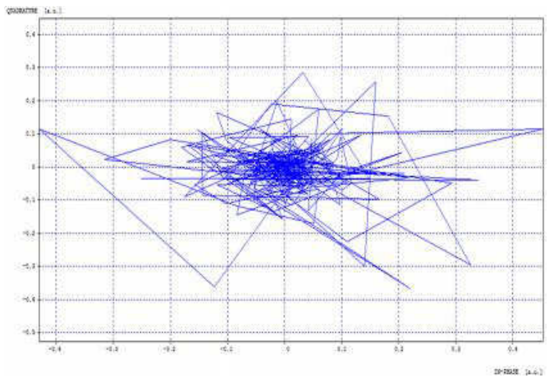


Fig.6. Scattering diagram of the system

12× 100-GHz WSSs Using CO-OFDM and Single Coherent Detection with 80-GS/s ADCs", Optical Society of America, 2011.

[3] Jing Huang, Deming Liu & Cheng Zeng, "10 Gb/s WDM-PON Using Downstream OFDM and Upstream OOK", Canadian Center of Science and Education, Vol. 4, No. 2; March 2011.

[4] Shen Wen, Deng Hui, Wang Wei, Hou Gong, "A Novel Power Communication Access Network with OFDM-WDM-PON", Atlantis Press, International Conference on Advanced Information and Communication Technology for Education, 2013.

[5] Jeong-Min Joo, Moon-Ki Honga, Dung Tien Phama, Chun Ju Youn, Yong-Hwan Kwon, Eun-Soo Nam, and Sang-Kook Hana, "19-Gb/s adaptively modulated optical OFDM transmission for WDMPON using 1 GHz RSOAs with separated I/Q baseband delivery" , IEEE Asia Communications and Photonics, Vol. 8309, 2011.

[6] Gholamreza Baghersalimi, "A Comparative Performance Study of Optical Subsystem Equalization in OFDM-based and OWDM-based Radio-over-Fiber Systems", 11th International Conference on Telecommunications, 2011.

[7] Yufeng Shao, Bo huang, Nan Chi, Junwen Zhang, Wuliang Fang, Bo Liu, Xiangjun Xin, "A Novel Subcarrier OFDM-MSK WDM Passive Optical Network", 2010.

[8] E. Giacoumidis, J.L Wei and J. M. Tang, "Adaptive Modulation Induced WDM Impairment Reduction in Optical OFDM PONs" Future Network & MobileSummit 2010 Conference Proceedings, 2010.

[9] Manpreet Singh, Karamjit Kaur, "Coherent Detection WDM Optical OFDM System", International Journal of Advanced Research in Computer and Communication Engineering, Vol.2, Issue 12, December 2013.

Super-Channels: DWDM Transmission beyond 100 GB/s and Modulation Format

Sarbjit Singh
Dept. of ECE
Chandigarh College of Engineering and Technology
jit_sarb@yahoo.com

Hardeep Singh
Dept. of ECE
Thapar University, Patiala-Punjab, India
hsaujla@yahoo.com

Abstract—**For transferring data to increase performance and implement different analogue and digital techniques are used in fiber optic communication channel. Different digital modulation formats maximizes spectral efficiency and also improves tolerance to transmission impairments. Ever increasing demands of the internet traffic in optical networks has led to the evolution of higher order modulation format for DWDM and the choice of the modulation format greatly affects the performance of the DWM transmission system. In this paper we have discussed the advanced modulation format for DWDM system. For 100 Gbit/s and higher bitrates, the paper provides a performance comparison together with the main characteristics of the modulation formats and indicates appropriate application areas of transport technologies for future networks. This paper also reviews the single carrier modulation format for future technologies at 400GB/s or 1TB/s.**

Keyword – DPSK; RZ-DQPSK, M-QAM, DPSK-3ASK, OSNR, PM-M-QAM, 400GB/s.

I. INTRODUCTION

Conventional on-off-key signals have been extensively employed in optical communication systems. However, OOK modulation is inadequate for transmission of 40 Gbit/s per channel or higher bit rates mainly because of its reduced robustness to fiber nonlinearity, chromatic dispersion and optical filtering at such bit rates. The DWDM transmission at 100 GB/s is known as Super-channel.

The choice of the modulation format greatly affects the performance of the transmission system The most promising modulation formats for future optical networks make use of the phase of the signals to carry information[1]. Among such formats, differential phase-shift-keying and differential quadrature phase-shift-keying are the ones more often referred.The main advantages of DPSK are approximately 3-dB improvement on optical signal-to-noise ratio when compared with conventional OOK, and improved dispersion and polarization mode dispersion tolerance. DQPSK shows also improved spectral efficiency.

With 40 Gbps DWDM [2] now in deployment, carrier and technology developer attention has turned to 100 Gbps and more. So beyond 100 GB/s is the super-channel. In fact super channel is an evolution in DWDM in which several optical carriers are combined to create a composite line side signal of the desired capacity, and which is provisioned in one operational cycle. Some of the most asked questions in the DWDM industry today focus around 100 Gbps: What will be the technology winners? Do capacity growth rates still present a sustainable business case? How quickly will the development and deployment of 100-Gbps DWDM occur? The optical transmission market has been growing dramatically for the last several years, driving up demand for ever higher DWDM transmission rates and system capacity.

The 100 Gb/s system technology using multi-level modulation formats.

In order to relax the high speed bandwidth requirements of the electronic circuits and opto-electronic components multi-level coding like DQPSK is utilized. Multi-level formats coding of several bits in one symbol enable a reduction of the symbol rate of the system on the expense of an increased transmitter and receiver complexity. On the other hand multi-level coding reduces the optical bandwidth consumption of the channel and enables WDM transmission with a narrower DWDM channel spacing.

In this paper, we are focusing on multi-leveling modulation formats which have been applied for field trials together with system suppliers and have been already developed and deployed.

As we know that bit rate-distance product is a figure-of-merit of lightwave systems. To increase the capacity of lightwave systems, or bit rate-distance product, high speed data rate per channel and tighter channel spacing in DWDM systems are the possible solutions. 100 GB/s or more than 100 GB/s DWDM system would be the next generation of lightwave systems. In such high speed DWDM systems, linear and nonlinear impairments become severe. Those linear impairments include chromatic dispersion, and first order polarization mode dispersion; nonlinear impairments include self-phase modulation, cross-phase modulation and four-wave mixing .In fact performance of the DWDM system is governed by the several nonlinear impairments.

Table1:below shows various options of 100GB/s modulation formats which are of research interest and have been already deployed by their main properties in terms of reception (coherent or non-coherent), bits per symbol, symbol rate, and finally the related spectral efficiency. Obviously with the reduction of symbol rate, the modulation format realization becomes more and more complex.

Table I. Modulation Format for 100 GB/s DWDM

Modulation format	OOK	OOK-VSB	RZ-DQPSK	PM-QPSK
Coherent/Non-coherent	Non-coherent	Non-coherent	Non-coherent	Non-coherent
Bits/symbol	1	1	2	2 x 2
Spectral Efficiency	0.5	1	1	2
Symbol Rate(Gbd)	112	112	56	28

II. RZ-DQPSK Format

In most DQPSK systems in classical digital communications the four-symbol signal is generated by superposing two bipolar modulated carrier frequencies that exhibit the same frequency but have a 90° phase shift to each other [3]. It is difficult to adapt this concept to optical communications because the required phase shift of 90° has to be guaranteed until the signals are merged by a 3dB coupler. In a recent optical DQPSK proposal [4] this problem is overcome by O/E converting part of the transmitter signal and feeding it to a servo-control loop. We avoid this problem by a DQPSK transmitter setup shown in figure. The first two Mach-Zehnder modulators (MZM) form a chirp-free binary DPSK signal with RZ pulse shape. The subsequent phase modulator (PM) with a phase shift of either 90° or 0° generates four symbols out of the binary PSK signal. Figure 1 shows the transmission set up for RZ-DQPSK modulation format.

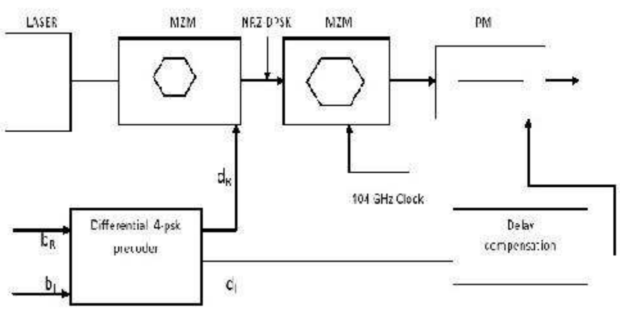


Fig.1. Transmission set up for RZ-DQPSK

A. *Modulation formats for systems beyond 100 Gb/s*

Transmission of optical signals beyond 100 Gb/s by increase of spectral efficiency are currently of high interest at research. The main focus is on multi-level modulation format based on M-QAM and coherent reception applied at single carrier as well as at multi-subcarrier modulation formats. The major target is to maximize their spectral efficiency.

With the deployment of 100 Gbit/s systems and the increasing demand for data transmission capacity, systems with bit rates of 400 Gbit/s and 1 Tbit/s are now being investigated [5]. Possible solutions include single-carrier multi-level modulation with a high symbol rate [2-6] for a bit rate of 400 Gbit/s, and multi-carrier modulation with a low symbol rate for each subcarrier, such as Nyquist wavelength division multiplexing [7] and coherent optical orthogonal frequency division multiplexing , for bit rates of 400 Gbit/s and 1 Tbit/s. Single carrier solutions for a bit rate of 400 Gbit/s are attractive because of the simple structure in the transmitter and a receiver complexity that is similar to those for multi-carrier solutions [8].A per-channel bit rate of 1 Tbit/s can also be achieved using a super channel comprised of several modulated optical subcarriers that are considered as a single entity for network routing [9]. This allows the subcarriers to be more closely spaced together compared to the individual channels in a conventional wavelength division multiplexed system. The number of optical carriers, the bit rate per carrier, and the modulation format are key aspects of the system design[10]. Super channels with a bit rate of 1 Tbit/s have been demonstrated using DP 8-QAM and 9 subcarriers with bit rates of 138 Gbit/s [11], DP 16-QAM and 4 subcarriers with bit rates of 320 Gbit/s, and DP 64-QAM and 10 subcarriers with bit rates of 128.4 Gbit/s . Transmission of a 1.28 Tbit/s dual-carrier DP 16-QAM signals possible over 3200 km of ULAF using hybrid Raman/EDFA amplification.

B. *Single carrier modulation formats*

To achieve bitrates of 400 GB/s or 1TB/s on a single carrier, higher level modulation schemes have to be applied. Recently QAM scheme together with polarization multiplexing is utilized to achieve a channel rate of 200 GB/s with 16 QAM. In an M-QAM or 2m QAM signal, m bits are transmitted in a single time slot or symbol, where m is an integer value. Adding polarization multiplexing to make PM-2m-QAM format [12], 2 m bits are transmitted per symbol. A PM-M-QAM signals can be realized in principle by parallel arrangements of PM-QPSK modulators, where the modulators are driven with binary data signals respectively.

C. *M-QAM realizations and demonstrations*

For realization of 16QAM, a 4-level electrical modulation signal is needed at each electrode. This can either be realized by passive combination of two electrical data signal with different amplitude or by using digital signal processing and D/A conversion (DAC). Polarization multiplexed 16QAM signals have been realized by multilevel generation using passive combination of binary signals to achieve 224 Gb/s channel rate.Using polarization multiplexing and QAM modulation format various high capacity DWDM transmission experiments with high spectral efficiency have been performed. Channel rate of 240 Gb/s is achieved by 8PSK [13] and transmission over 320 km line is possible.Using DP-16QAM transmission, lengths between 670 up to 1500 km are achievable. RF-assisted optical Dual-Carrier 112 Gb/s polarization-multiplexed 16-QAM is applied to achieve 112 Gb/s channel rate. QAM modulation is reported for lower bitrate channels of 100 Gb/s using 32QAM, 100 Gb/s using 35QAM, 112 Gb/s and 120 Gb/s using 64QAM, 56 Gb/s with a spectral efficiency of 11.8 bit/s/Hz using DP-256QAM, 54 Gb/s using DP-512QAM [14].

D. DPSK-3ASK modulation format

By transmitting 2.5 information bits in one symbol, 100-Gb/s serial transport can be realized with components currently available for 40-Gb/s optical transport. One modulation format that meets this requirement is a DPSK-3ASK modulation, where the symbol value is selected out of a set of six values. Two consecutive symbols are combined to transmit five information bits. Information is carried in three different amplitude levels as well as in the phase difference between two symbols.To minimize hardware requirements and create a cost-effective solution, a single Mach-Zehnder modulator can be used to create the optical DPSK-3ASK signal after encoding [15] the 100-Gb/s data stream into a 40-Gb/s.

III. CONCLUSION

At 100 Gb/s line transport the dominant advantages of the DQPSK modulation format together with the coherent receiver have been widely recognized as the best and most cost effective solution for Metro and long haul transport. Thus, this 100 Gb/s modulation format and the transponder realization has been defined by an OIF framework and multi-source agreement. For the next hierarchy of 400 Gb/s line transport there is a high desire to reuse current electronic technologies supporting symbol rates up to 32 Gbaud, but also to be compatible with optical technologies supporting a fixed grid of 50 GHz.

The need of flexible grid requirements have been argued as future 400 Gb/s and 1 Tb/s bitrates will not fit into the fixed ITU-T grid, but this argument has been based on format solutions with challenging symbol rates being twice or four times higher than currently feasible. The currently most promising solution for 400 Gb/s line transport is based on two carriers with 200 Gb/s PM-16QAM modulated with symbol rate of 32 GBaud, supporting a spectral efficiency of 4.

Towards 1 Tb/s line transport, an O-OFDM based solution with multiple optical super-channels with or without additional electrical subcarriers appears promising as single carrier options requires unrealistic high symbol rates that might technologically not feasible within the next 10 years.

REFERENCES

[1] T. Matsuda, A. Naka and S. Saito ,"Comparison between NRZ and RZ signal formats for In-line Amplifier Transmission in the zero-dispersion regime," Journal of Lightwave Technology, Vol. 16, No. 3, March 1998.

[2] "RZ Versus NRZ in Nonlinear WDM systems," F. Forggieri, P.R. Prucnal, R. W. Tkach and A. R. Chraplyvy, IEEE Photonics Technology Letter, Vol. 9, No.7, July 1997.

[3] "Four-photon mixing and high-speed WDM systems," R. W. Tkach, A. R. Chraplyvy, F. Forghieri, A. H. Gnauck and R. M Derosier, Journal of Lightwave Technology, Vol.13, pp.841-849, May 1995.

[4] " Enabling Technologies for 40 Gb/s Long Haul DWDM Transport", B. Mikkelsen, P. Mamyshev, C. Rasmussen, C. Ketchian, NFOEC, 2000.

[5] B.-E. Olsson et al., RF-assisted optical dual-carrier 112Gb/s polarization multiplexed 16-QAM transmitter, in: OFC 2010, paper OMK.

[6] A. Sano , " 240-Gb/s polarization-multiplexed 64-QAM modulation and blind detection using PLC-LN hybrid integrated modulator and digital coherent receiver", in: ECOC 2009.

[7] Y. More , "200-km transmission of 100-Gb/s 32-QAM dual-polarization signals using a digital coherent receiver", in: ECOC 2010.

[8] Govind P. Agrawal, Fiber Optic Communication Systems, Third edition, 2002.

[9] Le Nguyen Binh, Digital Optical Communications, 2009.

[10] Gerd Keiser, Optical Fiber Communications, Third edition, 2000.

[11] R. Sabella, "Tutorial: Key elements for WDM Transport Networks", Photonic Network Communications, 2:1, 7-14, 2000.

[12] Ashwin Gumaste, Tony Antony, DWDM Network Designs and Engineering Solutions, 2003.

[13] Harold Kolimbiris, Fiber Optics Communications, 2004.

[14] Micheal S. Borella, Jason P. Jue, Dhritiman Banerjee, Byrav Ramamurthy, and Biswanath Mukherjee, "Optical Components for WDM Lightwave Networks", Proceedings of the IEEE, Vol. 85, No.8, August 1997.

[15] Gabriel Charlet, "Progress in Optical Modulation Formats for High-Bit Rate WDM Transmissions" IEEE Journal of Selected Topics In Quantum Electronics, Vol. 12, No. 4, July/August 2006.

Different Method of Channels Allocation to Linearity in WDM system

Veerpal Kaur
Deptof ECE
Punjabi University Patiala
Patiala, Punjab
Jalandhar,punjabkaur.veerpal617@gmail.com

Kamaljit Singh Bhatia
Dept. of ECE
S.G.G.S.W.U,
Fatehgarh Sahib
Punjab
kamalbhatia.er@yahoo.com

Amrit Kaur
Dept of ECE
Punjabi university Patiala
Patiala, Punjab
amrit_aks@yahoo.com

Harsimart Kaur
Dept of ECE
CTIEMT, Jalandhar
simrat6@yahoo.com

***Abstract*-In long-haul WDM communication system, transmission characteristics are degraded by four wave mixing. In this paper we analysis the FWM power equal channel spacing ,unequal spacing and equal spacing with polarisation at different value of chromatics dispersion in wave length division multiplexed system. We obverse that unequal spacing have lower FWM power as compared equal spacing and equal spacing with polarisation. Similarly at maximum value of dispersion having minimum value of FWM power.**

Keyword-Four wave mixing,Wavelength-Division Multiplexing, BER, Dispersion.

I. INTRODUCTION

The latest trend in light wave communication system is towards huge information transmission capacity by explosive growth in data rate communication by using wave length division multiplexing. [1-2] The performance of DWDM limited by dispersion and non linearity, When signal power is high. Non-linearity is two type one is stimulated scattering response for loss and gain due to stimulated process, other non linearity are knowkerr effect due to change in refractive with change in input power, and linear response of optical system is lost.[1].It is mainly three type self phase modulation(SPM),cross phase nodulation(CPM) and four wave mixing(FWM). Non linearFWM process in which three frequencies f_i,f_j,f_k (k i,j)interact through third order nonlinearity of optical fiber to generate new frequency[1]

$$f_{ijk} = f_i + f_j - f_k \qquad (1)$$

These new frequency components falls on the same transmission window and original signal get degraded and effect the system performance. Thus three wave give rise nine new frequency component. The number of side bands use to FWM increase is given as [2].

$$M = \frac{N^3 - N^2}{2}. \qquad (2)$$

Where N is number of channels and M number of newly generated side bands. Performance of WDM system degraded by low chromatic dispersion and non-linear effect of fiber link. [2-3] The cause of FWM analyzed in and different methods for mitigating the impact of fiber nonlinearities on WDM/DWDM systems including], dispersion management schemes[14]and the use of nonzero dispersion fibers [19]. Different channel allocation techniques like equal-channel spacing and unequal-channel spacing techniques [17-20], Reduction of FWM investigated in [6] optical phase conjugator as wave front reversal, which is used to suppress the FWM in OFDM system. [6] OPC support all the data type format, OPC simultaneously used to cancel the fiber loss, dispersion loss, Kerr effect in long communication. We have already achieved good results within Monitoring and Compensation of Optical Telecommunication MIMO-OFDM system with OADM [11-17].

In this paper, section II presents theoretical analysis of FWM performance degradation in intensity–modulated direct detection (IM-DD) WDM system with equal channel spacing and with varying chromatic dispersion, section III describe the experiment simulation setup for analyse the effect of equal unequal and polarisation with equal spacing, result and discussion summarized in section IV and conclusion presented in section V.

II. FUNDAMENTAL OF ANALYSIS

FWM is one of major limiting factor in WDM optical communication system due low dispersion, equal channel spacing .As a result, Estimating power of FWM is very important for design and evaluation of WDM system. A well know formula used to estimate the Power of FWM derived by Hill et al[3] and later reformulated by efficiency depends upon phase mismatch factor by Shibata et al.[4].This new formula used for estimating FWM crosstalk is recently used in WDM system.[3].Power of FWM can be written as

$$P_{ijk} = \frac{\eta}{9} D^2 \gamma^2 P_I P_j P_k exp(-\alpha L) \left\{ \frac{[1-exp(-\alpha L)]^2}{\alpha^2} \right\} \qquad (3)$$

Where $P_i P_j P_k$ are signal input power at frequency f_i ,f_j,f_k,(k i,j) respectively, L is the length of fiber, is the fiber attenuation coefficient the degeneracy factor D takes value equal to three or six for degenerate and non degenerate FWM, respectively. The nonlinear coefficient is given [1] by

$$\gamma = 2\pi n_2 / \lambda A_{eff} \qquad (4)$$

Where A_{eff} effect area of fiber , vacuum wave length andn_2 non linear is refractive index. Also η is the FWM efficiency which can be expressed as[4]

$$= \frac{\alpha^2}{^2+\Delta k^2}\left\{1+\frac{4\exp(-\alpha L)\sin^2(\Delta kL/2)}{1-\exp(-\alpha L)^2}\right\} \qquad (5)$$

In this expression k is phase match factor which depends upon dispersion of fiber and channel spacing and written as [4]

$$\Delta k = \frac{2\pi\lambda_k{}^2}{c}\Delta f_{ij}\Delta f_{jk}\left[D_c + \frac{\lambda_k{}^2}{2c}(\Delta f_{ij}+\Delta f_{jk})\frac{dD_c(\lambda_k)}{d\gamma}\right](6)$$

Where $f_{ij} = f_i - f_j$ channel spacing, D_c chromatic dispersion, $dD_c/d\lambda$ dispersion slope, and $_k$ is the wavelength at frequency f_k. Where k mismatch factor not a function of signal powers, so it call linear phase matching factor. FWM efficiency depends on the fiber dispersion, the channel separation, and the fiber length, but not on the transmitted power [4].So as above expression dispersion increase phase match factor increase and efficiency decreases. Power of FWM directly proportional to the efficiency so power of FWM decrease by increasing dispersion.[10]So similarly power of FWM reduces directly by increasing channel spacing.

III. SIMULATION SETUP AND DESCRIPTION

Work is carried out using simulation to observe the effect of FWM in WDM system to evaluate the impact of varying channel spacing between the input channels of a DWDM system in the presence of four wave mixing is shown in Fig. It consisting of eight CW lasers externally modulated by 10 Gbps NRZ data for each channel with equal channel spacing and unequal channel spacing varying in the range of 0.01 THz to 0.015THz launched into 200 km having two spans of 100 km of 0.2 dB/km attenuation factor and core effect area is $67.43 * 10^{12}\, m^2$ .The changes in transmitter side to discuss the various FWM suppression methods under channel allocation are given below

(a)Equal channel spacing-In the case of equal channel spacing the central frequencies of CW laser are taken as 193.025, 193.035, 193.045, 193.055, 193.065,193.075, 193.085 and 193.095 THz having channel spacing of 0.01 THz.

(b) Polarization -In the case of polarization method the polarization for alternate Channel is rotated 180 around the s2 axis of pioncare sphere.

(c) Unequal-channel spacing-In the case of unequal-channel spacing the central frequenciesof laser source are taken as 193.025, 193.04, 193.045, 193.055, 193.07,193.08, 193.085 and 193.095 THz.

Simulation parameters used were as followed in Table1 to obverse the effect of FWM at centre of channel.

TABLE I. SIMULATION PARAMETER AND THEIR VALUES

Parameters	Values
Pump frequency (THz)	193.025to193.095
Probe frequency(THz)	193.060
Channel spacing (THz)	0.010
Pump power(dBm)	-10
Reference bit rate(Gbit/s)	10
Attenuation (dB/km)	.2
Fiber length(km)	200
Amplifier gain (dB)	20
Preamplifier gain (dB)	35

The dispersion is completely compensated by using ideal fiber Bragg grating at each end of fiber span and loss are compensation by fixed amplifier. In order to observe the impact of suppression method the fiber dispersion value is varied from 0 to 16 ps/nm/km during simulation. At receiver end the optical power meter is used to measure of FWM component.

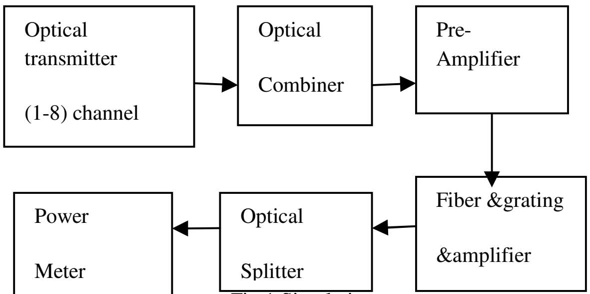


Fig.1.Simulation setup

IV. RESULT AND DISCUSSION

After simulation, it obverse that FWM signal degrade the system performance, analyse in term of FWM signal power, input output spectrum and eye opening. We make to analyse the performance of WDM system based upon different way to allocated the channel at input side and dispersion of fiber we make comparison between equally spaced, equally spaced with orthogonal polarization and repeated unequal channel spacing at varies value of dispersion shown in fig.2

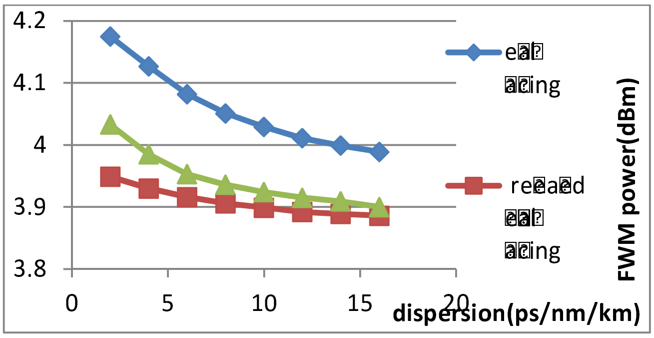


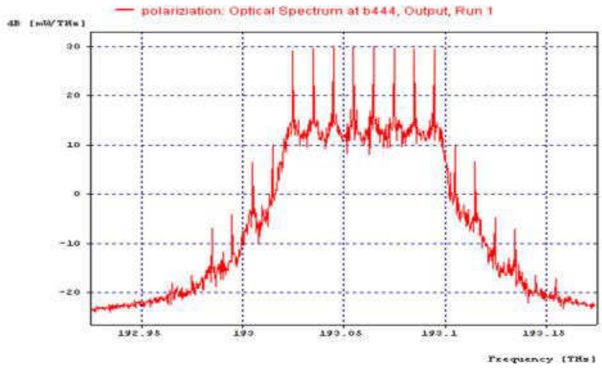
Fig.2. FWM power verses dispersion

We observe that unequal spacing having lower FWM power as compared to other equal spacing and orthogonal polarisation at equal

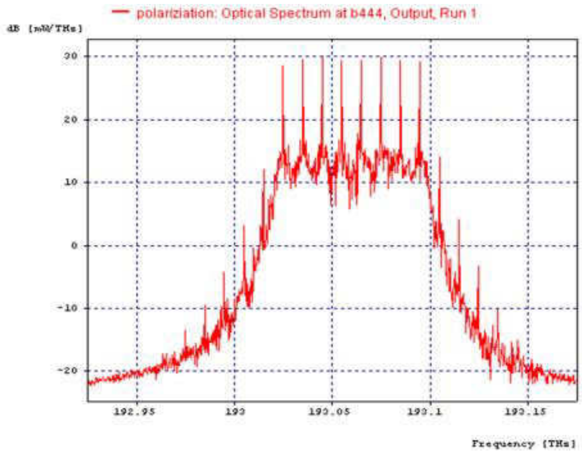
TABLE II. FWM POWER AT RESPECTIVE CHANNEL ALLOCATION

Channel allocation	**FWM power(dBm)**	
	Lower limit(dBm)	***Upper limit(dBm)***
Equal	3.989	4.175
Equal spacing (With polarisation)	3.900	4.050
Unequal spacing	3.886	3.949

Similarly we obverse that as dispersion increase from (0 to 16) power of FWM reduces as due to phase match factor increase,decreasing efficiency which directly reduces FWMPower. As follow the equations in section II. We make analysis output spectrum as shown in fig.3



(a) Output Spectrum (D=4)



(b) Output spetrum (D=6)

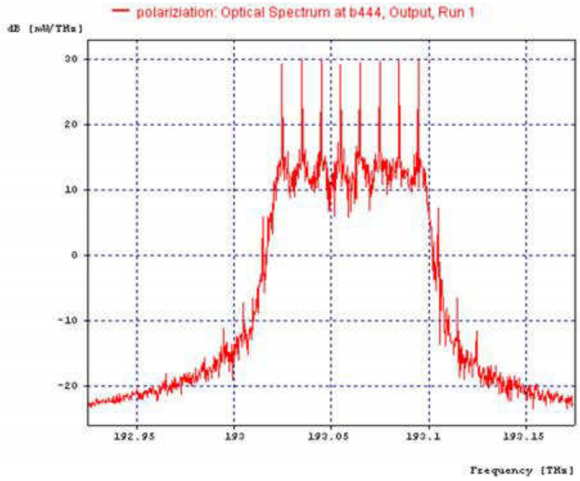
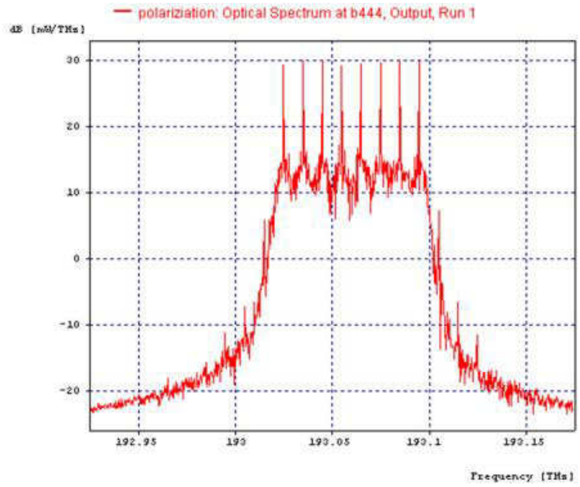


Fig.3. (c) Output spetrum (D=8)



(d) Output Spectrum (D=10)

Fig.3. output spectrum at different value of dispersion.

Above all output spectrum examine that new peaks appeared in spectrum decrease as dispersion increased at equal spacing.

V. CONCLUSIONS

The proposed methods for FWM suppression like equal channel spacing, equal channel spacing with orthogonal polarization and unequal-channel spacing has been found best to the existing methods like dispersion management, using non-zero dispersion fiber. Moreover the existing above FWM suppression method like unequal-channel spacing requires a complex system design and the proposed methods are superior to commonly used methods when the dispersion is fully compensated. Similar increase the dispersion of fiber power of FWM signal is decrease, new peaks appeared in output spectrum decreased.

REFERENCES

[1] G. P. Agrawal, Fiber-Optic communication system, 3rd Ed.Wiley, New York, 2002.

[2] G.P.Agrawal, Nonlinear fiber optics,3rdEd,Academic press, an Diego,CA,2001.

[3] R.W. Tkach, A.R. Chraplyvy,F.Forghieri,A.H.Gnack and R.M.Derosier, “Four-photon mixing and high speed WDM system,” J. Lightwave Technol.,vol13,pp.841-8849,May 1995.

[4] K.O.Hill, D.C.Joshon, B.S.Kawasakiand R.I. MacDonald, “CW three-wave mixing in single mode fiber,”*J. Appl. Phys.,*vol.49, pp. 5098-5106, Oct. 1978.

[5] N. Shibata, R. P. Braun, and R. *G.* Waarts, “Phase-mismatch dependence of efficiency of wave generation through four-wave mixing in a single mode optical fiber,” *IEEE J. Quantum Electron.*vol. 7, pp. 1205-1210,July 1987.

[6] Xuefeng Tang andZongyan Wu, “Reduction of Intra channel non linearity using OPC,” Journal of Photonics Technology, volume17, no.9, September 2005.

[7] Vladimir pechenkin,jvanj.fair, “On FWM suppression in DM-OFDM system with an optical phase conjugation (OPC) module,” J. Light wave Tech., Vol. 29, No.11, june1, 2011.

[8] Mr. T. Sabapathi, A. Selvamani, “Suppression of FWM by OPC conjugation in DWDM fiber link,” International journal on recent advancements in electrical, electronics and control engineering, 2011.

[9] Sugumaran, Arulmozhivarman, “Effect of chromatic dispersion on FWM in WDM system and it suppression,” Journal of international conference and workshop on emerging Trend in technology 2011

[10] B.K Mishra, Loch, An. Jolly , KarishmaMhatre, “Analysis of four wave mixing effect in WDM communication system for different channel spacing,” International journal of computer application., Dec.2013.

[11] Kamaljit Singh Bhatia, R.S. Kaler, T.S. Kamal and Rajneesh Kaler, “Monitoring and Compensation of Optical Telecommunication Channels by using Optical Add Drop Multiplexers for Optical OFDM System”, 2012 De Gruyter.

[12] E.K. Singh Bhatia, T.S. Kamal, R.S. Kaler, “Peak-to-average power ratio reduction using coded signal in optical-orthogonal frequency division multiplexing systems”, IET Optoelectronic., Elsevier Science. (2012), Vol. 6, Iss. 5, pp. 250–254

[13] Kamaljit Singh Bhatia, R.S. Kaler, T.S. Kamal, “Design and simulation of optical-OFDM systems”, Journal of Russian Laser Research, Springer Science+Business Media (2012), Volume 33, Number 5, pp. 202–208, September, 2012,

[14] K.S. Bhatia, T.S. Kamal, R.S. Kaler, “An adaptive compensation scheme-based coded direct detection optical–orthogonal frequency division multiplex (OFDM) system” Computers and Electrical Engineering 38, Elsevier Science (2012) 1573–1578,

[15] Kamaljit Singh Bhatia, R.S. Kaler, T.S. Kamal, RajneeshRandhawa, “Simulative analysis of integrated DWDM and MIMO-OFDM system with OADM” Optik 124, Elsevier Science (2013) 117–121

[16] Kamaljit Singh Bhatia, T.S. Kamal, “Modeling and simulative performance analysis of OADM for hybrid multiplexed Optical-OFDM system” In Press Optik, Elsevie Science (2013),

[17] Amarpal Singh, Ajay K. Sharma, T.S. Kamal, “ Investigation on ModifiedFWMsuppression methods in DWDM optical communication system, Opt. ” Commun.282 (February) (2008) 392–395.

[18] Rajneesh Kaler, R.S. Kaler, “Investigation of four wave mixing effect at differentchannel spacing. ” In press Optik,123Elsevier Science (2012), 352–356.

[19] Neokosmidis, T. Kamalakis, A. Chipouras, T. Sphicopoulos, “ New techniques for the suppression of the four-wave mixing induced distortion in nonzero dispersion fiber WDM systems. ” J. Lightwave Technol. 23 (March) (2005).

[20] Vishal Sharma, RamandeepKaur, “Implementation of DWDM system in the presence of four wave mixing (FWM) under the mpact of channel spacing.”Optik 124 (2013), 3112-3114..IET Optoelectronic, ElsevierScience.

Li-Fi:An Unexampled and Invincible Technology for Indoor Wireless Technology

Kamaljeet Kainth
Dept. of Electronics and Communication
Guru Nanak Dev University, Regional Campus
Jalandhar, India
kainth108@gmail.com

***Abstract*--Light Fidelity (Li-Fi) is a new wireless technology to provide the connectivity with in localized network environment. Whether you're using wireless internet in a home, stealing it from the guy next door, or competing for bandwidth at a conference, you may have face the problem of slow speed as more than one device is tapped into the network. Radio spectrum is getting clogged day by day and demand for wireless data is increasing exponentially every year and capacity is dying up. In order to overcome this problem in future, One germen phycist. Harald Haas has come up with a solution he calls "data through illumination" –taking the fibber out of fiber optic by sending data through an LED light bulb that varies in intensity faster than the human eye can follow[1]. He coined the term "light fidelity" or Li-Fi and set up a private company, Pure VLC, to exploit the technology. He talks about a future where data for laptops, smart phones, and tablets is transmitted through the light in a room. And security would be more – if you can't see the light, you can't access the data.**

Keywords-- LED (Light emitted diode), Wi-Fi, VLC

I. INTRODUCTION

Li-Fi is transmission of data through illumination which send data through a LED light bulb that varies in intensity faster than the human eye can follow. Li-Fi is is the optical version of Wi-Fi. So LiFi is the term some have used to label the fast and cheap wireless-communication system. LED can be switched on and off very quickly, which gives nice opportunities for transmitted data. If the LED is on, you can transmit a digital 1, if it's off you can transmit a 0[1]. This brilliant idea was first showcased by Harald Haas from University of Edinburgh, UK, in his TED Global talk on Visible Light Communication. the other hand Wi-Fi is great for general wireless coverage within buildings, and Li-Fi is ideal for high density wireless data coverage in confined areas and for relieving radio interference issues.

Wireless Fidelity (Wi-Fi) has been in use from almost years to provide the internet services to all the required places right from home to humungous organizations.

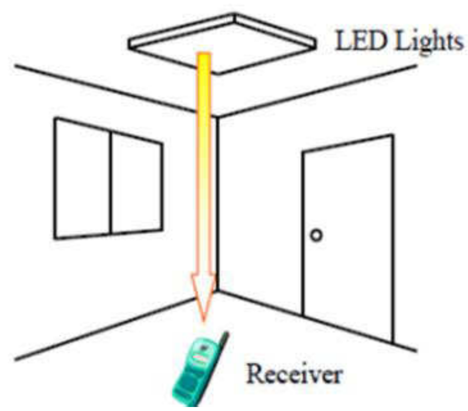


Fig.1.Typical Example of Visible light communication[1]

But this technology has certain drawbacks. As the demand for wireless data is increasing day by day it cause the congestion in radio spectrum. Also there are issues regarding Radio Spectrum like capacity, efficiency, availability and security issues. Lifi can overcome with all these issues of Radio Spectrum. The technology was demonstrated at the 2012 Consumer Electronics Show in Las Vegas using a pair of Casio smart phones to exchange data using light of varying intensity given off from their screens, detectable at a distance of up to ten meters. Li-Fi Consortium (group made by a number of companies and industry groups in October 2011) believes it is possible to achieve more than 10 Gbps, theoretically allowing a high-definition film to be downloaded in 30 seconds [2]

II. WORKING TECHNOLOGY

Operational procedure of Li-fi is very simple, if the LED is on, you transmit a digital 1,if off you transmit a 0.As LED can flicker very quickly and flickering data we want to code. But only fast data rates and depleting bandwidths worldwide are not the only reasons that give this technology an upper hand. Since Li-Fi uses just the light, it can be used safely in aircrafts and hospitals that are prone to interference from radio waves. This can even work underwater where Wi-Fi fails completely, thereby throwing open endless opportunities for military operations. The LIFI product consists of 4 primary sub-assemblies: [3]

- Bulb
- RF power amplifier circuit (PA)
- Printed circuit board (PCB)
- Enclosure

The PCB controls the electrical inputs and outputs of the lamp and houses the microcontroller used to manage different lamp functions. An RF (radio-frequency) signal is generated by the solid-state PA and is guided into an electric field about the bulb. The high concentration of energy in the electric field vaporizes the contents of the bulb to a plasma state at the bulb's center; this controlled plasma generates an intense source of light. All of these subassemblies are contained in an aluminum enclosure.

A. *Visible light communication (VLC)*

Visible light communication (VLC) is a data communications medium using visible light between 400 and 800 THz (780–375 nm). VLC data rate can be increased by parallel data transmission using LED arrays where each LED transmits a different data stream. There are reasons to prefer LED as the light source in VLC while a lot of other illumination devices like fluorescent lamp, incandescent bulb etc. are available. In Fig2, A overhead lamp fitted with an LED with signal-processing technology streams data embedded in its beam at ultra –high speeds to the photo-detector. A receiver dongle then convert the tiny changes in amplitude into and electrical signal, which is then converted basic into a data stream and transmitted to a computer or mobile device.

The importance of using VLC is that Visible light LED can be used almost everywhere. Infrared light is already used for communication such as wireless remote control, Infrared wireless LAN and Infrared inter-building communication. Visible light LED's are beginning to be used in every home and office, which makes visible light LED's ideal for ubiquitous data transmitter. Growth rate of LED lighting is expected to triple from 2009 to 2012 & market share of LED lighting will be more than 30 % of total lighting market in 2016.

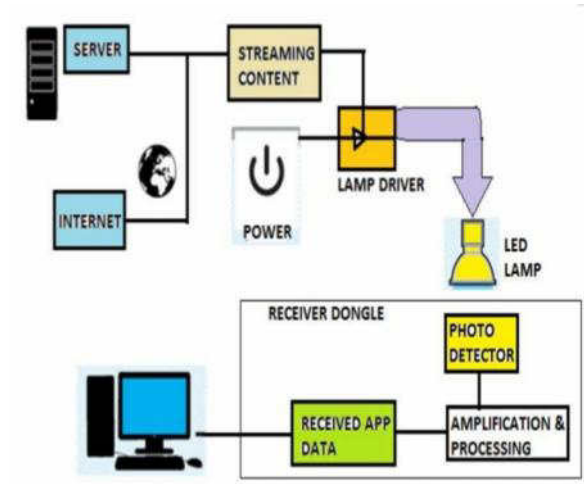


Fig. 2. Data transmission using LED[3]

B. *Different Models for Giga-Speed Technology*

Li-Fi Consortium defined different models for the gigabits speeds. GigaShower and Giga-MIMO for in house communication [3]. There are Transmitter or reciver is mounted into the celling connected to a media server. Giga Shower.

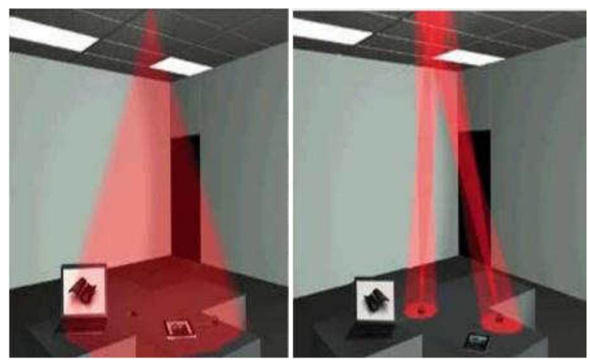


Fig.3.Gigashower and Giga-MIMO Models[7]

provide unidirectional data services via several channel to multiple users with gigabit-class communication. This is like watching TV channels or listening to different radio stations where no uplink channel is needed .GigaSpot and GigaMIMO are optical wireless single-and multiple hotspot solutions offering bidirectional gigabit class communication in a room, hall or shopping malls.

III. COMPARISON BETWEEN LI-FI AND WI-FI

Li-Fi is a terminology which is used to describe visible light communication technology applied to high speed wireless communication. Wi-Fi is great for general wireless coverage within buildings, and Li-Fi is ideal for high density wireless data coverage in confined area and for relieving radio interference issues. It is a complimentary technology that should eventually help free up much needed space within the radio wave spectrum. If the light signal is blocked switch back over to radio waves[4] .VLC/Li-Fi is not in Competition with Wi-Fi. Some parameters are compared here.

Table I. Comparsion between Li-Fi and Wi-Fi [4]

S. No.	Parameters	Wireless Technologies	
		Light Fidelity	***Wi-Fi***
1.	Speed for data transfer	Faster transfer speed (>1 Gbps)	Data Transfer speed (150 Mbps)
2.	Medium	Used Light as a carrier	Used Radio spectrum
3.	Network topology	Point to point	Point to point
4.	Spectrum Range	10,000 time broad spectrum than Radio frequency	Radio frequency spectrum range is less than visible light spectrum.
5.	Operating frequency	Hundreds of Tera Hz	2.4 GHz
6.	Cost	Cheaper than Wi-Fi(free band)	Expensive in comparison to Li-Fi (Radio Spectrum)

If we compare current and future wireless technologies with each then itself Li-Fi has great advantages over all.

Table II. Comparsion between Li-Fi and Wi-Fi [4]

Technology	Speed	Data density Rating
Wireless(current)		
IrDA	4 Mbps	***
Bluetooth	3 Mbps	**
Wi-Fi – IEEE802.11n	150 Mbps	*
Wireless (future)		
WiGig	2 Gbps	*
Giga-IR	1 Gbps	**
Li-Fi	>1Gbps	***

Fig.5. Comparison between current and future wireless technology [4]

IV. FUTURE APPLICATIONS

Visible light communication advantages in hazardous or tough conditions.

- Li-Fi used in sensitive areas such as aircraft for data transmission without causing interference. However light waves can't penetrate walls [5].
- It is used in places where it is difficult to lay optical fibers like operation theaters and at traffic signal. In traffic signals,
- Li-Fi can be used to communicate with the LED lights of cars and might alert drivers when other vehicles are too close so accident numbers can be decreased. Traffic lights could better regulate traffic flow using data.
- It can also be used in hospitals and aero planes where radio signals are prohibited [6].

Thousands and millions of street lamps can be transferred to Li-Fi lamps to transfer data. It can be used in petroleum and chemical plants where other transmission or frequencies could be hazardous [6].

V. LIMITATIONS AND CHALLENGES

Despite of a large no. of advantages of VLC communication over Radio Spectrum, there are certain areas which need special attention for a better future of this technology. Some of them a prescribed here .First of all light can't pass through objects. Also interferences from external light sources like sun light, normal bulbs and opaque materials in the path communication. High installation cost of the VLC System also a major issue. Far Field issues in Led arrays are also an area of concern. Besides of all these a major challenge facing Li-Fi is how the receiving device will transmit back to transmitter [9][10].

VI. CONCLUSION

The concept of Li-Fi is currently attracting a great deal of interest, not least because it may offer a genuine and very efficient alternative to radio-based wireless. The possibilities are numerous and can be explored further. If Li-Fi can be put into practical use, every bulb can be used something like a Wi-Fi hotspot to transmit wireless data and we will proceed toward the cleaner, greener, safer and brighter future. Wi-Fi is great for general wireless coverage within buildings, and Li-Fi is ideal for high density wireless data coverage in confined area and for avoid radio interference issues VLC/Li-Fi is not in Competition with Wi-Fi. It is a complimentary technology that should eventually help free up much needed space within the radio wave spectrum.

REFERENCES

[1] http://www.digplanet.com/wiki/Li-Fi

[2] Visible-light communication: Tripping the light fantastic: A fast and cheap optical version of Wi-Fi is coming', The Economist, dated 28 Jan 2012

[3] Gordon Povey, 'Li-Fi Consortium', dated 19 October 2011

[4] 'Li-Fi: Data through Light', The Institute of Engineers, Technorama Magazine, Volume 62, pp. 41, December 2012.

[5] Seminarprojects.com/s/seminar-report-on-lifi

[6] Harald Haas, 'Wireless data from every light bulb', TED Global, Edinburgh, July 2011.

[7] http://www.lificonsortium.org/

[8] Will Li-Fi be the new Wi-Fi?, New Scientist, by Jamie Condliffe, dated 28 July 2011

[9] http://www.ispreview.co.uk/index.php/2013/01/tiny-led-lights-set-to-deliver-wifi-style-internet-communications.html, January 2013

[10] Harald, B., Nikola, S., Dobroslav, T., Stefan, V., and Harald, H., "VLC: Beyond Point-to-Point Communication," IEEE Communications Magazine pre-print available at: http://www.eng.ed.ac.uk/drupal/hxh/publications/ (2014).

FCON -A Cognitive Approach to 4th Generation Optical Networks

Robin Chadha
ECE Deptt
CT Institute
Jalandhar, Punjab, India
Chadha_robin@yahoo.com

Sansar Chand
ECE Deptt
CT Institute
Jalandhar, Punjab, India
sumisansar@gmail.com

Sunil Kumar
ECE Deptt.
CT Institute
Jalandhar, Punjab,india
sunil_rattu@yahoo.com

Jatinderpal Singh
ECE Deptt.
CT Institute
Jalandhar, Punjab, India
jhsingh11@gmail.com

***Abstract*— To fulfill the ever-swelling traffic demand, Flexible Optical networking is the inspiration for the current expansion of Internet as it marks better use of optical network resources. Recent advancement in this field enables far better quality of service and experience for the end users, empowered through the much higher capacities supported than the earlier technology. However, not all traffic demands necessitate such high bit rates and operators are seeking for networks that are not wasting resources but are cost-effective and therefore versatile this is achieved by employing flexibility in the optical networks. In this paper, we discuss flexible and cognitive optical networks from the perspective of future standardization and benefits of employing cognitive technology in the future optical networks.**

Keywords*— *Optical Networks, Optical Transport, Cognitive Networks, Flexible Optical Networks.

I. INTRODUCTION

As with the progression in technology, we are rapidly approaching the physical capacity limit of conventional optical fiber. So to accommodate the ever-increasing traffic demand.it is essential to make better use of optical network resources. One promising way to achieve this is to introduce elasticity and adaptation into the optical domain through more flexible spectrum allocation, where the required minimum spectral resources are allocated adaptively based on traffic demand and network conditions. Additionally, reconfigurable OADMs (ROADMs) and Optical Cross-Connects (OXCs) were implemented to achieve a higher degree of flexibility and to enable networks to adapt remotely and on-demand to the possible traffic changes, thus reducing the associated operational costs. Moreover, the introduction of high data-rate transmission technology aims to provide large trunks so as to accommodate the bandwidth-intensive new multimedia applications. Nevertheless, not all traffic demands require such high bit rates and operators are seeking for networks that are not wasting resources but are cost-effective and therefore versatile. Operators provide connections with capacity that fulfils the highest (worst case) demand (over- provisioning), while these connections remain underutilised for most of the time. For this purpose, the recent advances in coherent technology, software-defined optics and multicarrier transmission techniques, such as Orthogonal Frequency Division Multiplexing (OFDM) [1]-[2] and Nyquist WDM (N-WDM) [3], have introduced the possibility to achieve a significantly high spectrum-efficiency providing a fractional bandwidth feature. In fact, thanks to these technologies it is possible to dynamically tune the required bit-rate and the optical reachability by appropriately choosing the allocation of the spectrum and the modulation format. These new technologies will enable a new network architecture where any two nodes can be connected with the amount of bandwidth required, either providing a sub-wavelength service or super-channel connectivity [4]-[5]. In this context, to serve a given traffic demand, the network manager has to select the route, the channel, the bit-rate and the modulation format [5]. Hence, traditional Routing and Wavelength Assignment (RWA) algorithms are no longer applicable and it is transformed to a Routing, Modulation Level and Spectrum Allocation (RMLSA) problem where every connection request is assigned a spectrum fraction.

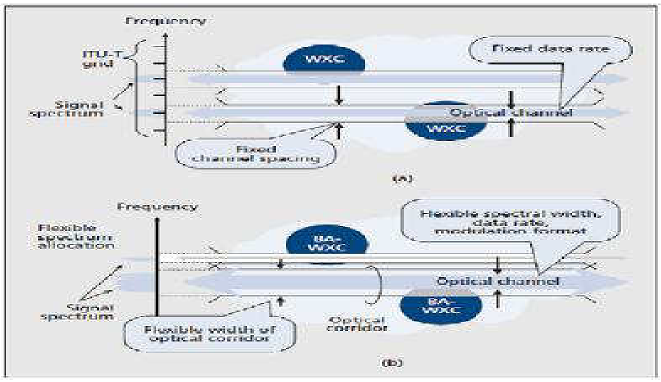


Fig. 1. (a) conventional optical network (b) elastic and adaptive optical network

A promising solution to tackle these challenges comes from exploiting cognition [6]. The use of cognitive techniques in optical networks brings about an extended level of "intelligence" to the optical layer by facilitating the adaptive tuning of various physical layer characteristics (modulation format, forward error correction, wavelength capacity, etc) and network layer parameters (bandwidth, number of simultaneous lightpaths, QoS, etc) depending on application or service requirements.

II. COGNITIVE OPTICAL NETWORKING

A cognitive network is defined as "a network with a process that can perceive current network conditions, and then plan, decide, and act on those conditions. The network can learn from these adaptations and use them to make future decisions [6]. Therefore, a cognitive network should provide better end-to-end performance than a non-cognitive network. In fact, cognition has already been tested and proven to be an excellent solution for wireless networks [7].

There are three main ingredients in such a network:

• Monitoring elements, which provide the network with the perception of the current conditions, and thus enable an aware network.

•Software adaptable elements, which provide the network with the capacity to modify its current configuration, thus enabling an adaptive network.

• Cognitive processes, which learn or make use of past history, so that even when facing two equivalent scenarios, the network (or the entity containing those cognitive processes) may act in a different way if its previous history is different. This third element is the main feature that enables a cognitive network.

Cognitive networks are thus closely related to autonomic networks[8]. An autonomic network relies on self-configuration, self-healing, self-optimization,and self-protection functionalities, so that it may make decisions without manual intervention, i.e., without having to consult with a human administrator [9]. In this way, anautonomic network is not only aware and adaptive, but alsoautomatic. Therefore, a cognitive network can be consideredas a variant of an autonomic network [8], but itemphasizes the self-optimization functionality as well as the use of learning mechanisms, in contrast with other types of autonomic networks, which generally rely on policy-based methods rather than on learning techniques to support the adaptations.

Fig.2 Cognitive loop structure

In the area of optical communications, cognitive techniques are exploited in the framework of CHRON [10] project so to enable "intelligence" in the optical layer. In particular, CHRON should be able to provide effective decisions, by relying on cognition, on:

- how to route new traffic demands, either through existing optical connections (light paths), through new light paths or by triggering a reconfiguration process of the virtual topology (i.e., by rearranging existing connections);

- how to assign resources, not only wavelengths or spectrum, but also the most appropriate transmission/switching technique, modulation format, bit-rate, etc.;

- how to ensure energy-efficient operation;

According to the definition of cognitive networks given above, those decisions must be made by taking into account current status and knowledge acquired through previous experience. Thus, the core element of the CHRON architecture is the cognitive decision system. Such a system is complemented with a network monitoring system, which provides traffic status and optical quality of transmission measurements, and with a set of control and management mechanisms to implement the decisions that are made by the cognitive decision system and to disseminate the monitored information. The interaction of those building elements is detailed in Fig. 3.

Since the cognitive decision system must deal with very diverse tasks, it is composed by five different modules, all of them exploiting cognition. Thus, it includes a RWA/RMLSA module to process optical connection (lightpath) requests; a QoT estimator module to predict the QoT of the optical connections before being established (and thus helping the RWA/RMLSA module to ensure that quality requirements are met); a virtual topology design module, which determines the optimal set of lightpaths that should be established on the network to deal with a given traffic demand, and a traffic grooming module, which is in charge of routing traffic through the lightpaths composing the virtual topology. Last but not least, a network planner and decision maker module coordinates and triggers the operation of the other modules and handles the communications with other network elements.

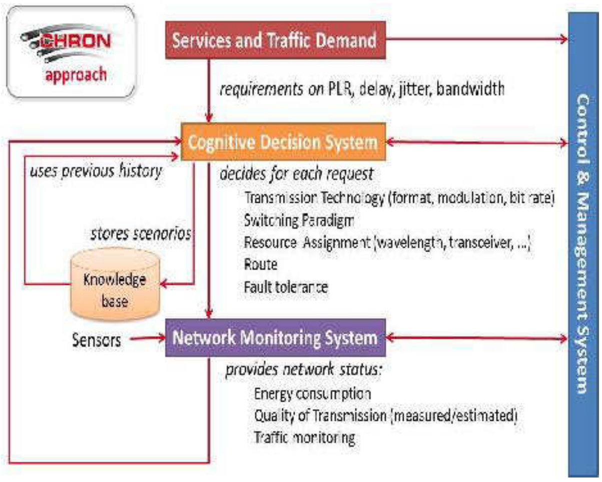


Fig. 3 Main elements of the CHRON approach

In the framework of this architecture, the advantages of cognition have already been demonstrated in a number of scenarios, such as on quickly and effectively assessing whether an optical connection (i.e., a lightpath) satisfies QoT requirements [11], or on determining which set of connections should be established on an optical network (i.e., the so-called virtual topology) in order to support the traffic load while satisfying QoT requirements and minimizing energy consumption and congestion [12].

In the former scenario, the utilization of Case-Based Reasoning techniques to exploit knowledge acquired through previous experiences leads to obtaining not only a high percentage of successful classification of lightpaths into high or low QoT categories (Fig. 4(a)), but also to a great reduction in the computing time (around three orders of magnitude) when compared to a previous tool for QoT assessment which does not employ cognition [11].

In the latter scenario, the inclusion of cognition in a multi-objective algorithm to determine the optimal set of virtual topologies with different trade-offs in terms of throughput and energy consumption brings great advantages. Since a multi objective algorithm provides a set of solutions (i.e., virtual topologies) in a single execution, we have joined the solutions provided by two versions of the same algorithm: one without cognition and the other with cognition. Then, the best set of solutions has been selected, which is called the common Pareto Optimal Set (POS). Fig. 4(b) shows that at the beginning (when there is no previous history that the cognitive method can exploit), both methods contribute approximately with the same number of solutions. However, once cognition really enters into play, i.e., when enough past history is used, most of the solutions contained in the common POS (i.e., the best solutions) are obtained by the cognitive method [12].

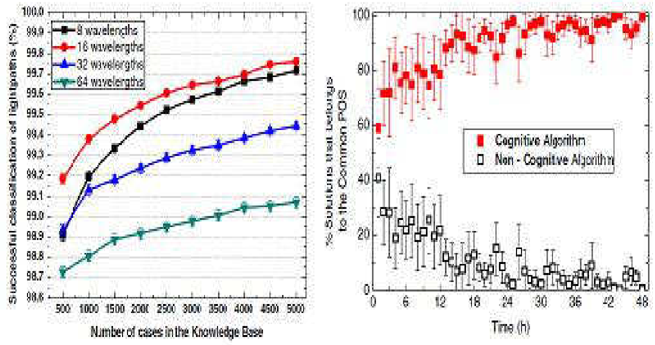


Fig 4 (a) successful classification of light paths into high/low categories (b) percentage of solutions with and without cognition

III. Advantages of cognition based Flexible Networks

Cognition is a useful tool capable of optimizing the design and control of an optical network. A cognitive network allows the introduction of a flexible transport to support the Future Internet, by pushing down to the optical layer some of the “intelligence” typically performed in the IP layer.

A. *Spectrum Allocation Advantages*

This study includes fixed WDM SLR networks that deliver either 40 Gb/s, 100 Gb/s or 400 Gb/s per channel and MLR [13] networks with data rates of 10 Gb/s, 40 Gb/s, 100 Gb/s and 400 Gb/s. Regarding the flex-grid solutions, two multi-carrier solutions have been considered; one refers to the technique reported in [1] (denoted as E-OFDM) while the other refers to the technique in [2] (denoted as O-OFDM). Both multi-carrier solutions can adapt the transmitted bit-rate from 10Gb/s-400Gb/s by modulating subcarriers with the necessary modulation level.

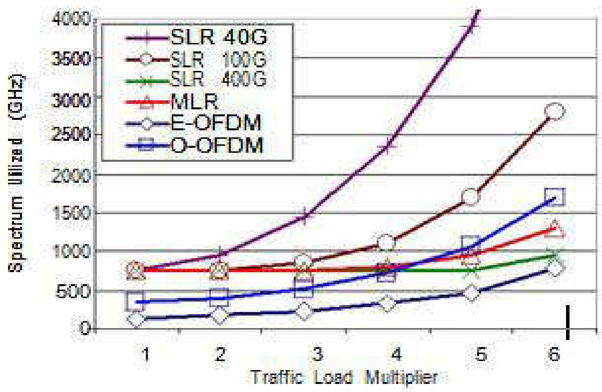


Fig. 5 Spectrum utilization for all solutions and different traffic loads

To calculate the bandwidth utilized by the various solutions the Deutsche Telekom core network (14 nodes, 23 bidirectional links) and the realistic traffic matrix of the DT network for 2010 scaled up to 11 times to obtain traffic ranging from 3.6 Tb/s up to 39.6 Tb/s has been utilized. Under the given assumptions, the flexible multi-carrier solutions offer the most efficient spectrum allocation as expected from the optimized packing of the connections in the frequency domain (Fig. 5).

B. *Energy Efficiency Advantages*

In addition to the capital cost of the future core network, power consumption is another parameter that becomes relevant in network planning, mainly due to the operational economic implications but also the growing ecological awareness, considering the pace at which traffic is increasing annually. Following the resource allocation of all solutions, the energy efficiency is estimated considering the power consumption needs of the associated networking elements. Hence, the considered solutions were compared with respect to the power consumption of the associated network elements, i.e., transponders, optical cross-connects (OXCs) and optical line amplifiers.

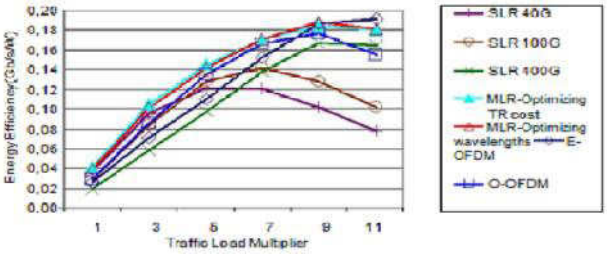


Fig. 6 Energy Efficiency under different traffic loads

The estimated energy efficiency (in Gb/s/W) for the various traffic loads is illustrated in Fig. 6. 400G SLR appears to be the least efficient for traffic load up to 5 although it tends to improve for higher loads. The other SLR solutions achieve better efficiency that decreases for high loads justified by the great number of transponders as depicted in Fig. 6. On the other hand, the granularity of 10G/40G/100G/400G in MLR and of the low-rate subcarriers in O-OFDM appears to be sufficient for the entire range of traffic loads optimizing the number and type of transponders and leading to low power consumption. Under the given power consumption assumptions, E-OFDM demonstrates lower energy efficiency for load up to 5. Moving up in traffic load, the transponders assumed run at higher bit rates leading to superior energy efficiency.

On the whole, in terms of the overall network energy efficiency, flex-grid solutions achieve low energy per bit as they use just the amount of network resources needed for given input traffic.

IV. CONCLUSION

We have provided an overview of cognitive optical networks.By means of network monitors, the network becomes aware of current conditions and thus can adapt itself in order to optimize network performance with the help of software-adaptable elements. However, these networks also rely on cognitive processes, which make it possible to learn from the past and thus get an advantage from knowledge acquired through experience for further improvements. There are many different alternatives for the implementation of cognition, and we have briefly described a number of architectures, mainly focusing on the CHRON approach. We have also analyzed their enabling techniques in terms of monitoring elements, software adaptable elements, and control and management plane solutions, taking into account the current trend toward the use of more flexible and heterogeneous optical technologies. Finally, we have shown how cognition can help in diverse optical networking tasks, such as assessing the QoT of optical connections, designing optimized virtual topologies in reconfigurable environments, or helping identify the incoming signal format at a receiver, thus facilitating the autonomous modification of the modulation format.

REFERENCES

[1] Klekamp, A., et al.: Transmission Reach of Optical-OFDM Superchannels with 10-600 Gb/s for Transparent Bit-Rate Adaptive Networks. In: Proceedings of ECOC 2011.

[2] Chandrasekhar, S., et al.: Transmission of a 1.2-Tb/s 24-Carrier No-Guard-Interval Coherent OFDM Superchannel over 7200-km of Ultra-Large-Area Fiber. In: Proceedings of ECOC (2009); paper PD2.6

[3] Gavioli, G., et al.: Investigation of the Impact of Ultra-Narrow Carrier Spacing on the Transmission of a 10-Carrier 1Tb/s Superchannel. In: Proceedings of OFC (2010); paper OThD3

[4] Borkowski, R., et al.: Experimental Demonstration of Mixed Formats and Bit Rates Signal Allocation for Spectrum-flexible Optical Networking. In: Proc. Optical Fibre Communications Conference, Los Angeles CA, USA (March 2012); paper OW3A.7

[5] Jinno, M., et al.: Spectrum-efficient and scalable elastic optical path network: architecture, benefits, and enabling technologies. IEEE Communications Magazine 47, 66–73 (2009)

[6] Thomas, R.W., Friend, D.H., DaSilva, L.A., MacKenzie, A.B.: Cognitive networks: adaptation and learning to achieve end-to-end performance objectives. IEEE Communications Magazine 44, 51–57 (2006)

[7] Mahmoud, Q.H. (ed.): Cognitive Networks: Towards Self-Aware Networks. John Wiley & Sons, Ltd. (2007)

[8] Z. Movahevic, M. Ayari, R. Langar, and G. Pujolle, “A survey of autonomic network architectures and evaluation criteria,”IEEE Commun. Surv. Tutorials, vol. 14, no. 2, pp. 464–490,2012.

[9] J. Strassner, “The role of autonomic networking in cognitive networks,” in Cognitive Networks: Towards Self-Aware Networks, Q. H. Mahmoud, Ed. Wiley, 2007, pp. 23–52.

[10] CHRON project, http://www.ict-chron.eu

[11] Jiménez, T., et al.: A Cognitive System for Fast Quality of Transmission Estimation in Core Optical Network. In: Proc. OFC/NFOEC (2012); paper OW3A.5.

[12] Fernández, N., et al.: Cognition to Design Energetically Efficient and Impairment Aware Virtual Topologies for Optical Networks. In: 16th International Conference on Optical Networking Design and Modeling, ONDM 2012. University of Essex, Colchester (in press, 2012).

[13] Agraz, F., et al.: Experimental Demonstration of Centralized and Distributed Impairment-Aware Control Plane Schemes for Dynamic Transparent Optical Networks. In: OFC/NFOEC (2010); paper PDPD5.

[14] R. W. Thomas, D. H. Friend, L. A. DaSilva, and A. B. MacKenzie, "Cognitive networks: Adaptation and learning to achieve end-to-end performance objectives," IEEE Commun. Mag., vol. 44, pp. 51–57, Dec. 2006.

[15] Nicola sambu, et al:"towards high rate and flexible optical networks" IEEE Communications Magazine • May 2012 .

Optimization of Point to Point Optical Backhaul Network

Sunil Kumar
Dept. of ECE
CT Institutions
Jalandhar, Punjab, India
sunil_rattu@yahoo.com

Harpreet Dhami
Dept. of ECE
GES Polytechnic College
Hoshiarpur, Punjab,India
dhamihsp@gmail.com

Jatinder Pal Singh
Dept. of ECE
CT Institutions
Jalandhar, Punjab, India
jhsingh11@gmail.com

Sansar Chand
Dept. of ECE
CT Institutions
Jalandhar, Punjab, India
sumisansar@gmail.com

***Abstract*—Wavelength Divison Multiplexing (WDM) increases the carrying capacity of the physical medium. WDM assigns incoming optical signals to specific frequencies of light (wavelengths, or lambdas) within a certain frequency band. Dense Wavelength Division Multiplexing (DWDM)[2] has recently gained great popularity as it provides a cost effective way to increase the transmission capacity of existing fiber cable installations. Recently the performance of DWDM[9] components and frequency-stabilized lasers has substantially improved, while the costs have come down significantly. This makes a variety of new optical network architectures economically viable. Depending on application needs, different types of DWDM systems are deployed; the list includes point-to-point, tree and ring network, etc. Each of these systems needs different components. In this research work, a multichannel high speed DWDM point to point network has been designed with for 16 channels with channel spacing of 0.4nm and data rate of 10Gbps. Performance optimization of the modeled network using different modulation format has been obtained on modeled network.**

Keywords- Wavelength Divison Multiplexing (WDM), Dense Wavelength Division Multiplexing (DWDM) network, OADM.

I. INTRODUCTION

Wavelength Division Multiplexing (WDM) increases the carrying capacity of the physical fiber by assigning incoming optical signals to specific frequencies of light (wavelengths, or lambdas) within a certain frequency band. Like TDM, the resulting capacity is an aggregate of the input signals, but WDM carries each input signal independently of the others. This means that each channel has its own dedicated bandwidth; all signals arrive at the same time, rather than being broken up and carried in time slots .The difference between WDM and Dense Wavelength Division Multiplexing (DWDM) is fundamentally one of only degree. DWDM spaces the wavelengths more closely than does WDM, and therefore has a greater overall capacity. Recently, the channel bandwidth of commercial DWDM system has reached to OC-192(STM-64) and the total bandwidth of an optical network exceeds 20 Tbps . The term "dense" WDM or DWDM, was once used to signify the use of wavelengths with a channel spacing of 50 GHz in a single mode fiber system (SMF).

Optical Add/Drop Multiplexer (OADM)[1] is an important network element in DWDM-OADM system that can be used to reduce the number of conversions between the electrical and optical domains and frame processing time, thus decreasing the delay of the network. An OADM takes a multi-wavelength signal arriving in an input fiber, drops one or more preselected wavelengths from the signal, and adds one or more pre-selected wavelengths into the multi-wavelength signal that exits in an output fiber. An OADM may be well thought-out to be a certain type of optical cross-connect. The main factor in OADM is crosstalk which arises due to component imperfections and limits the performance of the system.

II. SIMULATION SET UP

The designed optical point to point (P2P) network architecture is based on a single unidirectional fiber topology having data rates of 10 Gbps. It consists of nodes as shown in figure 1 connected by non linear single mode fiber. Each node is converting the electrical data into the optical signal and transmitted the optical link . Each node is also equipped with tunable transmitter operating in multiband environment and compound receiver with multiple filters; each receiver takes care of a particular data channel which owns a unique specific wavelength.

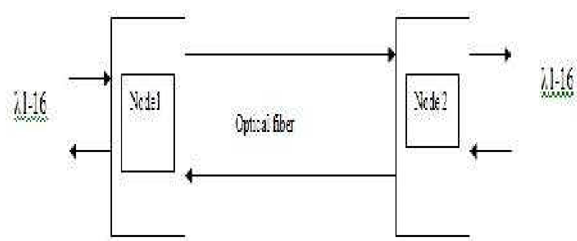


Fig 1. 16 Channel DWDM network

Each node has the ability to add/drop any wavelength of each data channel. EDFA (erbium doped fiber amplifier)[4][8] after each fiber span is inserted to compensate the fiber attenuation. LEAF fiber i.e. least effective area is used to compensate the nonlinearities of multichannel system.

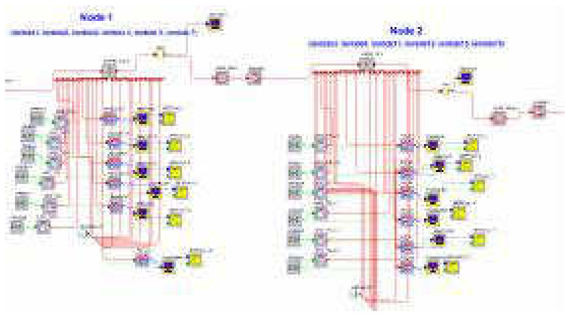


Fig 2. Simulation set up

The power per channel of -9 dBm was used at transmitters. We used 16 wavelengths at 50 GHz spacing ranging from 1550 to 1556 nm wavelength. After each node multiplot is used to observe optical performance matrix.

The simulation setup for DWDM network has been shown in figure 2. Each node has one input and one output for line signal, and also 16 inputs/outputs for added/dropped wavelengths (in- and out-clients interfaces) as shown in figure 4.3. The random data block generate the pseudo random bits which are change into electrical signals and then converted into 1550 nm optical signal to be transmitted output from these ports are connected to optical receiver blocks and then to MultiPlot visualizer which provide plots for electrical signal spectrum, and eye diagram Each node is consisting of demultiplexer and multiplexer which work in multiband environment. The optical multiplexer and demultiplexer are used to combines the add signal from the current node and to drop signal from current node which arrived from the other neighboring downstream node. The 16 wavelengths used in the architecture are given in table 1.

The node block consists of one 1x16 Demultiplexer, 16x1 Multiplexer, and 16 optical switches . The input from transport interface (Input1) is demultiplexed into 16 wavelengths and each of them goes to a switch with corresponding input from client interface (Input2-Input17).

TABLE I. 16 Wavelengths' used in architecture

Lambda(n)	Wavelength(nm)
1	1550
2	1550.4
3	1550.8
4	1551.2
5	1551.6
6	1552
7	1552.4
8	1552.8
9	1553.2
10	1553.6
11	1554
12	1554.4
13	1554.8
14	1555.2
15	1555.6
16	1556

The switch can be in either bar or cross state (is set by the switching array value). One output from the switch goes back to clients interface out ports (Output2-Output17) and the other output is being multiplexed with 15 other outputs and then sent to transport interface output (Output1). The OADM CC block can specify following parameters: crosstalk level [-30 dB], switching configuration [1,1,1,1,1,0,0,0,1,1,0,0,1,1,1,0], first channel wavelength [1.55E-6] and channel spacing [4.0E-10], optical filter bandwidth [3.2E-10] used in demultiplexing.In switch configuration 1 represent add/drop and 0 represent pass through

III. RESULTS AND DISCUSSIONS

The investigations have been carried out at a bit rate of 10Gb/s in an DWDM OADM ring designed to yield a performance analysis of data formats viz. NRZ, RZ, CRZ and DPSK [3][14][10] using BER,Q²dB and Eye diagram. Figures 3,4 & 5 shows the simulation set for comparing modulation formats.

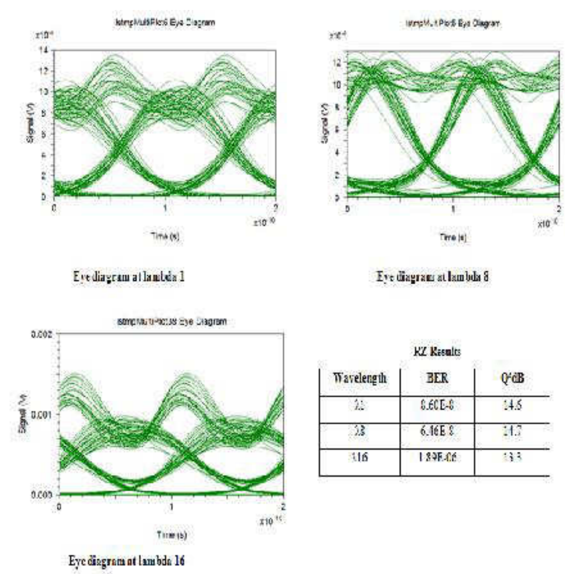
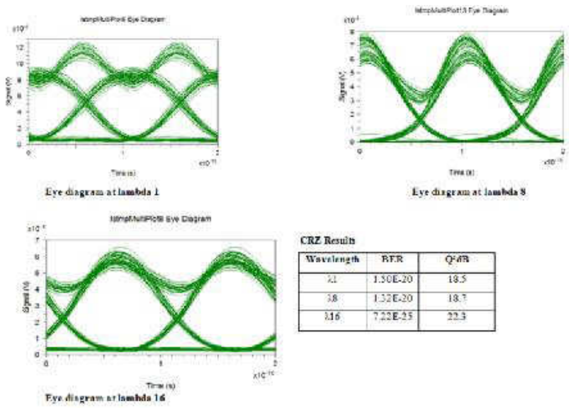


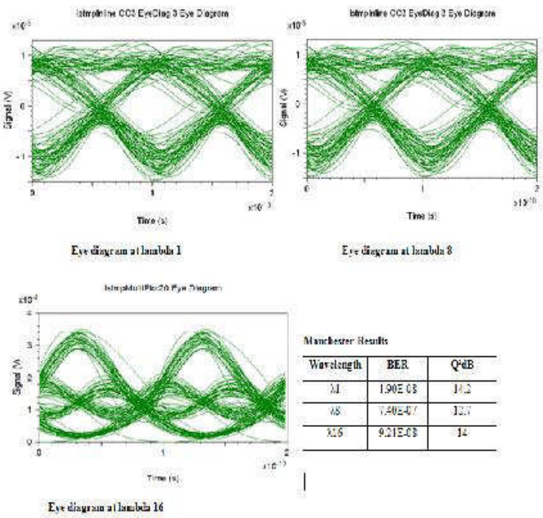
Fig 3.RZ modulation format



CRZ Results

Wavelength	BER	Q-dB
λ1	1.50E-20	18.5
λ8	1.32E-20	18.7
λ16	7.22E-25	22.3

Fig 4.CRZ modulation format



Manchester Results

Wavelength	BER	Q-dB
λ1	1.90E-08	14.2
λ8	[illegible]	12.7
λ16	9.21E-08	14

Fig 5.Manchester modulation format

IV. CONCLUSION

It is observed from performance that chirped return to zero (CRZ) modulation format has yielded better performance as compared to RZ and Manchester formats.BER of the order of e-25 and Q as high as 22.3 has been reported for CRZ. Hence it is concluded that CRZ is the optimal modulation format for the modeled multichannel, high speed DWDM network.

It is observed that EDC has yielded better performance. BER of the order of e-33 and Q as high as 22.2 has been reported for EDC with FFE six and DFE one. Hence it is concluded that EDC with FFE six and DFE one is the optimal combination to achieve better performance for the modeled multichannel, high speed DWDM .

The surviving signal power levels experience power excursions when some channels are dropped or added. This due to the fact that when channels are added or dropped by network's reconfiguration or failure, the power of the surviving channels decreases or increases due to cross saturation in the amplifiers. Power excursion of surviving channels can cause signal distortion by nonlinear effects or degradation of optical signal to noise ratio (OSNR).

REFERENCES

[1] Fujiwara, Masahiko "Optical Cross connect/ADM Based Networks", IEEE Communications Magazine, March 2004, pp. 111-112.

[2] Rosenkranz, Werner "Approaches for Bandwidth-Efficient Modulation Formats in DWDM- Systems", http://www.nit.eu/icton-2001,July 2001.

[3] Jaworski, Marek "Optical Modulation Formats for High-speed DWDM Systems", http://www.nit.eu/icton-2003,July 2003.

[4] Freitas, Marcio and Almeida,R. de "Transient control for cascaded EDFAs by using a multi-objective optimization approach", http://www.mendeley.com/research/transient-control- cascaded-edfas-using-multiobjective- optimization-approach, April 2004.

[5] Gorinevsky, Dimitry and Farbe, Gennady "System Analysis of Power Transients in Advanced WDM Networks", IEEE/OSA Journal of Light wave Technology, Vol. 22, No. 10, 2004, pp 2245- 2255.

[6] Bülow, Henning and Klekamp, Axel "Electronic Dispersion Compensation", IEEE/OSA Journal of Light wave Technology, Vol. 26, January 1, 2008.

[7] Yeh, J.R., Selker, M., Trail, J. D. and Levi, I. "DWDM Architectures for Video on Demand Transport and Distribution", 1998 IEEE LEOS Summer Topical Meetings Digest (1999), pp5.

[8] Richards, D. H., Jackel, J. L. and Ali, M. A. "Multichannel EDFA Chain Control: A Comparison of Two All-Optical Approaches", IEEE photonics technology letters, vol. 10, NO.1, January 1999.

[9] Stoll, Detlefand and Richter,Alexander "Metropolitan DWDM:A Dynamically Configurable Ring for the KomNet Field Trial in Berlin", IEEE Communications Magazine, February 2001.

[10] BOSCOR,G. and Poggiolini, P.' "An exact analysis of RZ versus NRZ sensitivity in a noise limited optical systems", Proc. 27th Eur. Conf. on Opt. Comm. ECOC'OI– Amsterdam (2002), pp-526-527.

[11] Bae, Soo-Jin and Hee Lee,Chang " Transient Control of EDFA using Recirculating loop for WDM Transmission System", CLEO'2000, May 2005, CWK5.

[12] O'Sullivan, M.S.and Roberts, K "Electronic Dispersion Compensation Techniques for Optical Communication Systems", ECOC 2005 Proceedings - Vol.2, 2005.

[13] Sen, Qian Yu and Shanbhag, Abhijit' "Electronic Data Processing for Error and Dispersion Compensation, Journal of light wave technology", Vol. 24, NO. 1DECEMBER 2006.

[14] Downie, John D. and Sauer, Michael "Behavior of MLSE-EDC with Self-Phase Modulation Limitations and Various Dispersion Levels in 10.7-Gb/s NRZ and Duobinary Signals", IEEE photonics technology letters, vol. 19, NO. 13, JULY 1, 2007.

Comparative Study of DWDM Topologies and Networks

Vikrant Sharma
ECE Department, CTIT
Jalandhar, India
Vikrant.sharma@yahoo.com

Amandeep Singh Sehmby
ECE Department, CTIEMT
Jalandhar, India
singh_aman099@yahoo.com

Narinder Pal Singh
ECE Department,CTIEMT
Jalandhar, India
er.narinderpal@gmail.com

Rajesh Sharma
CA Department, CTIEMT
Jalandhar, India
rajeshsharma1234@gmail.com

***Abstract*—This paper provides an overview of the topologies and networks of Dense Wavelength Division Multiplexing (DWDM) and its comparison. The network topologies and the recent development of two major DWDM-based networks, namely the backbone network and the residential access network, are also examined.. the paper consist of the detailed analysis of various DWDM links like with TDM and without TDM and it also classifies the back bone DWDM network. This paper looks into the future of broadband integrated service networks based on the DWDM technology.**

Keywords*—DWDM (Dense Wavelength Division Multiplexing), MONET (Metropolitan Optical Network), OADM(Optical Add-Drop Multiplexer).

I. INTRODUCTION

Dense wavelength division multiplexing (DWDM) networks are categorized into four main topological configurations: DWDM point-to-point with or without add-drop multiplexing network, fully connected mesh network, star network, and DWDM ring network with OADM nodes and a hub. Depending upon the application each topology can exhibit its own requirements and optical components.

Apart from these there are some hybrid topologies which can be the combination of above mentioned topologies. The Metropolitan Optical Network project (MONET) is a biggest example in which WDM network is developed for and funded by a number of private companies and by U.S. government agencies. It consists of two sub networks, one located in New Jersey and one in the Washington, D.C./Maryland area; the two are interconnected with a long-distance point-to-point optical link.

II. TOPOLOGIES

A. Point-to-Point Topology

Point-to-point topology is a topology that requires ultrahigh speed (10-40 Gb/s), ultrahigh aggregate bandwidth (in the order of several terabits per second), high signal integrity, great reliability, and fast path restoration capability and is used for long haul transport. To use this topology the distance between transmitter and receiver may be several hundred kilometers, and the number of amplifiers between the two end points is typically less than 10 (as determined by power loss and signal distortion).To drop and add the channels the add drop multiplexing techniques can be used in this topology. Number of channels, channel spacing, type of fiber, signal modulation method, and component type selection are all important parameters in the calculation of the power budget.

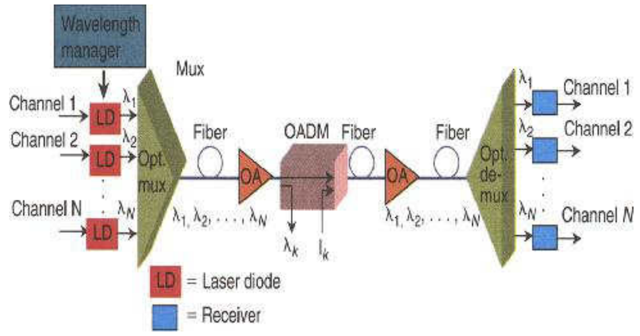


Fig. 1. A DWDM point-to-point with add-drop multiplexing enables the system to drop and add channels along its path.

In DWDM, each channel is carried over a specified wavelength (hi) also known as "optical channel." Different channels may carry different data (e.g., voice, data, video, data packets) at different bit rates. The transmitter-receiver optical link contains several optical components like optical amplifiers, fiber(s), optical filters, OADM, couplers, laser sources, and modulators and receivers. Each component exhibits its own signal-affecting characteristic, as described in Part II. In Figure 1 an end-to-end simplistic view of a DWDM point-to-point system that includes lasers, an optical multiplexer and de multiplexer, fibers, optical amplifiers (OA), and an optical add-drop multiplexer is shown.

B. Ring-Configured Mesh and Star Networks

In general, a DWDM ring network consists of a fiber in a ring configuration that fully interconnects nodes; some systems have two fiber rings for network protection. Such a ring may cover a local or a metropolitan area and span a few tens of kilometers. The fiber ring may contain few (4) to many wavelength channels, and few to many nodes. The bit rate per wavelength channel may be 622 Mb/s or lower, or 1.25 Gb/s or higher. One of the nodes on the ring is a hub station where all wavelengths are sourced, terminated, and managed; connectivity with other networks takes place at this hub station. Each node and the hub have optical add-drop multiplexers (OADM) to drop off and add one or more designated wavelength channels.

In DWDM ring networks, the hub station can also act as a source and can also terminate several types of traffic. It also manages all the channels between various nodes. At an OADM, one (or more) optical frequency is dropped off and added, whereas the remaining frequencies pass through transparently. As the number of OADMs is increased the signal becomes more prone to losses.

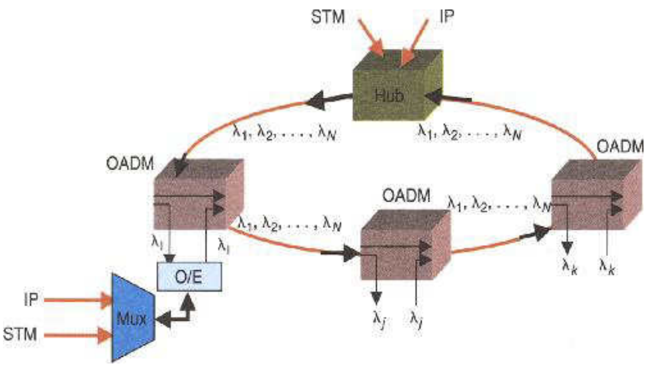


Fig. 2. A DWDM ring network; the hub station sources and terminates payloads of several types.

The number of wavelengths in the fiber is typically more than the number of nodes. Figure 2 depicts a basic configuration but does not address network survivability or ring fault avoidance.

To accomplish the fully connected network of nodes, the wavelength assignment is also managed. The hub may also provide connectivity with other networks. In addition, an OADM node may be connected with a multiplexer/ de multiplexer where several data sources are multiplexed. A simple ring topology with a hub and two nodes, A and B, linked via wavelength as is shown in Figure 3, where node A also multiplexes several data sources. All data sources are terminated by the corresponding OADM node (node B), however, since they are on the same channel (and the same wavelength).

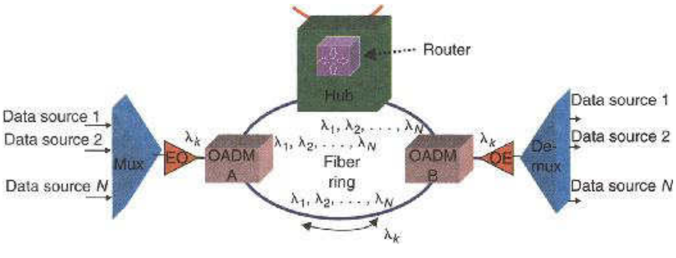


Fig. 3. In a DWDM ring topology, channel (wavelength) assignment may be managed by the hub station .

III. DWDM BACKBONE NETWORKS

The network structures of DWDM-based backbone networks can be classified into three classes:

• Simple point-point DWDM link,

• DWDM wavelength routing with electronic TDM (time domain multiplexing) and switching/routing backbone network, and

• All-optical DWDM network.

This section discusses the network architecture and the mechanisms of these three types of DWDM backbone networks.

A. Point-To-Point DWDM links

The point to point link backbone network is the simplest application of DWDM technology. Figure 4 shows the architecture of the networks using 4 network switching/routing nodes as an example. In this architecture, the electronic nodes can be SONET/SDH switches, Internet routers, ATM switches, or any other type network Wavelength multiplexer / de multiplexer (light wave grating devices) and a pair of optical-electrical/electrical-optical converters constitutes the components of DWDM node. Each wavelength channel is used to transmit one stream of data individually. All of the light wave channels are multiplexed into one light beam and pumps it into one single fiber. The combined light of multiple wavelengths is separated by the de multiplexer at the receiving end. The signals carried by each wavelength channel are then converted back to the electrical domain through the O/E convertors (photo detectors). One wavelength channel is generally equivalent to a traditional fiber in which one light beam is used to carry information in this way. The dotted-lines in Figure 4 represent the wavelength channels. It is worth noting that the wave length channels in one fiber can be used for both directions or two fibers are used with each for one direction.

The increase in the bandwidth by creating multiple channels with low costs is the biggest advantage of DWDM links. The limitation of this approach, however, is that the bandwidth of each wavelength channel may not be fully utilized due to the speed of the electrical devices, which is referred to as the well-known electro-optic bottleneck. Also, the use of the wavelength channels may not be optimal due to the fact that the meshes formed by the wavelength channel are all identical, which can be seen in Figure 4.

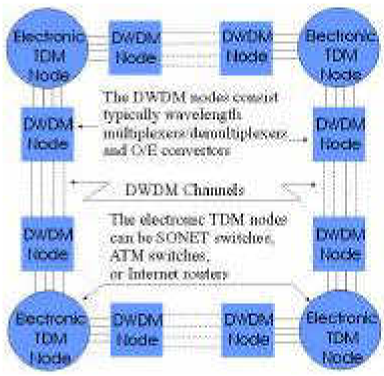


Fig. 4.DWDM point-to-point link backbone network

B. Wavelength Routing with Electronic TDM

Figure 5. represents the second type of DWDM application in backbone networks, in which to configure or reconfigure the network topology within the optical domain the wavelength routers are used and the TDM (Time Domain Multiplexing) network nodes can be used to perform multiplexing and switching in the electrical domain. This combined optical and electrical network architecture can be applied in SONET/SDH in which the electrical TDM network nodes would be SONET switches, or in the Internet in which the electrical TDM network nodes would be the Internet routers. The architecture can also be used in an ATM network where the electrical TDM network nodes would be ATM switches.

The optimization of DWDM wavelength channels by reconfiguring the mesh formed by the wavelength channels is the biggest advantage of this combined architecture. The topology of reconfiguration can be dynamic in which network topology is reset periodically according to the traffic with the time period in the order of seconds or milliseconds. The reconfiguration can also be static in which the mesh is set for a longer period of time. The wavelength router in the optical domain is the enabling technology. Different types of wavelength routers are available commercially which range from mechanically controlled to thermally controlled and to semiconductor wavelength switches. The ability to utilize the bandwidth capacity to the level that the electronics can handle is the biggest advantage of this technology, because of the reconfiguration of the mesh by the wavelength routers but the problem of electro-optic bottleneck still prevails. But still the problem is reduced in comparison with the point-to-point DWDM links. Several technical issues are required to be dealt with for this architecture, which include the control system for the mesh reconfiguration, the traffic evaluation among the DWDM channels and among the fibers, as well as the technology for constructing the wavelength routers. These technical issues have been the topics of research for some time and all are still undergoing development at the current time.

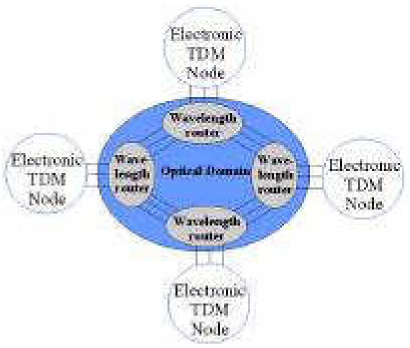


Fig. 5.Wavelength routing with electronic TDM DWDM networks

C. All-Optical DWDM Networks

To eliminate the conversions between electricity and light is the main goal of all the DWDM networks. The all-optical network is also referred to as the transparent network. Two types of all-optical DWDM backbone networks can be proposed, which are:

• Wavelength switching DWDM networks without TDM, and

• DWDM with optical domain TDM.

These two types of DWDM backbone networks are discussed in the following two subsections.

1). Wavelength switching without TDM

Use of wavelength switches (also called wavelength routers) can also allow the use of wavelength switching .Architecture of the network, in which wavelength switches are used to establish connections between the two communicators is shown in figure 6. This is very much the same as the old PSTN (Public Switched Telephone Network) system where the crossbar type of electrical switches was used to establish the circuit for the two users. The wavelength switches for this type of networks can switch among the wavelength channels of multiple fiber input and output ports. A wavelength router contains the additional capacity of changing the wavelength of the signal between routers resulting in high utilization of wavelength channels. In the case of a wavelength router without wavelength conversion, the two users involved in the communication are connected by one signal wavelength across all of switches in the light path. However, with the wavelength conversions, the two sides of the communication can be connected by different wavelengths in different fiber links between switches. Figure 7 shows an example of a wavelength router structure with multiple fiber inputs and outputs. The simplicity in the switching mechanism is the advantage of DWDM networks. However the under utilization of the bandwidth capacity of the wavelength channels is the major problem of these kind of networks. This is due to the reason that once the light path is established between the two sides of the communication parties, it is up to the user to use the available bandwidth of the wavelength channel which can be tens of GHz or even hundreds of GHz. It is obviously inefficient for voice communications. Hence, this type of all-optical DWDM backbone networks will unlikely be used to replace the SONET/SDH. But, it can be useful for Enterprise intranets where different types of communications can share the same network with low costs of building it. Also, because of the transparency of the network, it provides many advantages, such low error rate and low maintenance costs.

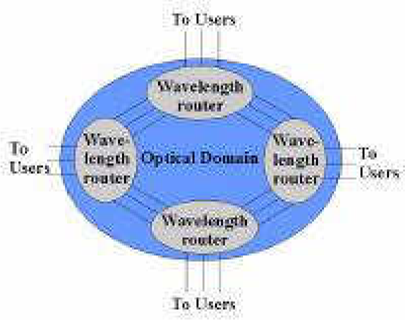


Fig. 6.Wavelength switching without TDM

2). Wavelength switching with optical TDM

The wavelength routing all-optical network has the problem of low efficiency in utilizing the bandwidth of wavelength channels with each having the capacity of hundreds of Gigabits per second. Although the combination wavelength routing and electronic time domain multiplexing may increase the bandwidth utilization to some extent, it brings out the O/E conversions that may limits the speed and due to which packet delays are caused. Therefore, it is natural to implement optical TDM in future optical networks, which eliminates the O/E conversions resulting in a transparent high-speed all-optical network. Replacing the electrical TDM nodes in the DWDM with electrical TDM architecture (Figure 5) by optical TDM nodes, we obtain a DWDM wavelength routing with optical TDM architecture, as shown in Figure5. This indicates that the all-optical TDM nodes in the all-optical architecture can be optical SONET/SDH switches, or all-optical ATM switches, or all-optical internet routers. Different types of all-optical TDM/switch nodes can also be in one network, provided the protocol conversions are implemented.

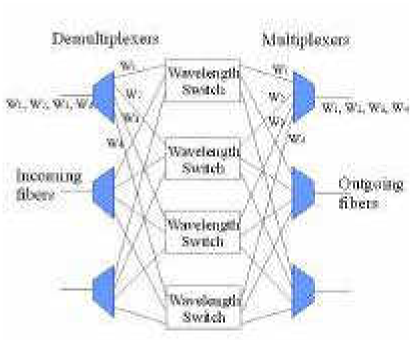


Fig. 7.A Wavelength Router

In fact, the combination of optical TDM/switch node and the wavelength router in one routing site (Figure 8) into one all-optical switching node that not only forwards packets through time domain multiplexing but also selects the light path intelligently according to the availability and traffic loads of the links. The only limitation and the question so far is how can we build such all-optical TDM/ switch nodes. This is the problem that is still unsolved. To build the three major types of the current electrical switches, namely the SONET/SDH switch, the ATM switch, and the Internet router, the SONET/SDH switch is the simplest one among the three since it is for circuit switching. The most complex one among the three is the internet router since there is requirement of a digital optical processor and all-optical memories. However, digital optical logic devices are generally required to build these switches. Hence, it is clear that there is always a need of optical digital processing power before the implementation of any of these TDM switches, unless other mechanism may emerge to revolutionize the existing concept of packet switching networks.

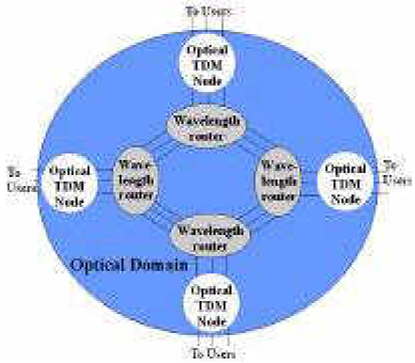


Fig. 8.All-optical TDM/switch with wavelength router

IV. CONCLUSION

This paper provided the summary of the DWDM networks and topologies. The DWDM point-to-point topology played a vital role in the backbone networks and it will continue to be installed for existing and new fiber links. However, for the all-optical DWDM network to become viable, we may have to wait till the optical processing power becomes available. This may create a time gap between the DWDM point-to-point applications and the all-optical DWDM transparent networks. The cost barrier has been weakened through replacing the fiber by DWDM channels in the active access network architecture. The TDM devices in the system may still be too high at the present time. On the other hand, the fiber cost (along with the costs of the passive devices) in the passive architecture is probably still not cheap enough to make it ahead of the active architecture. However, the very high channel-count DWDM may change the landscape of the access network world, and it may become even cheaper than the combined costs of twisted pair copper and coaxial cables.

REFERENCES

[1] S.V. Kartalopoulos, Understanding SONETISDH and ATM: Communications Networks for the Next Millennium, IEEE Press, Piscataway, NJ, 1999.

[2] B. Furht, Handbook ofInternet and Multimedia : Systems and Applications, IEEE Press, Piscataway, NJ, 1999.

[3] L.G. Raman, Fundamentals ofTelecommunications Network Management, IEEE Press, Piscataway, NJ, 1999.

[4] J.e. Palais, Fiber Optic Communications, 3rd ed., Prentice-Hall, Englewood Cliffs, NJ, 1992.

[5] A. Borella, G. Cancellieri, and F. Chiaraluce, Wavelength Division Multiple Access Optical Networks, Artec House, Boston, 1998.

[6] B.T. Doshi, S. Dravida, P. Harshavardhana, O. Hauser, and Y. Wang, "Optical Network Design and Restoration," Bell Labs Technical Journal, vol. 4, no. 1, 1999, pp. 58-84.

[7] L.Y. Lin, E. Karasan, and R.W. Tkach, "Layered Switch Architectures for High-Capacity Optical Transport Networks," IEEE Journal on Selected Areas in Communications, vol. 16, no. 7, September 1998, pp. 1074-1080.

[8] M. Koga et aI., "Large-Capacity Optical Path Cross-Connect System for WDM Photonic Transport Network," IEEE Journal on SelectedAreas in Communications, vol. 16, no. 7 September 1998, pp. 1260-1269.

[9] D. Banerjee, 1. Frank, and B. Mukherjee, "Passive Optical Network Architecture Based on Waveguide Grating Routers," IEEE Journal on Selected Areas in Communications vol. 16, no. 7, September 1998, pp. 1040-1050.

[10] S. Chatterjee and S. Pawlowski, "All-Optical Networks," Communications of the ACM, vol. 47, no. 6, June 1999, pp. 74-83.

[11] S.V. Kartalopoulos, "A Manhattan Fiber Distributed Data Interface Architecture, Globecom'90, San Diego, CA, December 2-5, 1990.

[12] S.V. Kartalopoulos, "Disaster Avoidance in the Manhattan Fiber Distributed Data Interface Network," Globecom'93, Houston, TX, December 2, 1993.

[13] S.R. Johnson and V.L. Nichols, "Advanced Optical Networking-LUCENT's MONET Network Elements," Bell Labs Technical Journal, vol. 4, no. 1, 1999, pp. 145-162.

[14] L.D. Garrett et aI., "The MONET New Jersey Network Demonstration," IEEE Journal on Selected Areas in Communications, vol. 16, no. 7, September 1998, pp. 1199-1219.

[15] Y. Chen,· M.T. Fatehi, H.J. LaRoche, J.Z. Larsen, and B.L. Nelson, "Metro Optical Networking," Bell Labs Technical Journal, vol. 4, no. 1, 1999, pp. 163-186.

[16] A.R. Chraplyvy, "High-Capacity Lightwave Transmission Experiments," Bell Labs Technical Journal, vol. 4, no. 1, 1999, pp. 230-245.

[17] D.B. Buchholz et aI., "Broadband Fiber Access: A Fiber-to-the-Customer Access Architecture," Bell Labs Technical Journal, vol. 4, no. 1, 1999, pp. 282-299.

[18] G.C. Wilson et aI., "FiberVista: An FTTH or FTTC System Delivering Broadband Data and CATV Services," Bell Labs Technical Journal, vol. 4, no. 1, 1999, pp. 300-322.

[19] A. Jourdan et al., "Key Building Blocks for High-Capacity WDM Photonic Transport Networks," IEEE Journal on Selected Areas in Communications, vol. 16, no. 7, September 1998, pp. 1286-1297.

[20] M. Berger et aI., "Pan-European Optical Networking Using Wavelength Division Multiplexing," IEEE Communications Magazine, vol. 35, no. 4, 1997, pp. 82-88.

[21] B. Fabianek, K. Fitchew, S. Myken, and A. Houghton, "Optical Network Research and Development in European Community Programs: From RACE to ACTS," IEEE Journal on Selected Areas in Communications, vol. 16, no. 7, September 1998, pp. 1061-1073.

Comparative Analysis of Different Hybrid Optical Amplifiers

Harminder Kaur
CT Institute of Technology & Research, Maqsudan, Jalandhar
harminder.kajal@gmail.com

Navneet Walia
CT Institute of Technology & Research, Maqsudan, Jalandhar

Amanjot Singh
CT Institute of Technology & Research, Maqsudan, Jalandhar

Sanjeev Mahajan
CT Institute of Technology & Research, Maqsudan, Jalandhar

***Abstract*-In this paper we have compared the performance of different hybrid optical amplifiers on the basis of BER and eye diagram. We have observed that under varying length from 50 km to 150 km with bit rate of 10Gbps.SOA hybrid amplifier gives better performance of 1.9170e-033 BER at fiber length 70 km.**

Keywords-DWDM, BER, EDFA, SOA.

I. INTRODUCTION

In the fiber optic communication, there is deprivation in transmission signal with the increase in distance. To compensate signal deprivation optoelectronic regenerators were used before the introduction of optical amplifier. In opto-electronic regenerators, the optical signal is first rehabilitated into electric current and then regenerated by using a transmitter. But this reduces the consistency of networks as regenerator in an active device. Therefore, up gradation of multichannel WDM network will require optical amplifier. To remove loss restrictions and to amplify the signal, the optical amplifiers are used. So, there is a demand to examine the unrepeated all optical transmission and ultra fast broadband transmission over long distances. The hybrid optical amplifier have attracted much attention as they are amplifies the broad bandwidth. The hybrid optical amplifier has wide gain spectrum ease of assimilation with other devices and low cost.

Optical amplifier can serve several purposes in the design of fiber optic communication system. Optical Amplifier amplify incident light through stimulated emission[1] . The light augmentation occurs, when incident photon and emitted photon are in phase and release two more photons. The use of optical amplifiers is attractive for Dense WDM (Wavelength Division Multiplexing) light wave system as all channels can be amplified simultaneously. In WDM multiple optical carrier at different wavelength are modulated by using independent electrical bit stream and transmitted over same fiber.

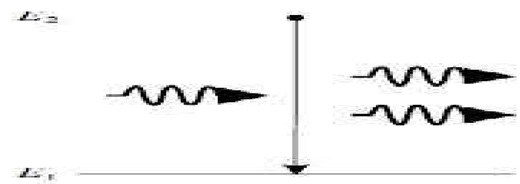


Fig. 1. Stimulated Emission[2]

The main aim of device is to achieve optical amplification and population. This can be achieved by exciting electron into higher energy level by external source called pumping.

A semiconductor laser amplifier is a modified semiconductor laser which typically has different facet reflectivity and different device length [2]. Semiconductor optical amplifier (SOA) is very similar to a laser except it has no reflecting facets. A weak signal is sent through the active region of the semiconductor which results in a stronger signal emitted from the semiconductor.

The EDFA consists of three basic components: length of erbium doped fiber, pump laser and wavelength selective coupler to combine the signal and pump wavelengths. The optimum fiber length used depends upon the pump power, input signal power, amount of erbium doping and pumping wavelength [3] . Erbium doped fiber amplifiers (EDFAs) have higher compatibility with optical fiber. Theoretically EDFA is capable of amplifying all the wavelengths ranging from 1500 to 1600 nm. Practically there are two windows of wavelength namely C and L band which allows the data signal to stimulate the excited atoms to release photons [4]. The 980-nm pump wavelength shows gain efficiencies of around 10dB/mW, while the 1480-nm pump wavelength gives efficiencies of around 5dB/mW. Usually, gains are on the order of 25 dB.

Typically noise figure lies between 4-5 dB with forward pumping and equivalent figures for backward pumping are 6-7 dB assuming 1480 nm pumping light was used.

II. SIMULATION SETUP

The different combinations of amplifiers in optical communication system can provide a better result and better feasibility for long distance transmission. In this, diagram shows that 4 channels DWDM are transmitted at 10Gb/s spaced at 100GHz with optical frequency 193.4THz. Figure 2 shows the block diagram of transmitter of DWDM 4 channel link control system in which the PRBS block uses the bit rate user variable to set the simulation bit rate, and each of the CW laser model blocks use a formula utilizing the user variables to set their wavelengths. This allows parameter scans to be performed to view the BER vs. properties such as the bit rate and frequency spacing. The PRBS generator generates the four different binary sequences. The output is fed to the electrical generator which generates the carrier signal. The laser and electrical generator output is fed to the modulator. The modulator modulates the signal and the modulated signals are multiplexed together and are passed through the single mode fiber through EDFA amplifier. As shown in figure 3, at the receiver side the signal is demultiplexed. The most efficient way to assess the quality of transmission is to analyze the eye diagrams, which show patterns of the electrical signal after detection, and to evaluate BER values of the transmitted signal as a parameter featuring best the signal distortions arising during transmission. To estimate the transmitted signal distortions caused by fiber nonlinearity, we will observe its optical spectrum.

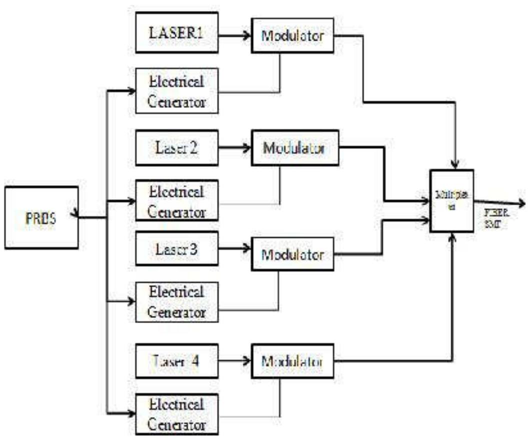


Fig. 2. Transmitter of DWDM 4 Channel Link Control

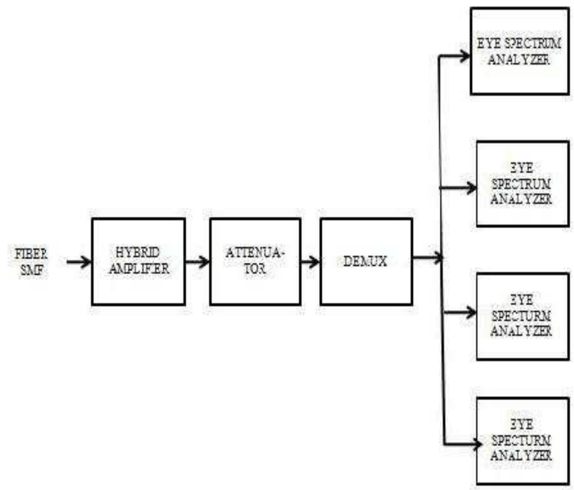


Fig. 3. Receiver of DWDM 4 Channel Link Control

III. RESULT AND DISCUSSIONS

In this work we note the performance of different amplifiers by using their combinations. The optical signal is transmitted and measured over different distance for 50, 70, 90, 110, 130 and 150 km. The BER of combination of SOA and SOA amplifier is not good. The BER is degrades with increase in the fiber length and the eye is not fully open. Table 1 shows that the bit error rate is most degrade at 150km fiber length. The BER is e-004 at 50km fiber length and 70km fiber length respectively.

Table I. BER at different fiber length

FIBER LENGTH (in kms.)	BER
50	9.1976e-004
70	1.3266e-004
90	4.5763e-006
110	1.0490e-007
130	2.0323e-007
150	4.2655e-003

Fig. 4. shows the eye spectrum of the SOA-SOA hybrid amplifier. It shows that the eye is not maximum opened. Fig. 5 shows graph between fiber length and BER.

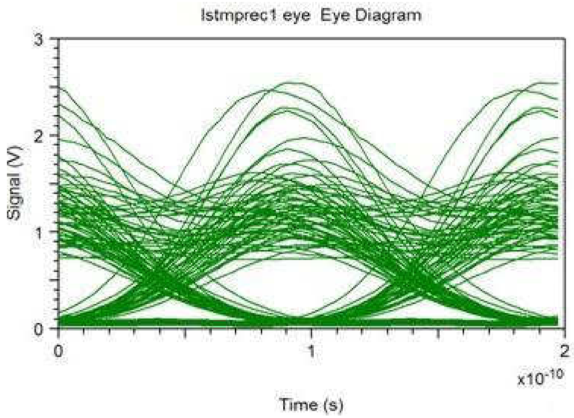


Fig. 4. Eye Diagram

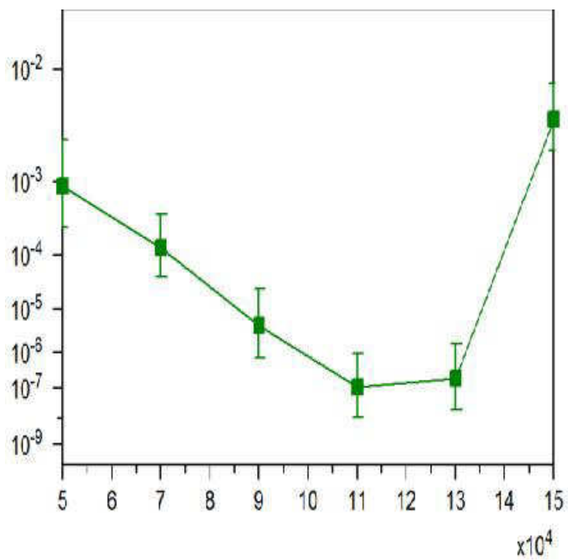


Fig.5. Graph between Fiber length and BER

Table II. BER at different Fiber Length

FIBER LENGTH	BER
50	3.2405e-031
70	1.9170e-033
90	2.1318e-020
110	6.6248e-016
130	4.4981e-007
150	5.6833e-003

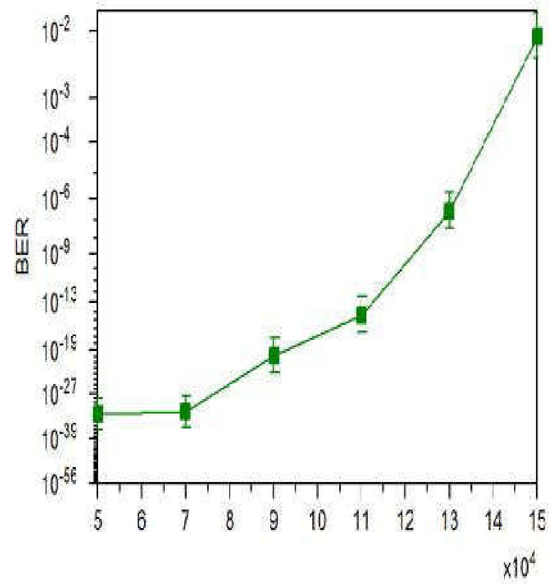


Fig. 6. Graph between Fiber length and BER

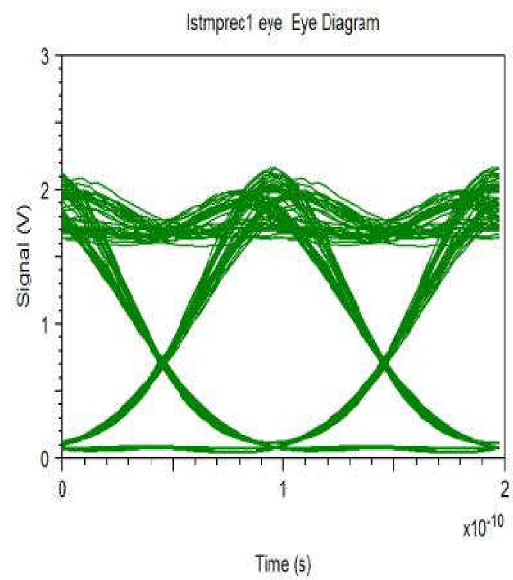


Fig. 7. Eye spectrum at 50km

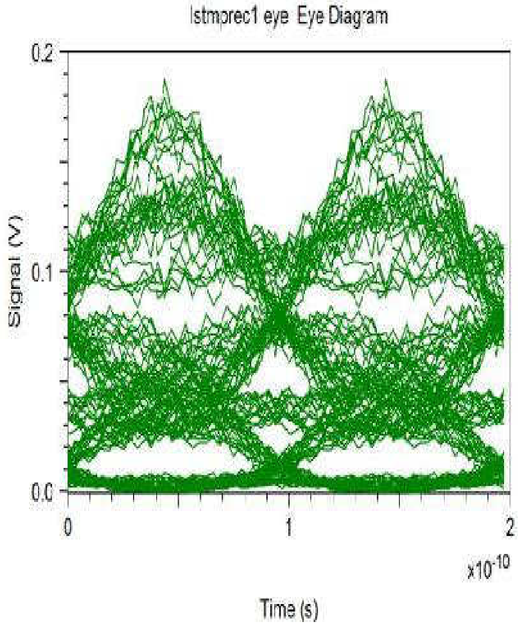


Fig. 8. Eye spectrum at 150km

The BER of combination of EDFA and EDFA is less than e^{-09} with fiber length of 150km. The bit error rate is e^{-33} at 50km. As the fiber length increases there is degradation in the signal. Table III shows the BER at different fiber length.

Table III. BER at different fiber length

FIBER LENGTH(in kms.)	BER
50	1.5479e-033
70	1.7593e-019
90	4.6210e-006
110	2.1750e-002
130	1.6441e-001
150	3.0539e-001

The eye spectrum of the system using EDFA-EDFA combination at fiber length 50 km is shown in figure 9 whereas figure 10 shows the eye spectrum of EDFA at fiber length of 150km. We observe maximum eye opening at 50km and minimum at 150km.

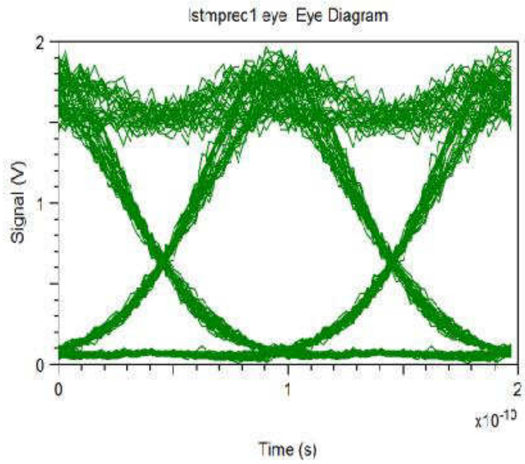


Fig. 9. Eye spectrum at 50km

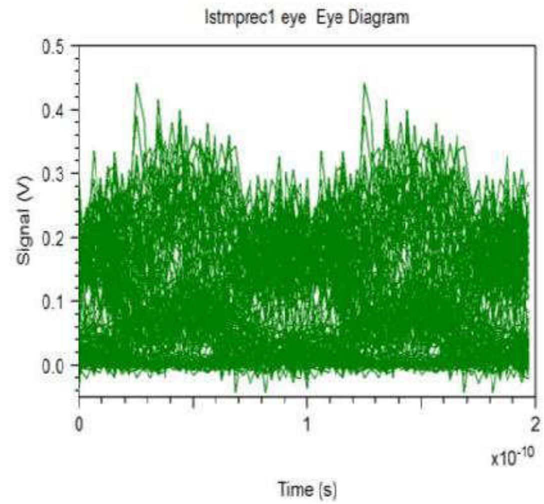


Fig. 10. Eye spectrum at 150km

IV. CONCLUSION

In this paper the performance of different combination of optical amplifier i.e. hybrid optical Amplifier for 4 channel Dense WDM link with channels spaced at 100 GHz starting at an optical frequency. These comparisons have been done by varying transmission distance in between 50 to 150km with minimum dispersion. The performance was evaluated on the basis of parameter like Min log of BER & eye opening factor. Now it is observed that SOA-EDFA provides better results for output power.

REFERENCES

[1] Cedric F. Lam, "Passive Optical Networks- Principles and Practice", Elsevier Science and Technology, 2007.

[2] Biswanath Mukherjee,"Optical WDM Networks", Springer, New York, 2006.

[3] Banerjee A., Park Y., Clarke F., Song H., Yang S., Kramer G., Kim K., Mukherjee B., “Wavelength-division-multiplexed passive optical network (WDM-PON) technologies for broadband access”, Journal of optical networking, Vol. 4, No. 11, 2005.

[4] B. Nagaraju, M.C. Paul, M. Pal, A. Pal, R.K. Varshney, B.P. Pal, S.K. Bhadra, G. Monnomc and B. Dussardier, “Design and fabrication of an intrinsically gain flattened Erbium doped fiber amplifier”, Joural of Optics Communication, Vol. 282, Pages 2335–2338,2009.

[5] J.M. Senior, “Optical Fiber Communications”, Prentice Hall, New York, 1992. Biswanath Mukherjee, “Optical WDM Networks”, Springer, New York, 2006.

[6] GP Agarwal, Fiber-optic communication systems”. New York ,Pages 546:226-260.

[7] Bobrovs V, Ivanovs G, “Comparison of Different Modulation Formats and Their Compatibility With WDM Transmission System”, 2008.

[8] M. M. J. Martini, C. E. S. Castellani, M. J. Pontes, M. R. N. Ribeiro, and H. J. Kalinowski., “Multi-pump Optimization for Raman+EDFA Hybrid Amplifiers under Pump Residual Recycling”, IEEE International Conference on Microwave and Optoelectronics

[9] Payne D.N, Mears R.J., Reekie L. and Jauncey I.M., “Low-noise erbium-doped fiber amplifier at 1.54 ^m’, Electron Letters”, Volume 23, 1987, Pages 1026-1028.

[10] Desurvire E., Giles C.R. and Simpson J.R., “Gain saturation effects in high speed, multichannel erbium doped fiber amplifiers at 1.53 p,m wavelength”, IEEE Journal of Lightwave Technology, Volume 7, 1989, Pages 2095-2104.

Sub-bandage Compression Therapy using Optical Fiber Sensors

Taranbir Kaur
Department of CSE
CTITR, Maqsudan

Manpreet kaur
Department of CSE
CTITR, Maqsudan

Gulwatanpreet Singh
Department of CSE
CTITR,Maqsudan

Geetika Mannan
Department of CSE
CTITR,Maqsudan

***Abstract*--Compression bandage therapy is used to cure the most common problem of chronic leg ulcer in human beings .The pressure exerted by the bandage can be estimated using Laplace equation given by Thomson the distribution of the pressure in the legs varies from top to bottom due to gravity .The treatment uses the three layer bandage with optical fiber sensor and Fiber Bragg grating to measure in-vivo resting and working pressure second layer of the bandage is the highest gradient pressure layer Working pressure measures during simple movements in leg is evaluated with respect to time**.

Key terms--Compression, FBG, Optical Fiber Sensor.

I. Introduction

Chronic leg ulceration is the most common problem in 1 % of the Australian population and about 5% of adult population of United States is affected by chronic ulcer every year. Chronic leg ulcer is a long lasting sore which occurs on the leg or ankle. It is basically a break in the skin which cannot heal and area of breakdown increases with time. The problem is due to badly managed diabetes, high cholesterol, smoking and poor blood circulation. In the survey it has been found that the women are more affected by this problem than men. The symptoms of a venous leg ulcer can be detected by pain, itching and swelling in the affected leg. The available treatment options include compression bandages, medication, surgery and hyperbolic oxygen therapy. The most effective way to treat the chronic leg ulcer is the bandage application or stockings because it controls the flow of blood in the leg and keeps the pressure at the ankle low when standing up. Different types of bandages are used for the treatment of leg ulcer. These bandages are made up of two, three or four layers. The efficiency of bandage therapy can be evaluated according to the amount and distribution of compressive pressure that is applied.

The Laplace Equation is used to estimate the pressure exerted by the bandage on a cylindrical object which states that the compression is directly proportional to the number of layers of bandage and bandage tension, and inversely proportional to the circumference of the object and the width of bandage. But the use of Laplace Equation in the compression therapy remains in controversy due to the irregularity of the human limb. Although different types of sensors are available for the measurement of sub bandage pressure such as Pneumatic type sensor, Liquid filled sensor, resistive type strain gauges but the Fiber Optic Sub-bandage Force/Pressure sensor has been preferred due to its high spatial resolution and adjustable sensing area that is required for an ideal sub bandage pressure sensing.

The key feature of sub bandage pressure for maintaining compression treatment needs the sensors that are thin, flexible and according to the curvature of limb. The sensor should have high resistivity, high sampling rate with low hysteresis and should be sensitive to temperature and patient movement. The key factor is to have high spatial resolution system by which a multiple of closely spaced sensing elements help concurrent measurements along the whole length of the treated limb.

The distribution of pressure in the legs varies from top to bottom due to the gravity. The effect of gravity on the flow of blood in the veins depends on how the leg is placed whether in the horizontal or blending position. Fiber Brag Grating sensors are resonant periodic structures inside the optical fiber in which index of refraction varies along the fiber. FBG sensors offer higher accuracy, longer scalability, smaller size, immunity to electromagnetic interference. FBGs are strain measuring devices which are used to analyses the stress and to show high pressure dependency.

$$\text{Pressure} = \frac{\text{Tension*no of bandage layers}}{\text{Radius*Bandage Width}} \qquad (1)$$

Bandage is an interface pressure that exerts on the limbs. This pressure depends on the type of bandage and whether the limb on which measurement is taken, immobile , giving rise to a resting pressure (RP) produced directly by passive tension exerted by compression system or mobile, giving rise to walking pressure (WP) produced intermittently on exertion when muscle is exercised.

II. Compression Bandage Used

A. Three Layer bandage:

During compression treatment optical fiber sensor array is used for measuring in-vivo resting and working pressure . A single layer compression is considered as stocking but not a bandage, to measure the exerted compression here a three layer bandage is applied to the healthy leg with sensor array. The first innovative layer of the bandage is a compression wadding using unique technology enabling binder-free combination of a wadding with a knitted layer, which offers a degree of light compression to protect the skin and avoids high pressure points on bony and slippage of the subsequent layers and The second layer provides the high pressure and from toe to knee exerted graduated pressure bandage is applied .and last third layer use to keeps the lower layer on place .Fig 1 shows the implementation of three layer bandage on affected leg with optic sensor .

Fig. 1. The three layer bandage applied on affected leg with optic sensor

B. Bandage Pressure measurement:

Preliminary results indicate that venous pressure at the ankle of the patient who is lying supine is around 10 mmHg, but due to an increase in hydrostatic pressure while standing pressure will rise by 80 mmHg .and because of low spatial resolution sensor the theoretic gradient profile to the actual healing rate has not verified .but the both resting and working sub bandage pressure gradients across the treated limb can be measured by using this. Compression is generally considered to be the standard

The research was done by an experienced chronic wound care nurse a histogram is plotted to display the record of sub bandage pressure by the sensor array with and without real time feedback with respect to sensor.

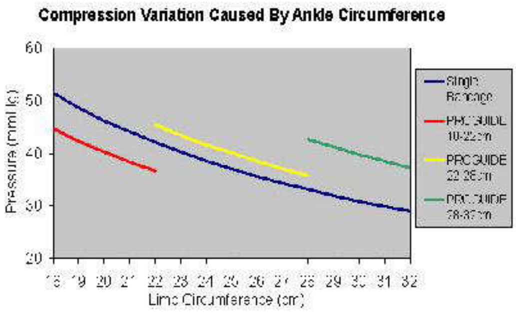


Fig. 2. Compression variations by Ankle circumference

III. CHARACTERSTICS OF OPTICAL FIBER SENSOR

FBGS technologies of Germany provide low bend loss DTG (draw tower grating)arrays formed in sensor array ,each consist 33 spectrally separated FBG elements having wavelength range 1571 nm to 1590 nm .the physical length of FBG is around 3mm with separation 10.4 mm as shown in figure here two optical fiber are entwined into a double helix form having pitch same as the separation of FBG s at the point when both the FBG s coincides it form a FBG Sensing element .each FBG in the sensing pair describes equal arc with opposite radius to provide opposite response to sideway deformation .the double helix form can handle the common variations caused by temperature fluctuation. To infer the applied force degree of Bragg Wavelength is used The compressive force applied from the top will reduces the Bragg Wavelength while the bottom FBG is elongated and the wavelength is increased and due to high rigidity and elasticity of the optical fiber low hysteresis is formed .

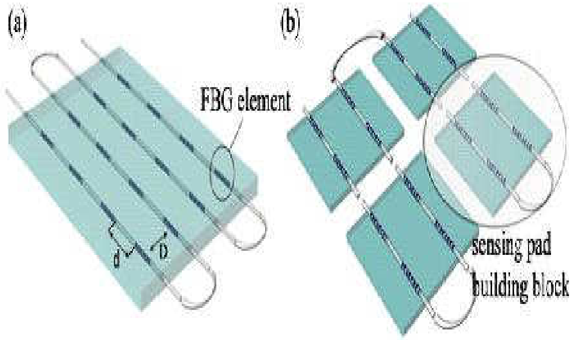


Fig. 3 a) PDMS 2D sensing surface with FBG-array optical sensing elements b) elementary 2D sensors used as building blocks for large scale surface sensors

The orientation of all the FBG pairs in the array was kept aligned with the placement of the substrates. a substrate metal disc having diameter 7.5mm and thickness of 0.9 mm is used to fix the FBG pair The substrate was hollow in the center and had a groove cut through the top such that it allows the vertically aligned FBG pair to be located in the cavity with freedom of sideways movement while supporting the entwined fibers at the horizontally aligned points. A non-stretchable tape is used to minimize the perturbation during sensor handling .the final thin flexible assembly sensor is placed under compression layer of bandage. And to simultaneously monitor the wavelength peak FBG Scan 804 interrogator is used.

The figure 4. Shows the Working pressure measurement during simple movements. During dorsiflexion the circumference of the leg changes which will increase the pressure with respect to time .It has been shown in the graph that the value of pressure in lying down position remain constant while moving the leg will change the pressure accordingly and while standing the pressure is increased to 80 mmhg

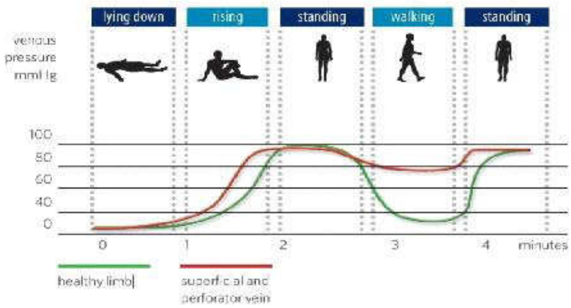


Fig. 4 change of pressure with respect to movement of Legs

IV. CONCLUSION

The most common problem of leg ulcer in human beings can be treated by applying high graduated pressure on the wound area Compression is generally considered to be the standard therapy for the treatment of venous leg ulcers. The three layer sub bandage is used with Fiber Optic sensor consisting two FBG array twined FBG sensors offer higher accuracy, longer scalability, smaller size , immunity to electromagnetic interference . FBGs are strain measuring devices which are used to analyses the stress and to show high pressure dependency. Fiber Optic Sub-bandage

Force/Pressure sensor provides high spatial resolution and adjustable sensing area that is required for an ideal sub bandage pressure sensing. An experienced wound care nurse has plotted and displayed the record of sub bandage pressure by the sensor array with and without real time feedback with respect to sensor. During dorsiflexion the circumference of the leg changes which will increase the pressure with respect to time.

REFERENCE

[1] David Hsiao-Chuan Wang, Neil Blenman, Simon Maunder, Vicki Patton , and John Arkwright "An optical fiber Bragg grating force sensor for monitoring sub-bandage pressure during compression therapy," Optics Express, Vol. 21, Issue 17, pp. 19799-19807, 2013.

[2] John J. Bergan, M.D., Geert W. Schmid-Schönbein, Ph.D., Philip D. Coleridge Smith, D.M., Andrew N. Nicolaides, M.S., Michel R. Boisseau, M.D., and Bo Eklof, M.D., Ph.D. N Engl J Med, vol. 355, pp. 488-498,August 2006.

[3] Fletcher, A., Cullum, N., and Sheldon, T. 'A systematic review of compression treatment for venous leg ulcers',Br. Med. J.,vol. 315, pp. 576–580, 1997

[4] J. Schuren and K. Mohr, "The efficacy of Laplace's equation in calculating bandage pressure in venous leg ulcers," Wounds UK4, 38–47, 2008

[5] J.W. Arkwright, S. N. Doe, M. C. Smith, N. G. Blenman, I. D. Underhill, S. A. Maunder, B. Glasscock, M. M. Szczesniak, P. G. Dinning, and I. J. Cook, "A Fibre Bragg Grating Manometry Catheter for In-vivo Diagnostics of Swallowing Disorders," TuI-2, OECC/ACOFT 2008, Sydney, Australia, 7 - 10 July (2008).

[6] Dale, J., Callum, M., andRuckley, C. V. (1983): 'How efficient is a compression bandage?', Nursing Times, November 16, pp. 49–518

Track-1
Technical Session 3-(A & B)
SOFTWARE ENGINEERING/CLOUD COMPUTING/BIOMEDICAL SIGNAL PROCESSING

Survey on Ant Colony Optimization as Computational Intelligence Technique

Preet Walia
Dept. of CSE, BFCET, Bathinda
waliaz1990@gmail.com

Bharat Naresh Bansal
Dept. of ECE,MIMT,Malout
bharat.mimt@gmail.com

Harleen Kaur
Dept. of IT,BFCET,Bathinda
harleengrover@gmail.com

***Abstract*—Fifteen years ago, when the first Ant Colony Optimization (ACO) algorithm was introduced, which was taking inspiration from ants that depends on foraging, division of labor, brood sorting, and cooperative transport etc for finding food for designing optimization algorithms. In this paper, the main idea about main ACO Algorithms like Ant System (AS), MAX-MIN Ant System (MMAS) and Ant Colony System (ACS) defined and also shows changes between them. Initially, this ACO algorithm used to solve the Travelling Salesman Problem (TSP) and finds feasible results in computational time with this algorithms. After that, this algorithms used for many different applications like NP-hard Problems, Telecommunication networks and industrial works and these were briefly described in this paper. This paper helps to better understand the theoretical properties of ACO in different fields.**

Keywords—ants foraging behavior, double bridge experiment, traveling salesman problem, metaheuristic, ant system, max-min ant system, ant colony system.

I. INTRODUCTION

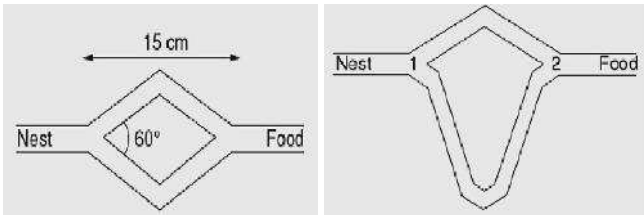
The Ant Algorithms derived from observation of real ants behavior and uses this process used as source of inspiration for design of ant algorithms for finding solution of optimization and distributed problems. The main principle is self organizing which allow coordinated behavior of real ants that can be exploited to coordinate population of artificial agents to solve computational problems. Several different aspects of the behavior of ant colonies have inspired different kinds of ant algorithms such as foraging, division of labor, brood sorting, and cooperative transport. In all these examples, ants coordinate their activities via stigmergy, a form of indirect communication mediated by modifications of the environment. For example, a foraging ant deposits a chemical on the ground which increases the probability that other ants will follow the same path. Therefore idea behind ant algorithms is to use form of artificial stigmergy to coordinate society of artificial agents.

A. Ants Foraging behavior

From research we find out that most of communication among individuals or between individual and environment are based on the use of chemicals produced by the ants. These chemicals called pheromones. This is different from humans whose important senses are visual or acoustic. On other side, ants uses pheromone trails as describe marking paths on ground that used by other society members.

B. Double Bridge Experiment

Deneubourg et al. [1] investigated the pheromone laying and following behavior of ants. In an experiment known as the “double bridge experiment'', the nest of a colony of Argentine ants was connected to a food source by two bridges of equal lengths. Ants first start to explore the surroundings of the nest and eventually reach the food source. Along their path between food source and nest, Argentine ants deposit pheromone. Initially, each ant randomly chooses one of the two bridges. However, due to random fluctuations, after some time one of the two bridges presents a higher concentration of pheromone than the other and, therefore, attracts more ants. This brings a further amount of pheromone on that bridge making it more attractive with the result that after some time the whole colony converges toward the use of the same bridge. This colony-level behavior, based on autocatalysis, that is, on the exploitation of positive feedback, can be used by ants to find the shortest path between a food source and their nest [2].



(a) (b)

Fig1. Experimental setup for the double bridge experiment. (a) Branches have equal length. (b) Branches have different length

Goss et al. [3] considered a variant of the double bridge experiment in which one bridge is significantly longer than the other. In this case, the stochastic fluctuations in the initial choice of a bridge are much reduced and a second mechanism plays an important role: the ants choosing by chance the short bridge are the first to reach the nest. The short bridge receives, therefore, pheromone earlier than the long one and this fact increases the probability that further ants select it rather than the long one [2].

II. DIFFERENT ANT COLONY OPTIMIZATION ALGORITHMS

In Ant Colony Optimization (ACO), a number of artificial ants build solutions to the considered optimization problem and exchange information on the quality of these solutions via a communication scheme that is reminiscent of the one adopted by real ants[2]. Different ant colony optimization algorithms have been proposed. The original ant colony optimization algorithm is known as Ant System [4]–[6] and was proposed in the early nineties. Since then, a number of other ACO algorithms were introduced. All ant colony optimization algorithms share the same idea, which is best illustrated through an example of how ACO algorithms can be applied. Given table shows different algorithms as:

TABLE I

Sr. No.	List of successful Ant Colony Optimization Algorithms			
	Algorithms	***Year***	***Authors***	***Reference***
	ANT SYSTEM	1991	DORIGO ET AL.	[4]-[6]
	ELITIST AS	1992	DORIGO ET AL.	[5],[6]
	ANT-Q	1995	GAMBARDELLA & DORIGO	[7]
	ANT COLONY SYSTEM	1996	GAMBARDELLA & DORIGO	[8]-[10]
	MAX-MIN AS	1996	STUTZLE & HOOS	[11]-[13]
	RANK-BASED AS	1997	BULLNHEIMER ET Al.	[14],[15]
	ANTS	1999	MANIEZZO	[16]
	BWAS	2000	CORDON ET AL.	[17]
	HYPER-CUBE AS	2001	BLUM ET AL.	[18],[19]

A. *ACO for the Travelling Salesman Problem(TSP)*

In the TSP, a set of cities and the distance between each of them is given. The goal is to find the shortest tour that allows each city to be visited once and only once. In ACO, the problem is tackled by simulating a number of artificial ants moving on a graph that encodes the problem itself: each vertex represents a city and each edge represents a connection between two cities. A variable called pheromone is associated with each edge and can be read and modified by ants.

ACO is an iterative algorithm. At each iteration, a number of artificial ants are considered. Each of them builds a solution by walking from vertex to vertex on the graph with the constraint of not visiting any vertex that she has already visited. At each step of solution construction, an ant selects the following vertex to be visited according to a stochastic mechanism that is biased by the pheromone. Suppose artificial ant on i vertex. In particular, if j has not been previously visited, it can be selected. At the end of an iteration, on the basis of the quality of the solutions constructed by the ants, the pheromone values are modified in order to bias ants in future iterations to construct solutions similar to the best ones previously constructed.

B. *Ant Colony Optimization Metaheuristic*

TABLE II

Algorithm 1	**Ant Colony Optimization Metaheuristic**

Set parameters, initialize pheromone trails

while termination condition not met **do**

ConstructAntSolutions

ApplyLocalSearch (optional)

UpdatePheromones

Endwhile

In Algorithm 1, ACO Metaheuristic basic structure is defined. Where first initialization occur after that metaheuristic iterates over three phases: At each iteration, number of solutions are constructed by number of ants, these solutions then improved through local search and at last, pheromone is updated. Description of three phases defined as:

- ConstructAntSolutions: A colony of ants build a set of solutions from elements of finite set of available solution elements. Solutions are extended with more feasible solutions with the help of objective function after each iteration.

- ApplyLocalSearch: Once solutions have been constructed and before updating the pheromone, then to improve the solutions obtained by the ants through a local search. This phase, which is highly problem-specific, Optional centralized actions, e.g.: Local optimization, Ant elitism.

- UpdatePheromones: The aim of the pheromone update is to increase the pheromone values associated with good or promising solutions, and to decrease those that are associated with bad ones. Two opposite mechanisms:(i) Pheromone deposit: Ants increase pheromone values on visited components and/or connections.(ii) Pheromone evaporation: Decrease pheromone trails on all components/connections by a same value.

III. MAIN ACO ALGORITHMS

In the Literature Survey, we studied many ACO algorithms, and describe original AS & two most successful variants MAX-MIN AS, ACS defined in detail:

A. *Ant System (AS)*

AS is the first ACO algorithm proposed in the literature [4]–[6]. Its main characteristic is that, at each iteration, the pheromone values are updated by all the m ants that have built a solution in the iteration itself. The pheromone $_{i,j}$ associated with the edge joining cities i and j, is updated as follows:

$$\tau_{i,j} \leftarrow (1-\rho).\tau_{i,j} + \sum_{k=1}^{m} \Delta\tau_{i,j}^{k}, (1)$$

Where, is evaporation rate, m is number of ants, Δ is quantity of pheromone laid on edge (i,j) by ant k:

$$\Delta\tau_{i,j}^{k} = \begin{cases} \frac{Q}{L_K} & if\ ant\ k\ used\ edge\ (i,j)in\ its\ tour \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

Where, Q is constant, is length of tour constructed by ant k.

In the construction step, ants select the city to be visited through a stochastic mechanism. When ant k is in city i and has so far constructed the partial solution s^p, the probability of going to city j is given by:

$$p^k_{(i,j)} = \begin{cases} \frac{\tau^{\alpha}_{i,j}\eta^{\beta}_{i,j}}{\sum_{c_{il}\in N(s^p)} \tau^{\alpha}_{i,j}\eta^{\beta}_{i,j}} & if\ c_{il} \in N(s^p), \\ 0 & otHerwise \end{cases} \quad (3)$$

Where, $N($ is set of feasible components, edges (i,l) l is city which is not visited by k ant, and control relative importance of heuristic value which is given by:

$$\eta_{i,j} = \frac{1}{d_{i,j}} \quad (4)$$

Where, is distance between cities I and j.

B. MAX-MIN Ant System (MMAS)

In this algorithm some improvements performed in ant system. The main principle that was followed is that only best ant will update the pheromone trails. Therefore change in update formula in (1) as:

$$\tau_{i,j} \leftarrow \left[(1-\rho).\tau_{i,j} + \sum_{k=1}^{m} \Delta\tau^{best}_{i,j}, (1)\right]^{\tau_{max}}_{\tau_{min}} \quad (5)$$

And also in (2) as:

$$\Delta\tau^{best}_{i,j} = \begin{cases} \frac{1}{L_{best}} & if\ edge\ (i,j)\ belong\ to\ best\ tour \\ 0 & otherwise, \end{cases} \quad (6)$$

Where L is the length of best tour by the best ant. Concerning the lower and upper bounds on the pheromone values τ and τ that are typically obtained and tuned on specific problem [20].

C. Ant Colony System

Local Pheromone Update is new term that added to pheromone update and performed by the all ants at the end of each construction process. This Local Pheromone Update performs great contribution in ACS and formula given as:

$$\tau_{i,j} = (1-\varphi).\tau_{i,j} + \varphi.\tau_0 \quad (7)$$

Where $\varphi \in ($ is the pheromone decay coefficient, is the initial value of pheromone.

The main goal of the local update are: decreasing the pheromone concentration, ants encourage subsequent ants to choose other edges and to produce different solutions.

Pheromone update in ACS is similar to MMAS, which can be either iteration-best or best-so-far, slightly change in formula defined as:

$$\tau_{i,j} \leftarrow \begin{cases} (1-\rho).\tau_{i,j} + \rho.\Delta\tau_{i,j}\ if\ (i,j) belongs\ to\ best\ tour, \\ 0\ otherwise. \end{cases}$$

One another important difference between ACS and AS is in decision rule used by ants during construction process. In ACS pseudorandom proportional is used.

IV. Applications of ACO Algorithms

- Major advantage of ACO algorithm is that it helps to find out high quality solution in less time.
- ACO is also applicable to dynamic problems e.g. in telecommunication networks.
- ACO used as computational intelligence technique in real world applications.

TABLE III

Sr. No.	**Applications of ACO algorithms with Problem Type**	
	Problem Type	***Problem Name***
	ROUTING	TRAVELLING SALESMAN PROBLEM VEHICLE ROUTING SEQUENTIAL ORDERING
	ASSIGNMENT	QUADRATIC ASSIGNMENT COURSE TIME TABLING GRAPH COLORING
	SCHEDULING	PROJECT SCHEDULING TOTAL WEIGHTED TARDINESS OPEN SHOP
	SUBSET	SET COVERING L-CARDINALITY TREES MULTIPLE KNAPSACK MAXIMUM CLIQUE
	OTHER	CONSTRAINT SATISFACTION CLASSIFICATION RULES BAYESIAN NETWORKS PROTEIN FOLDING PROTEIN-LIGAND DOCKING

A. Application to NP-hard(Static) problem

This approach will first try to apply to various problems and also compare the results with existing ones then find out ACO gives better results. Initially performance of ACO will be tested on Travelling Salesman Problem. Then other NP-hard problems also considered more than hundred. Many examples of Problem Type defined in Table 2 i. e. 1) Routing Problems: which defined as the distribution of goods, 2) Assignment Problems: where a set of items has to be assigned to a given number of resources, 3) Scheduling Problem: which are concerned with the allocation of resources to tasks over time, and 4) Subset Problems: where a solution to a problem is considered to be a selection of a subset of available items.

B. Application to Dynamic problems(like telecommunication network)

ACO is very effective approach for routing problems which first used in telephone networks then local area or internet. A well-known example is AntNet [21] that provided by Schoonderwoerd et al. the concept of ant-inspired routing algorithms for telecommunication networks that improved to the point of being state-of-the-art in wired networks. AntNet has been extensively tested, in simulation, on different networks and under different traffic patterns, proving to be highly adaptive and robust. A comparison with state-of-the-art routing algorithms has shown that, in most of the considered situations, AntNet outperforms its competitors. Many other algorithms provided for wired network scenario for survey see [22], [23], [24], [25].

C. Application to Academic (like industrial) problems

Number of companies have started to use ACO for real world applications and gaining success on academic problems. First e.g is on the ACO metaheuristic i.e. EuroBios (www.eurobios.com). Other one is the Company that has played, and still plays, a very important role in promoting the real world application of ACO is AntOptima (www.antoptima.com). AntOptima's researchers have developed a set of tools for the solution of vehicle routing problems whose optimization algorithms are based on ACO. Particularly successful products based on these tools are (i) DYVOIL, for the management and optimization of heating oil distribution with a non homogeneous fleet of trucks, used for the first time by Pina Petroli in Switzerland, (ii)AntRoute, for the routing of hundreds of vehicles of companies like Migros.

V. DIFFERENT POSSIBLE TOPICS IN ACO

More research on ACO there are defined more challenging problems involve multiple objectives, dynamic modifications of data, and stochastic nature of objective function and of the constraints. We focus on the extension of the applicability of ACO algorithms from discrete to continuous optimization problems and to the study of parallel implementations of ACO algorithms.

A. Dynamic Optimization problems

Here important fact is search space changing during time. In such situation, it is crucial that the algorithm be able to adjust the search direction, following the changes of the problem being solved and best e.g is telecommunication network. ACO algorithms have also been applied to dynamic versions of the TSP, where either the distance between some pairs of cities changes [26]–[28], or cities are dynamically added or removed from the set of cities to be visited. More recently, an ACS algorithm has also been applied to dynamic vehicle routing problems [29], showing good behavior on randomly generated as well as real world instances.

B. Stochastic Optimization problems

Stochastic Travelling Salesman Problem (STSP) was first stochastic problem tackled by ACO. Each city has a given probability of requiring a visit and the goal is to find an a priori tour of minimal expected length over all the cities, with the strategy of visiting a random subset of cities in the same order as they appear in the a priori tour. The first ACO algorithm for this problem was proposed by Bianchi et al. [30]. Further ACO for PTSP have been proposed by Branke & Guntsch [32], Gutjahr [33], [34], Birattari et al. [35].

C. Multi Objective Optimization

In the two-colony ACS algorithm for the vehicle routing problem with time window constraints [35] and in the MMAS for the bi-objective two-machine permutation flow shop problem [36], the multi-objective optimization problem is handled by ordering the objectives; differently, Doerner et al. [37] apply ACO to a bi-objective transportation problem and combine the objectives in a weighted sum. On the other hand, if preferences or weights cannot be given a priori, the goal is to find a set of non-dominated solutions that are optimal in the Pareto sense. The first ACO algorithm for finding non-dominated solutions was proposed by Iredi et al. [38] for the bi-objective scheduling problem. Other applications include portfolio optimization [39] & quadratic assignment [40].

D. Parallel Implementation

Any parallel models used in other population-based algorithms can be easily adapted to ACO. Two main strategies are: 1) fine-grained parallelization, 2) coarse-grained approaches. Research on parallel ACO algorithms has quickly shown that fine-grained parallelization results in a very significant communication overhead. Therefore, the focus has mostly turned to coarse-grained parallelization schemes, where p colonies run parallel on p processors [41]–[45].

E. Continious Optimization

Algorithm that designed for combinatorial optimization that tackle the continuous problem that is simplest approach which divide domain of each variables into set of intervals. ACO algorithms have been developed, which are specifically designed for continuous and mixed continuous-discrete variables [46], [47]. Many other ant inspired algorithms that based on many factors like:

- Inspired by foraging behavior, path marking
- Inspired by division of labor
- Inspired by Brood Sorting
- Inspired by cooperative transport

VI. CONCLUSION

First, These algorithms seemed as a crazy idea, when we study that these algorithms takes inspiration from natural behavior of ants. But with the help of these algorithms in many applications feasible results derived in less time, then it will be proved that it is one of the most promising approaches to the approximate solution of difficult optimization problems. Study about best use of ACO algorithms in field of dynamic space and in stochastic aspects as well as in multiple objectives are the major research work directions in the future work. A better understanding of the theoretical properties of ACO algorithm is certainly another research direction that will be pursued in the future.

REFERENCES

[1] J.-L. Deneubourg, S. Aron, S. Goss, and J.-M. Pasteels, "The self-organizing exploratory pattern of the Argentine ant," Journal of Insect Behavior, vol. 3, p. 159, 1990.

[2] Marco Dorigo, Mauro Birattari, and Thomas St¨utzle, "Ant Colony Optimization," IEEE computational intelligence magazine | nov. 2006.

[3] S. Goss, S. Aron, J.-L. Deneubourg, and J.-M. Pasteels, “Self-organized shortcuts in the Argentine ant,” Naturwissenschaften, vol. 76, pp. 579–581, 1989.

[4] M. Dorigo, V. Maniezzo, and A. Colorni, “Positive feedback as a search strategy,” Dipartimento di Elettronica, Politecnico di Milano, Italy, Tech. Rep. 91-016, 1991.

[5] M. Dorigo, “Optimization, learning and natural algorithms (in italian),” Ph.D. dissertation, Dipartimento di Elettronica, Politecnico di Milano, Italy, 1992.

[6] M. Dorigo, V. Maniezzo, and A. Colorni, “Ant System: Optimization by a colony of cooperating agents,” IEEE Transactions on Systems, Man, and Cybernetics—Part B, vol. 26, no. 1, pp. 29–41, 1996.

[7] L.M. Gambardella and M. Dorigo, “Ant-Q: A reinforcement learning approach to the traveling salesman problem,” in Proc. Twelfth International Conference on Machine Learning (ML- 95), A. Prieditis and S. Russell, Eds., Morgan Kaufmann Publishers, pp. 252–260, 1995.

[8] M. Dorigo and L.M. Gambardella, “Ant colonies for the traveling salesman problem,” BioSystems, vol. 43, no. 2, pp. 73–81, 1997.

[9] “Ant Colony System: A cooperative learning approach to the traveling salesman problem,” IEEE Transactions on Evolutionary Computation, vol. 1, no. 1, pp. 53–66, 1997.

[10] L.M. Gambardella and M. Dorigo, “Solving symmetric and asymmetric TSPs by ant colonies,” in Proc. 1996 IEEE International Conference on Evolutionary Computation (ICEC’96), T. Baeck et al., Eds. IEEE Press, Piscataway, NJ, pp. 622–627, 1996.

[11] T. St ¨utzle and H.H. Hoos, “Improving the Ant System: A detailed report on the MAX–MIN Ant System,” FG Intellektik, FB Informatik, TU Darmstadt, Germany, Tech. Rep. AIDA–96–12, Aug. 1996.

[12] T. St¨utzle, Local Search Algorithms for Combinatorial Problems: Analysis, Improvements, and New Applications, ser. DISKI. Infix, Sankt Augustin, Germany, vol. 220, 1999.

[13] T. St¨utzle and H.H. Hoos, “MAX–MIN Ant System,” Future Generation Computer Systems, vol. 16, no. 8, pp. 889–914, 2000.

[14] B. Bullnheimer, R.F. Hartl, and C. Strauss, “A new rank based version of the Ant System— a computational study,” Institute of Management Science, University of Vienna, Tech. Rep., 1997.

[15] “A new rank-based version of the Ant System: A computational study,” Central European Journal for Operations Research and Economics, vol. 7, no. 1, pp. 25–38, 1999.

[16] V. Maniezzo, “Exact and approximate nondeterministic tree-search procedures for the quadratic assignment problem,” INFORMS Journal on Computing, vol. 11, no. 4, pp. 358–369, 1999.

[17] O. Cord´on, I.F. de Viana, F. Herrera, and L. Moreno, “A new ACO model integrating evolutionary computation concepts: The best-worst Ant System,” in Proc. ANTS 2000, M. Dorigo et al., Eds., IRIDIA, Universit´e Libre de Bruxelles, Belgium, pp. 22–29, 2000.

[18] C. Blum, A. Roli, and M. Dorigo, “HC–ACO: The hyper-cube framework for Ant Colony Optimization,” in Proc. MIC’2001—Metaheuristics International Conference, vol. 2, Porto, Portugal, pp. 399–403, 2001.

[19] C. Blum and M. Dorigo, “The hyper-cube framework for ant colony optimization,” IEEE Transactions on Systems, Man, and Cybernetics—Part B, vol. 34, no. 2, pp. 1161–1172, 2004.

[20] M. Dorigo and T. Stutzle, Ant Colony Optimization, MIT Press, Cambridge, MA, 2004.

[21] G. Di Caro and M. Dorigo, “AntNet: Distributed stigmergetic control for communications networks,” Journal of Artificial Intelligence Research, vol. 9, pp. 317–365, 1998.

[22] K.M. Sim and W.H. Sun, “Ant colony optimization for routing and load-balancing: Survey and new directions,” IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2003.

[23] G. Di Caro, “Ant colony optimization and its application to adaptive routing in telecommunication networks,” Ph.D. dissertation, Universit´e Libre de Bruxelles, Brussels, Belgium, 2004.

[24] F. Ducatelle, G. Di Caro, and L.M. Gambardella, “Using ant agents to combine reactive and proactive strategies for routing in mobile ad hoc networks,” International Journal of Computational Intelligence and Applications, vol. 5, no. 2, pp. 169–184, 2005.

[25] G. Di Caro, F. Ducatelle, and L.M. Gambardella, “AntHocNet: An adaptive natureinspired algorithm for routing in mobile ad hoc networks,” European Transactions on Telecommunications, vol. 16, no. 5, pp. 443–455, 2005.

[26] M. Guntsch and M. Middendorf, “Pheromone modification strategies for ant algorithms applied to dynamic TSP,” in Applications of Evolutionary Computing: Proc. EvoWorkshops 2001, ser. LNCS, E. J. W. Boers et al., Eds., Springer Verlag, vol. 2037, pp. 213–222, 2001.

[27] “A population based approach for ACO,” in Applications of Evolutionary Computing, Proc. EvoWorkshops 2002: EvoCOP, EvoIASP, EvoSTim, ser. LNCS, S. Cagnoni et al., Eds., Springer Verlag, vol. 2279, pp. 71–80, 2002.

[28] C.J. Eyckelhof and M. Snoek, “Ant systems for a dynamic TSP: Ants caught in a traffic jam,” in Proc. ANTS 2002, ser. LNCS, M. Dorigo et al., Eds., Springer Verlag, vol. 2463, pp. 88–99, 2002.

[29] R. Montemanni, L.M. Gambardella, A.E. Rizzoli, and A.V. Donati, “Ant colony system for a dynamic vehicle routing problem,” Journal of Combinatorial Optimization, vol. 10, pp. 327–343, 2005.

[30] L. Bianchi, L.M. Gambardella, and M. Dorigo, “An ant colony optimization approach to the probabilistic traveling salesman problem,” in Proc. PPSN-VII, ser. LNCS, J. J. Merelo et al., Eds., Springer Verlag, vol. 2439, pp. 883–892, 2002.

[31] J. Branke and M. Guntsch, “New ideas for applying ant colony optimization to the probabilistic TSP,” in Applications of Evolutionary Computing, Proc. EvoWorkshops 2003, ser. LNCS, G. R. Raidl et al., Eds., Springer Verlag, vol. 2611, pp. 165–175, 2003.

[32] W.J. Gutjahr, “A converging ACO algorithm for stochastic combinatorial optimization,” in Proc. SAGA 2003, ser. LNCS, A. Albrecht and T. Steinh¨ofl, Eds., Berlin, Germany: Springer Verlag, vol. 2827, pp. 10–25, 2003.

[33] “S-ACO: An ant based approach to combinatorial optimization under uncertainity,” in Proc. ANTS 2004, ser. LNCS, Dorigo et al., Eds., Berlin, Germany: Springer Verlag, vol. 3172, pp. 1–12, 2004.

[34] M. Birattari, P. Balaprakash, and M. Dorigo, “ACO/F-Race: Ant colony optimization and racing techniques for combinatorial optimization under uncertainty,” in MIC 2005. The 6th Metaheuristics International Conference, K. F. Doerner et al., Eds., Vienna, Austria: University of Vienna, Department of Business Administration, pp. 107–112, 2005.

[35] M. Birattari, P. Balaprakash, and M. Dorigo, “ACO/F-Race: Ant colony optimization and racing techniques for combinatorial optimization under uncertainty,” in MIC 2005: The 6th Metaheuristics International Conference, K. F. Doerner et al., Eds., Vienna, Austria: University of Vienna, Department of Business Administration, pp. 107–112, 2005.

[36] V. T’kindt, N. Monmarch´e, F. Tercinet, and D. La¨ugt, “An ant colony optimization algorithm to solve a 2-machine bicriteria flowshop scheduling problem,” European Journal of Operational Research, vol. 142, no. 2, pp. 250–257, 2002.

[37] K. Doerner, R.F. Hartl, and M. Reimann, “CompetAnts for problem solving—the case of full truckload transportation,” Central European Journal for Operations Research and Economics, vol. 11, no. 2, pp. 115–141, 2003.

[38] S. Iredi, D. Merkle, and M. Middendorf, “Bi-criterion optimization with multi colony ant algorithms,” in First International Conference on Evolutionary Multi-Criterion Optimization, (EMO’01), ser. LNCS, E. Zitzler et al., Eds., Springer Verlag, vol. 1993, pp. 359–372, 2001.

[39] K.F. Doerner, W.J. Gutjahr, R.F. Hartl, C. Strauss, and C. Stummer, “Pareto ant colony optimization in multiobjective project portfolio

selection with ILP preprocessing," European Journal of Operational Research, vol. 171, no. 3, pp. 830–841, 2006.

[40] M.L. Ib´a nez, L. Paquete, and T. St¨utzle, "On the design of ACO for the biobjective quadratic assignment problem," in Proc. ANTS'2004, ser. LNCS, M. Dorigo et al., Eds., Springer Verlag, vol. 3172, pp. 214–225, 2004.

[41] T. St¨utzle, "Parallelization strategies for ant colony optimization," in Proc. PPSN-V, ser. LNCS, A. E. Eiben et al., Eds., Springer Verlag, vol. 1498, pp. 722–731, 1998.

[42] E.-G. Talbi, O. Roux, C. Fonlupt, and D. Robillard, "Parallel ant colonies for combinatorial optimization problems," in Parallel and Distributed Processing, 11 IPPS/SPDP'99 Workshops, ser. LNCS, J. Rolim et al., Eds., vol. 1586, pp. 239–247, 1999.

[43] M. Middendorf, F. Reischle, and H. Schmeck, "Multi colony ant algorithms," Journal of Heuristics, vol. 8, no. 3, pp. 305–320, 2002.

[44] M. Manfrin, M. Birattari, T. St¨utzle, and M. Dorigo, "Parallel ant colony optimization for the traveling salesman problem," in Proc. ANTS 2006, ser. LNCS, M. Dorigo et al., Eds., Springer Verlag, vol. 4150, pp. 224–234, 2006.

[45] S. Benker, K.F. Doerner, R.F. Hartl, G. Kiechle, and M. Lucka, "Communication strategies for parallel cooperative ant colony optimization on clusters and grids," in Complimentary Proc. PARA'04 Workshop on State-of-the-Art in Scientific Computing, pp. 3–12, 2005.

[46] K. Socha, "ACO for continuous and mixed-variable optimization," in Proc. ANTS 2004,ser. LNCS, M. Dorigo et al., Eds., Springer Verlag, vol. 3172, pp. 25–36, 2004.

[47] K. Socha and M. Dorigo, "Ant colony optimization for continuous domains," European Journal of Operational Research, in press.

Hybrid Approach of Steganography for Advanced Security

Priya Joshi
Mody University of Science and Technology
Lakshmangarh, Rajasthan, India
priyaj.joshi1@gmail.com

Anand Sharma
Mody University of Science and Technology
Lakshmangarh, Rajasthan, India
anand_glee@yahoo.co.in

Abstract—**The main goal of this paper is to give new insight and direction on hiding secure messages by proposing a few changes in the steganographic techniques. Starting with the existing methods and techniques of Steganography, and the problems faced due to the use of same image repeatedly in the form of cover image, making it easier for the steganalyst to be suspicious about the transferred information. We then illustrate a new method for securing the data by using steganography twice as well as securing the message with the help of double keys.**

Keywords—Steganography, Steganalysis, Cover object, Stego-Object, Cryptography.

I. INTRODUCTION

Since the time information sharing has come to exist, people have been eager to hide secret messages in any way they can. For, useful information if gets into wrong hands, may lead to disaster and loss in terms unknown. There have been a numerous number of ways in which data is secured and is hidden while being transferred, such that its safety is most prior. Yet, it has been almost impossible to stop unauthorized access to useful information.

There are numerous cryptographic techniques for encryption of the message being passed, yet cryptanalysts have left no stone unturned in making their way out. Even in the use of steganography as we are going to explain further, there have been steganalyst who have ways un-hiding the hidden messages. We in this paper propose a new idea of hiding the secret information (message intended to be delivered from one person to other without being seen by the third party) in a way in which it does not create any suspicion in the eye of a staganalyst.

Very often, a message is encrypted before being hidden in a message in order to achieve a better level of secrecy (which provides a basic example on how to combine cryptography and steganography).Steganography embeds the secret message in a harmless looking cover, such as a digital image file. The need for steganography is obvious but what is more obvious is the need for more research in the field. Simple techniques are easily detectable and there is a whole field of defeating steganographic techniques called steganalysis. As it is always the case, advances in steganography are usually countered by advances in steganalysis which makes it a constantly evolving field. Since most steganographic system use digital images as cover, the whole field has borrowed methods and ideas from the closely related fields of watermarking and finger-printing which also manipulate digital audio and video, for the purpose of copyright. Even though, in principle, much aspect of images can be manipulated, in reality most stego systems aim for the preservation of the visual integrity of the image. Early stego systems goals were to make changes not detectable by the human eye. This feature is not enough because statistical methods can detect the changes in the image even if it is not visible. Image compression also plays a role in steganography because it was found that on many occasions the result depend on the compression scheme used. Steganographers struggle to find more efficient methods to embed a secret message in a cover object, only to be defeated by techniques derived by steganalysts. Anderson and Petitcolas [2] explore the theoretical limits of steganography, and show how much hard it is to get a scheme that gives unconditional covertness, in the sense that the one-time pad provides unconditional secrecy

For cryptography. More recently, Hopperet al. [3] used cryptographic and complexity theoretic proof techniques to show that the exis-tence of one-way functions and access to a channel oracle are both necessary and sufficient conditions for the existence of secure steganography relative to any channel. They constructed a steganographic protocol that is provably secure and has nearly optimal bandwidth under these conditions when compared with known provably secure constructions, which is the first known example of a general provably secure steganographic protocol. Despite this fact, we believe that a simple modification of existing steganographic protocols may enhance drastically the security of a steganographic system (from a practical point of view).

Since the time information sharing has come to exist, people have been eager to hide secret messages in any way they can. For, useful information if gets into wrong hands, may lead to disaster and loss in terms unknown. There have been a numerous number of ways in which data is secured and is hidden while being transferred, such that its safety is most prior. Yet, it has been almost impossible to stop unauthorized access to useful information.

There are numerous cryptographic techniques for encryption of the message being passed, yet cryptanalysts have left no stone unturned in making their way out. Even in the use of steganography as we are going to explain further, there have been steganalyst who have ways un-hiding the hidden messages. We in this paper propose a new idea of hiding the secret information (message intended to be delivered from one person to other without being seen by the third party) in a way in which it does not create any suspicion in the eye of a staganalyst.

Very often, a message is encrypted before being hidden in a message in order to achieve a better level of secrecy (which provides a basic example on how to combine cryptography and steganography).Steganography embeds the secret message in a harmless looking cover, such as a digital image file. The need for steganography is obvious but what is more obvious is the need for more research in the field. Simple techniques are easily detectable and there is a whole field of defeating steganogrphic techniques called steganalysis. As it is always the case, advances in steganography are usually countered by advances in steganalysis which makes it a constantly evolving field. Since most steganographic system use digital images as cover, the whole field has borrowed methods and ideas from the closely related fields of watermarking and finger-printing which also manipulate digital audio and video, for the purpose of copyright. Even though, in principle, much aspect of images can be manipulated, in reality most stego systems aim for the preservation of the visual integrity of the image. Early stego systems goals were to make changes not detectable by the human eye. This feature is not enough because statistical methods can detect the changes in the image even if it is not visible. Image compression also plays a role in steganography because it was found that on many occasions the result depend on the compression scheme used. Steganographers struggle to find more efficient methods to embed a secret message in a cover object, only to be defeated by techniques derived by steganalysts. Anderson and Petitcolas [2] explore the theoretical limits of steganography, and show how much hard it is to get a scheme that gives unconditional covertness, in the sense that the one-time pad provides unconditional secrecy

For cryptography. More recently, Hopperet al. [3] used cryptographic and complexity theoretic proof techniques to show that the exis-tence of one-way functions and access to a channel oracle are both necessary and sufficient conditions for the existence of secure steganography relative to any channel. They constructed a steganographic protocol that is provably secure and has nearly optimal bandwidth under these conditions when compared with known provably secure constructions, which is the first known example of a general provably secure steganographic protocol. Despite this fact, we believe that a simple modification of existing steganographic protocols may enhance drastically the security of a steganographic system (from a practical point of view).

II. STEGANOGRAPHY

Three different aspects in information-hiding systems contend with each other: capacity, security, and robustness [4]. Capacity refers to the amount of information that can be hidden in the cover medium, security to an eavesdropper’s inability to detect hidden information, and robustness to the amount of modification the stego medium can withstand before an adversary can destroy hidden information. Information hiding generally relates to both watermarking and steganography. A watermarking systems primary goal is to achieve a high level of robustness. Steganography, on the other hand, strives for high security and capacity.

Modern steganography attempts to be detectable only if secret information is known namely, a secret key. This is similar to Kerckhoffs Principle in cryptography which holds that a cryptographic systems security should rely solely on the key material. Classical steganography concerns itself with ways of embedding a secret message. The embedding usually uses a secret key as we just explained. The terminology researchers use was agreed at the First International Workshop on Information Hiding [5]. We use a cover object (e.g. image, text, etc.) to embed a secret message. The resulting object is called a stego object (e.g. stego-image).

A. Steganographic System

A steganographic system is a mechanism that embeds a secret message m in a cover object c Using a secret shared k. The result is a stego object s which carries the message m. formally we defined the stegosystem as a pair of mappings (F; G) with F serves as the embedding function and G as the extraction function. s = F(c; m; k) m = G(s; k) If M is the set of all possible messages then the embedding capacity of the stegosystem is log 2 M 201 bits. The embedding efficiency is defined as

$$e = \log M\ d(c; s)\ 2 \qquad (1)$$

The set of all cover objects C is sampled using a probability distribution P(c) with c 2 C, giving the probability of selecting a cover object c. If the key and message are selected randomly then the Kullback-Leibler distance

$$KL(PjQ) = X\ c2C\ P(c) \log P(c)\ Q(c) \qquad (2)$$

Gives a measure of the security of the stegosystem.

The three quantifiers defined above: capacity, efficiency and security are the most important requirements that must be satisfied for any steganographic system. In reality, determining the best embedding function from a cover distribution is an NP-hard problem [6]. In addition, combining cryptography and steganography adds another layer of security [7]. Before embedding a secret message using steganography, the message is first encrypted. The receiver then should have both the stego-key in order to retrieve the encrypted information and the cryptographic key in order to decrypt it.

B. Image as Cover Media

Steganographic systems for the JPEG format seem more interesting than others (e.g. BMP) because the systems operate in a transform space and are not affected by visual attacks [8]. It would be helpful to review the encoding scheme of some image formats. The GIF format is a simple encoding of the RGB colors for each pixel using an 8-bit value. The color is not specified directly; rather the index into a 256 element array is selected. After the encoding the whole image is compressed using LZW lossless technique. In the JPEG format, first each color is converted from RGB format to Y C where the luma (Y) component representing the brightness of the pixel is treated differently than the chroma components (CBCR) which represent color difference. his difference of treatment is due to the fact that the human eye discerns changes in the brightness much more than color changes. Doing such a conversion allows greater compression without a significant effect n perceptual image quality. One can achieve higher compression rate this way because the brightness information, which is more important to the eventual perceptual quality of the image, is confined to a single channel. Once this is done for each component the discrete cosine transform (DCT) is computed to transform 8x8 pixel blocks of the image into DCT coefficients. The co-efficient are computed as:

$$F(u;v)= {}^{7}_{x=0}\ {}^{7}_{y=0}G(x;y)\{\cos(2x+1)\ u\}/16\{\cos(2x+1)\ v\}/16 \qquad (3)$$

After the DCT is completed the coefficients F(u; v) are quantized using elements from a table. Many different steganographic methods have been proposed during the last few years. Most of them can be seen as substitution systems (which are based on the Least Significant Bit (LSB) encoding technique). Such methods try to substitute redundant parts of a signal with a secret message.

III. IMAGE BASED STEGANOGRAPHY

The function denoted by F in Figure below represents the embedding function we are going to explain in this section. The symbol F −1 indicates the extraction function, since it is conceptually the inverse of embedding. We will call ISC (Image based Steganography and Cryptography) the algorithm which carries on such functions.

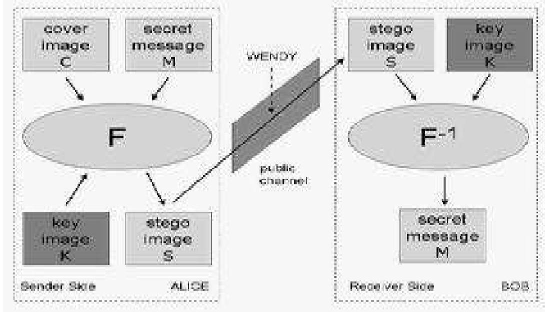


Fig.1. Image based Steganography and Cryptography

IV. PROPOSED METHOD

In various numbers of papers where steganography is combined with cryptography to provide a hidden as well as secure message, we still find them to be prone to attacks by steganalyst or cryptanalyst, any one of the analyst does their job well, the security of the message is challenged. As, observed from the papers (Combining Steganography and Cryptography: New Directions by Khalil Challita and Hikmat Farhat, Image Based Steganography and Cryptography by Domenico Bloisi and Luca Iocchi) the main problem arises is of the stegano-object which makes the steganalyst suspicious about the hidden message. Thus, the idea proposed is to cover the Stefano-object with one more image which is randomly chosen at instance hence giving no suspicion to the steganalyst.

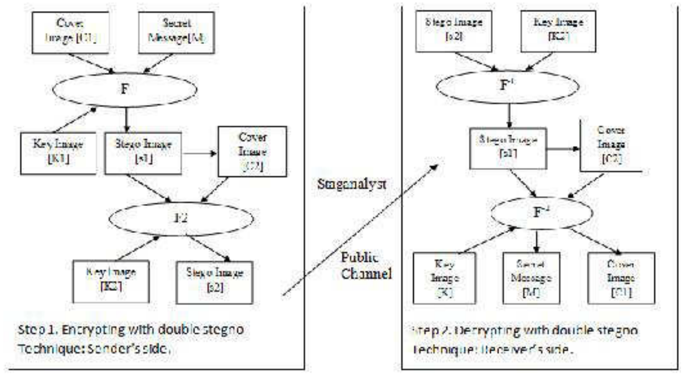


Fig 2.Steps for Encryption and Decryption

A. Step: 1 Sender's side:

We consider at first a given message which is encrypted using symmetric key cryptography, now we implement Steganography using a cover image C1. The function F encrypts the message with a Key K1 which is private and then we get our first stego Image. Further we use this stego image and again implement Steganography using second Cover Image C2. This is done to provide one more layer of security tour message... The function again does the same job, this time with a public key and the observed one is second stego image which is more secure.

B. Step: 2 Receiver's side:

Now we have stego image s2 received by the receiver, which is then decrypted using the public key K2. After which we receive the stego image s1 and the cover image C2 which are further decrypted using the private key K1. Which results in giving us the cover image C1 as well as the Secret message?

V. ALGORITHM

MK2CO(M, 2, c o v e r o b j e c t s , t e c h n i q u e) ; Hide the secret message M i n 2 c o v e r objects using two key concept , one public key and the other being a private Key (e. g. Mathematical Function) that distributes the bit of Mover the two cover objects, binding them with key and returning the stego object

M as follows (i.e. the secret key):

Store the first bit in C1, the second in C2 and covering it with a public Key, and so on until Ck=2; and then go back to C1...

For example, the first and second Assume that we need to store the secret message M = 110011 in k = 2 cover images C1; C2. We could for example use the well-known LSB technique to store

bits of M would be stored in C1, and only bit 0 in C2; Cryptanalyst will have to determine 4 key elements in order to completely uncover the secret message:

1. The number of cover objects used (i.e. *k*).
2. All the stego-objects used to hide *M*.
3. The public key for decrypting the message
4. The algorithm (i.e. secret key) used to hide *M* in the cover-objects.

Note that we did not include the *technique* used because it is usually well-known to the steganalyst.

We believe that this protocol achieves a very high level of confidentiality even if we use a relatively weak technique (such as LSB) to hide a message in several cover objects. This is because a steganalyst will not be able to recover *M* even if he/she suspects the presence of hidden information in some of the stego-objects used to hide *M*.

VI. CONCLUSION

The proposed technique of double stegnography relieves us from the suspicion of steganalyst as well as provides one more strong layer of hidden security on the message to be transmitted.Allows use of two cover images and implementing two keys making it hard to crack and increases the analysis time by any unauthorized user.

REFERENCES

[1] R. Anderson, F. Petitcolas. On the limits of steganography. In IEEE Journal on Selected Areas in Communications, volume 16, number 4, 1998.

[2] N. Hopper and L. Von Ahn and J. Langford., "Provably Secure Steganography", IEEE 2011.

[3] B. Chen and G.W. Wornell. Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding. In IEEE Trans. Information Theory, volume 47, no. 4, pp: 1423–1443, 2001.

[4] A.Kerckhoffs. La Cryptographie Militaire (Military Cryptography). In J. Sciences Militaires (J. Military Science, in French), 1983.

[5] Cormen, Leiserson, Rivest, and Stein. Introduction to algorithms. The MIT press, second edition, 2001.

[6] R. Krenn. Steganography and steganalysis.Whitepaper, 2004.

[7] K. Lee, A. Westfeld, and S. Lee. Generalised category attack: improving histogram-based attack on imaging. In Proceedings of the 9th international conference on Information hiding, IH'07, pages 378–391, Berlin, Heidelberg, 2007. Springer-Verlag.

Analysis and Simulation of Tinyos Application using Avrora and Tossim

Preeti Bisht
Computer Science and Engineering Department
Thapar University
Patiala, India
preetitheleo@gmail.com

Anil K. Verma
Computer Science and Engineering Department
Thapar University
Patiala, India
akverma@thapar.edu

***Abstract*—Finding a platform to run wireless sensor network platform is a difficult task as there is pros and cons of every platform. After study of each platform about their advantage and disadvantage and AVRORA, TOSSIM have been chosen to develop TinyOS applications. There are many other platforms where Wireless applications run like NS2, OMNeT++. In this paper there is a study of TinyOS and deployment of application over the platform like TOSSIM and AVRORA. Wireless embedded sensor network applications are supported by TinyOS. A sensor network may have large number of low power nodes. These nodes are executed in concurrent manner. Study of TinyOS applications such as Blink LEDs, Radio count to LEDs and Sense and deploy and simulate them on TOSSIM and analyze these applications via AVRORA using various commands.**

Keywords—TinyOS, JTossim, AVRORA, TOSSIM, NS2,OMNeT++.

I. INTRODUCTION

Wireless sensor network (WSN) [2] refers to a group of dispersed and dedicated sensors for recording and monitoring the physical conditions of the environment and organizing the collected data at a central dedicated location. WSN is also called Sensornet [3] or embedded sensor network.

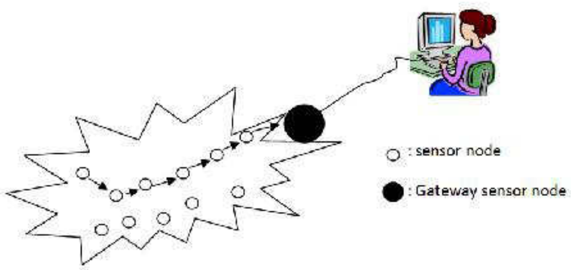


Fig. 1. Wireless sensor network

A Wireless sensor network [9] has any number of sensor nodes from a few hundreds to thousands. Sensor node is equipment that consist a radio transceiver along with an antenna, a micro-controller, an electronic circuit, and an energy source (i.e. battery). The size of the sensor nodes can also range from the size of piece of sugar to a size of telephone. Prices of sensor nodes may vary from few cents to hundreds and thousands of dollars which is based on the functionality parameters of sensor like energy consumption, computational speed rate, bandwidth, and memory. Every node in the network senses the environmental changes and passes this information to other nodes so that changes are in knowledge of every other node. Sensor nodes are great for deployment in hostile environment or very large geographical areas. As large number of sensor nodes are usually deployed on remote and inaccessible places. Deployment and maintenance is easy and scalable.

Work on TinyOS is extremely difficult that's why in the middle of the research many researchers switch their OS to other tools like NS2, OMNeT++ etc. But these tools have many limitations and we choose TinyOS over them.

To develop an application for wireless sensor network a platform is needed where deployment of application has been done. And a study on merits and demerits of various platforms like NS2, OMNeT++ has been done and chose TOSSIM and AVRORA over them.

A. Limitations of NS2 [10]

•As the layered protocols of WSN are simulated but their applications are not so they hardly acquire correct results.

•NS2 is designed is a general network simulator and doesn't consider characteristics of WSN like bandwidth, power consumption or energy saving in WSN.

•It doesn't support more than 100 nodes; if we increase number of node then tracing file will be too large to manage.

B. Limitations of OMNeT++ [10]

•OMNeT++ does not support the library of modules which is specifically used for WSNs simulation.

•Application developed in OMNeT++ have to face the compatibility problem, this makes combination of two models difficult and program may have high probability report bug.

After studying the limitations of various platforms TinyOS is chosen. But the research not ends here simulation and analysis on TinyOS is also a major field. So to simulate the application on TinyOS is through TOSSIM and JTossim and analysis is through AVRORA.

II. TINYOS

TinyOS is an open-source operating system useful for sensor networks [4][5]. As memory constraint is a severe problem so TinyOS use Component based approach which minimizes the size of code and quickly finds the result. Its component library consists of various network protocols, distributed services, drivers, and data collection tools (all of which can be used as it is or can be further refined for a customizing an application). TinyOS's event-driven execution model allows fine-grained power management still enables the scheduling flexibility necessary for the nature of wireless communication which cannot be guessed.

A. TinyOS programming concepts

The Tiny operating system [4][5], libraries, and applications are all written in NesC. NesC is a new structured component-based language. The NesC language is basically for embedded systems like sensor networks. NesC is very much similar to C, but supports as a backbone for concurrency model of TinyOS. NesC allows structuring, naming, and linking together all the components into efficient network embedded systems. The main agenda is to allow application program designers to build components that can be easily composed into complete, concurrent systems, and yet perform rigorous checking at compile time. TinyOS and NesC (Important concepts) as follow [4]

i. NesC Application

A TinyOS/NesC application has components which are linked ("wired") together to form a run-time executable .We can have more than one module file and a configuration file at the top level.

ii. Components

Components are the basic building blocks for NesC applications. There are two types of components: modules and configurations. A TinyOS component can provide and use interfaces.

iii. Module

A component that can implements one or more interfaces.

iv. Configuration

A component can wire other components; interfaces used by others components that are provided by others. This is called wiring. For building an application, there is a need of developing a configuration which consists of wiring of modules. Every nesC application has a top-level configuration file which consists of the components present in the application and how they call one another.

v. Interfaces

An interface is used to provide an abstract definition of the interaction of two components. It declares a set of functions that the interface's provider must implement, these functions are called commands. Another set of functions are also declared in interfaces, these are called the events, which the interfaces' require must implement.

vi. Commands

These are functions defined within an interface and implemented by the interface provider.

vii. Events

These are functions declared within an interface and implemented by the interface user or more specifically the component user.

viii. Concurrency Model

TinyOS executes its application in the form of threads. Threads are classified into two broad categories i.e. tasks and hardware event handlers.

- Task: They are the functions which are postponed or may be put off by some reason. If they are scheduled for execution then they run till they complete their execution, processes do not preempt each other. It enables component to perform background processing.
- Hardware event handlers*:* They are executed in response to the interrupt in hardware and also run till completion. Unlike the task thread they preempt the other hardware event handler. Commands and events must be declared with the async keyword, because they preempt each other so run in asynchronous manner.

B. Comparison between large embedded systems and TinyOS

TABLE I. A COMPARISON

Problem with Large Scale Deeply Embedded System[13]	**Why TinyOS is better option[13]**
❖ Large memory & storage requirement. ❖ Unnecessary and overkill functionality (address space isolation, complex I/O subsystem, UI) for given scenario. ❖ Relative high system overhead (e.g. context switch) Require complex and power consuming hardware support.	❖ Extremely small footprint ❖ Extremely low system overhead. ❖ Extremely low power consumption. ❖ NO Kernel therefore direct hardware manipulation. ❖ NO Process management therefore only one process on the fly. ❖ NO Virtual memory therefore single linear physical address space. ❖ NO Dynamic memory allocation i.e. Assigned at compile time and NO Software signal or exception but function Call instead.

	❖ Goal of TinyOS is to strip down memory size and system overhead

C. Architecture of TinyOS

TinyOS Architecture has following components which work together to make TinyOS applications work properly. So it's important to understand the details of TinyOS architecture before analysis and simulation of TinyOS applications.

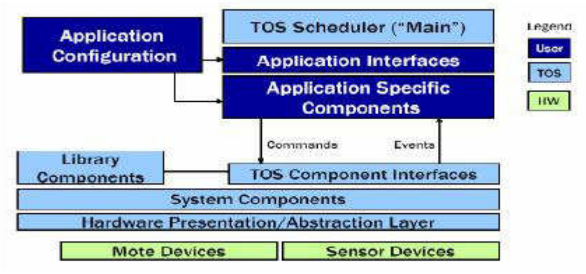


Fig. 2. TinyOS architecture [12]

i. Hardware(Mote /sensor devices)

Motes or sensor device are hardware devices on which simulation is done. Application has to be installed on these motes. Motes are of many types i.e. MicaZ, Mica2, telosb. MicaZ is used for simulation in this project. The micaZ [1] is a 2.4GHz Mote used for enabling low-power, wireless sensor networks.

ii. TinyOS

TinyOS inbuilt component are used to communicate with user application via commands and events such as led0On () etc. applications are supported by TinyOS libraries like "timer.h" and main scheduler schedule the tasks of TinyOS application. Abstraction layer hides the complexity of TinyOS and act as glue between hardware devices and TinyOS components.

iii. User

User deal with application parts, configuration file has wiring between the interfaces and application components so that they call each other.

III. Study of Simulators

After complete study of TinyOS, study of simulator is also important. Simulators used are TOSSIM and AVRORA.

A. TOSSIM

TOSSIM is a simulator designed for WSN applications running on TinyOS. TOSSIM was developed by US Berkeley's TinyOS project team in 2003. TOSSIM is a discrete event simulator. People run TOSSIM on Linux based operating system or on Cygwin on windows. [7] [10] TinyOS applications need thousands of nodes and to test the results of those nodes is an expensive process, so instead of directly simulate an application on live motes, and simulate them on virtual environment. TOSSIM allows debugging, testing, and analyzing the algorithms in a controlled environment. Users easily debug TinyOS applications using debuggers. TOSSIM is built in python and emphasize code readability. Main goal of TOSSIM is to focus on simulation and execution of TinyOS application, rather than simulating the real world problem. Recently a work on collection tree protocol is published in wireless sensor network via TOSSIM [6].

TABLE II. OPTIONS AND USAGE PROVIDES BY TOSSIM

OPTIONS	**USAGE**
-h	Help (Display this message)
-gui	pauses simulation waiting for GUI to connect
-a=<model>	specifies ADC model (generic is default)
-b=<sec>	motes boot over first <sec> seconds (default: 10)
-ef=<file>	use <file> for EEPROM; otherwise anonymous file is used
-l=<scale>	run sim at <scale> times real time (fp constant)
-r=<model>	specifies a radio model (simple is default)
-rf=<file>	specifies file input for lossy model (lossy.nss is default)
-s=<num>	only boot <num> of nodes
-t=<sec>	run simulation for <sec> virtual seconds
num_nodes	number of nodes to simulate

B. AVRORA

AVRORA [8][10] is a JAVA based simulation tool developed by University of California and designed for WSN applications analysis. AVRORA is used to simulate AVR-based microcontroller on MICA2 and MICAZ. It doesn't support GUI. AVRORA is an energy consumption analysis tool and deals with 10,000 or more number of nodes and performs as much as tens of times faster than other simulators with great accuracy. AVRORA handles 25 physical nodes in real-time while other support 5-6 physical nodes (i.e. motes). AVRORA is instruction level simulator which overcomes the limitations of TOSSIM.

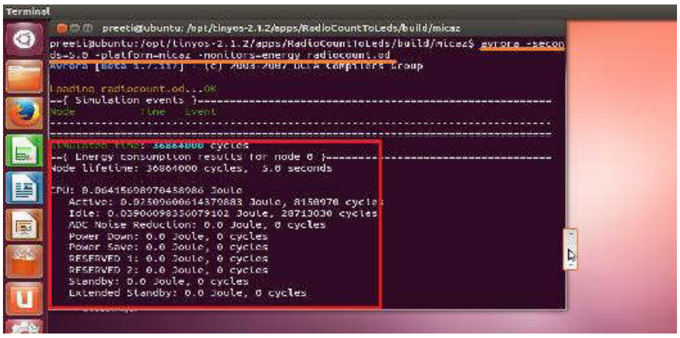


Fig. 3. Snapshot of AVRORA

To get help from AVRORA in TinyOS Avrora –help

Help has following categories:

- actions
- all
- architecture
- inputs
- monitors
- simulations
- topologies

Applications run on Avrora are BlinkLed, RadioCountToLeds and Sense. The main objective of study of these applications is they are sub module of the thesis module. Thesis is to implement and improve the existing WSN protocols. Radio count to LED application is a counter based application. In this application a AM packet is broadcast to a network and value of AM packet is updated every time it get updated. This application is useful to test basic AM communication and how timer works. Blink LED is useful to program the LED in such a way that they toggle on the given frequency.

IV. RESULTS

Results of AVRORA and TOSSIM were described in further section.

A. *AVRORA*

To execute commands over AVRORA, first we need to build our application. When running of application completes it creates a .exe file. And then we need to .exe file to .od file. Now our application is ready to run.

1. To get the result of how much RAM and ROM is used we run the command

   cd $TOSROOT/apps/Application_Name

   make micaz

2. To get the simulation time, run the following command

   cd $TOSROOT/apps/Application_name

   make micaz

   cd build/micaz

   convert –avrora main.exe app1.od

   avrora –seconds=5.0 –platform=micaz –monitor=sleep app1.od

3. To get consumption of energy , follow the commands

   cd $TOSROOT/apps/Application_name

   make micaz

   cd build/micaz

   convert –avrora main.exe app1.od

   avrora –seconds=5.0 –platform=micaz –monitor=energy-profile app1.od

4. To get LED chart use command

   cd $TOSROOT/apps/Application_name

   make micaz

   cd build/micaz

   convert –avrora main.exe app1.od

   avrora –seconds=5.0 –platform=micaz –monitor=leds app1.od

5. To get interrupts occur during execution of application

   cd $TOSROOT/apps/Application_name

   make micaz

   cd build/micaz

   convert –avrora main.exe app1.od

   avrora –seconds=5.0 –platform=micaz –monitor=interrupts app1.od

6. To get the stack size

   cd $TOSROOT/apps/Application_name

   make micaz

   cd build/micaz

   convert –avrora main.exe app1.od

   avrora –seconds=5.0 –platform=micaz –monitor=stack app1.od

Table 3 shows the comparison between TinyOS application using these AVRORA commands.

B. *TOSSIM*

Steps to get results from TOSSIM are:

Step 1: Simulate the application by writing “make micaz sim” in the TinyOS terminal.

Step 2: Call the python interpreter (writing “python” in terminal). “>>>” shows we are in python interpreter.

Step 3: Write following commands in the interpreter:

>>>from TOSSIM import *

>>>a=Tossim ([])

>>>a.runNextEvent ()

false

(False shows LED is off) then type commands:

TABLE III. RESULTS OF AVRORA

s.no.	Application Parameters	Blink LED	Radio count to LED	Sense
1	RAM used	51 bytes	321 bytes	47 bytes
2	ROM used	2644 bytes	12510 bytes	2862 bytes
3	Simulated time(at 5.0 seconds)	**Total cycles** =36864000 cycles **Active** 8037209 cycles **Idle** =0 cycles(rest in power save=28826791)	**Total cycles** =36864000 cycles **Active** 8150970 cycles **Idle** =28713030 cycles	**Total cycles** =36864000 cycles **Active**=8037209 cycles **Idle** 0cycles(rest in power 826791)
4	Energy consumed by CPU	0.026195705 Joule	0.0641569897Joule	0.026196705 Joule
5	Analyze-stack	**Maximum stack pointer**=0x10FD **Minimum stack pointer**=0x10E2 **Maximum stack size**=27 bytes	**Maximum stack pointer**=0x10FD **Minimum stack pointer**=0x0FF7 **Maximum stack size**=262 bytes	**Maximum stack pointer**=0x10FD **Minimum stack pointer**=0x10E2 **Maximum stack size**=27 bytes

Step 4: Now add debug message to Boot.booted and if we again run then we get the debug messages like

>>>import sys

>>>a.addChannel ("Boot", Sys.stdout);

>>>a.runNextEvent ()

DEBUG (1): Application booted.

DEBUG(1):Application booted a third time.

Step 5: Add a noise trace file to the application and a code to generate the noise model as follows:

```
#! /usr/bin/python
from TOSSIM import *
import sys
n=open ("noise_sequence.txt", "r")
l=n.readlines ()
for line in l:
        p=line.strip ()
        in p:
                s=int (p)
                for i in range (7):
                a.getNode (i).addNoiseTraceReading(s)
for i in range (7):
        a.getNode (i).createNoiseModel ()
```

Step 6: Add topology so that nodes get to know in which manner they have to communicate via sending messages.

>>>b=a.getNode (1)

>>>b.bootAtTime (20000)

>>>a.runNextEvent ()

true

(True shows now the new node with ID 1 is on).

Fig. 4. Topology file (topo.txt)

Step 7: now after this run the test.py file and terminal shows sending and receiving of messages. Fig. 5 shows the same.

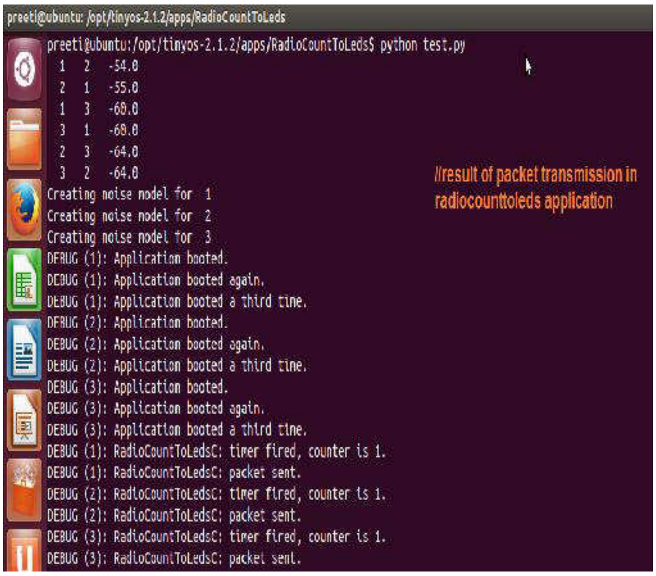


Fig. 5. Results of TOSSIM

V. CONCLUSION

TinyOS is an operating system designed for wireless embedded sensor network, the success rate of application is high as it works in sync with simulations and the real-time. Tools used to simulate the applications of TinyOS are TOSSIM, where TOSSIM is command line simulator. Analysis of TinyOS application is done by AVRORA. Future work is to deploy the Wireless sensor network protocol like LEACH, AODV and PEGASIS on TinyOS and simulate these protocols over it and analyze and compare any two protocols based on some parameters like – reliability, effectiveness, Scalability and security.

REFERENCES

[1] http://capsil.org/capsilwiki/index.php/MICAz accessed ,2014.
[2] www.ent.mrt.ac.lk/dialog/documents/ERU-2-wsn accessed ,2013.
[3] http://www.springerreference.com/docs/html/chapterdbid/63462.html accessed on 5/12/2013.
[4] P. Levis, D. Gay, "*TinyOS Programming*", http://csl.stanford.edu/~pal/pubs/tos-programming-web.pdf, July, 2009.
[5] P. Levis, S. Madden, et al., "TinyOS: An operating system for Sensor Networks," *Ambient intelligence, Springer*, pp. 115-148, 2005.

[6] R. Nath, "A TOSSIM based implementation and analysis of collection tree protocol in wireless sensor networks," *International conference on communication and signal processing, IEEE*, pp. 484-488, 2013.
[7] P. Levis, N. Lee, *"TOSSIM: A Simulator for TinyOS Network"*, http://www.cs.berkeley.edu/~pal/pubs/nido.pdf, version 1.0, June 26, 2003.
[8] http://compilers.cs.ucla.edu/avrora/,2013.
[9] http://www.techopedia.com/definition/25651/wireless-sensor-network-wsn accessed on 30/12/2013.
[10] http://www.cse.wustl.edu/~jain/cse567-11/ftp/sensor/index.html accessed on 25/1/2014.
[11] D. Gay, P. Levis, D. Culler , E. Brewer, *"nesC 1.3 Language Reference Manual*", http://tinyos.stanford.edu/tinyos-wiki/index.php/Main_Page, July, 2009.
[12] http://dc128.4shared.com/doc/aUnNeNPW/preview.html accessed on 7/1/2014.
[13] www.cs.virginia.edu/~cl7v/cs851-talks/tinyos_tian.ppt ,2013.

Measuring Cohesion

Rahul Kumar Pandey
Surguja University
Chhattisgarh, India
ravi.care18@gmail.com

***Abstract*--Cohesion is an important factor used in evaluating software design quality and modularity. The cohesion of a module refers to the relatedness of the module components. In software engineering, highly cohesive modules are highly desirable due to their high reusability and maintainability. Cohesion is classified according to levels. Being Cohesion is a prominent attribute of software services and is used as a major quality factor in service design. Cohesion is the degree of functional relevance of activities which are performed by a service to realize a business process. It shows how much an individual service is instrumental in performing one single task. High cohesion brings about ease of understanding of the design model and makes the system more agile. The design-level cohesion measures are formally defined, has been implemented, and can support software design, maintenance, and restructuring. This paper discusses metrics to measure cohesion at design level along with some issues related to it.**

Keywords—Cohesion, software measurement and metrics, software design, software maintenance,.

I. INTRODUCTION

Module cohesion was defined by Yourdon and Constantine as "how tightly bound or related its internal elements are to one another" . They describe cohesion as an attribute of designs, rather than code, and an attribute that can be used to predict properties of implementations such as "ease of debugging, ease of maintenance, and ease of modification.

We follow two approaches that have been used to develop objective, automatable methods for measuring module cohesion. The first approach, an *association-based* approach, is used by Lakhotia to formalize the notion of the associations between processing elements as a set of rules concerning data dependencies in module code. Lakhotia's methods require the analysis of Code-level information; they can be adapted for use on design-level constructs. The second approach, a *slice-based* approach, is used by Bieman and Ott They measure functional Cohesion in terms of the connections between code data tokens on module output slices. The Computation of functional cohesion also requires code level information class cohesion measures for object-oriented software have also been defined using a slice-based approach, and by analyzing the connectivity between methods through common references to instance variables Method bodies are needed to apply these code-level class cohesion measures.

Cohesion is only one of many attributes that designers strive to improve. For example, designers also aim to reduce coupling. We need to be able to objectively measure such attributes in order to evaluate alternative design options.

Some design goals may be in conflict—high cohesion with low coupling appears to be competing goals. In this paper, we show that module cohesion can be objectively assessed using only design-level information. We develop and compare a set of association-based and slice base design-level cohesion measures, and we describe how these measures can be applied as design. Along with some issues related with it.

II. ASSOCIATION-BASED COHESION MEASURES

Stevens, Myers, and Constantine define module cohesion (SMC Cohesion) on an ordinal scale. SMC Cohesion includes *coincidental*, *logical*, *temporal*, *procedural*, *communicational*, *sequential*, and *functional* cohesion where coincidental cohesion is the weakest and functional cohesion is strongest cohesion. SMC Cohesion is determined by inspecting the association between all pairs of a module's processing elements. A *variable dependence graph* models the control and data dependencies between module variables. The rules for designating a cohesion level are defined using a strict interpretation of the association principles of SMC Cohesion. Because the rules are formal, a tool can automatically perform the classification. However, the technique, as originally defined, can be applied only after the coding stage since it is defined upon the implementation.

A. A Design-Level View of a Module

The input/output dependence graphs (IODG), adapted from the variable dependence graph of Lakhotia [9], is based on the data and control dependence relationships between input/output components of a module. Input components of a module include in-parameters and referenced global variables. Output components include out-parameters, modified global variables, and 'functions return' values. The term 'component' refers to a static entity. An array, a linked list, a record, or a file is one component rather than a group of components. We define data and control dependence informally using the notation of Lakhotia.

i. DEFINITIONS

- A variable *y* has a *data dependence* on another variable *x* (*x y*) *d* ® if *x* 'reaches' *y* through a path consisting of a 'definition-use' and 'use-definition' chain from Lakhotia

- A variable *y* has a *control dependence* on another variable *x* if the value of *x* determines whether or not the statement containing *y* will be performed (from Lakhotia.
- A variable *y* is *dependent* on another variable *x* (*x* *y*) when there is a path from *x* to *y* through a sequence of data or control dependence. We call the path a *dependence path*.
- A variable *y* has *condition-control dependence* on another variable *x* (*x* *y*) if *y* has a control dependence on *x*, and *x* is used in the predicate of a decision (i.e., if-then-else) structure.
- A variable *y* has *iteration-control dependence* on another variable *x* (*x* *y*) if *y* has a control dependence on *x*, and *x* is used in the predicate of an iteration structure.
- A variable *y* has *c-control dependence* on another variable *x* (*x* ,*y*) if the dependence path between *x* and *y* Contains condition-control dependence but no iteration control dependence.
- A variable *y* has *it-control dependence* on another variable *x*(*x* *y*) *i* if the dependence path between *x* and *y* contains an iteration-control dependence.

ii. IODG DEFINITION

The *input/output dependence graph* (IODG) of a module *M* is a directed graph, *GM = (V, E)* where V is a set of input/output components of *M*, and *E* is a set of edges labeled with dependence types such that *E* = {(*x*, *y*) € *V* ×*V* | y *has data*, c-*control*, *and/or* i-*control dependence on* x}. The IODG shows the relationship between input and output components of a module. Each input contributes to one or more outputs; it is used to compute output(s), as input data, decision invariant, and/or loop invariant. The IODG is used to define a design-level cohesion measure.

B. Design-Level Cohesion (DLC) Measure

In a manner similar to the approach used to develop SMC Cohesion, we use six relations between a pair of output components based on the IODG representation:

i. *Coincidental relation (R1):*

*R*1(*o*1, *o*2) = *o*1 *o*2 ^¬ (*o*1 *o*2) ^¬(*o*2 *o*1) ¬□ □ *x* [(*x* *o*1) (*x* *o*2)] Two outputs *o*1 and *o*2 of a module have neither dependence relationship with each other, nor dependence on a common input.

ii. Conditional relation (R2):

*R*2(*o*1, *o*2) = *o*1 *o*2 □ *x* [(x^c *o*1] [x^c *o*2] Two outputs are c-control dependent on a common input.

iii. Iterative relation (R3):

*R*3(*o*1, *o*2) = *o*1 *o*2 □ *x* [(x^i *o*1] [x^i *o*2] Two outputs are i-control dependent on a common input.

iv. Communicational relation (R4):

*R*4(*o*1, *o*2) = *o*1 *o*2 □ *x* [(x^d *o*1) (x^d *o*2)] v ((x^p *o*1) (x^p *o*2)], where *p*, *q* □{*d*, *c*, *i*}, and *p* *q*. Two outputs are dependent on a common input. An input is used to compute both outputs, but as neither a condition flag to select one of two outputs nor a loop invariant to compute both outputs.

v. Sequential relation (R5):

*R*5(*o*1,*o*2) = *o*1 *o*2 ((*o*1 *o*2) (*o*2 *o*1))One output is dependent on the other output.

vi. Functional relation (R6):

*R*6(*o*1,*o*2) = (*o*1 =*o*2)

There is only one output in a module. Our relations are derived from Lakhotia; relations 1, 4, and 5 are essentially identical to the corresponding relations of Lakhotia. The remaining relations are defined to fit the IODG model. Cohesion strength increases from relation *R*1 to *R*6. The six relations correspond to six association principles (temporal cohesion is not included) of SMC Cohesion with some degree of overlap.

C. DLC MEASURE DEFINITION

The cohesion level of a module is determined by the relation levels of output pairs. For each pair of outputs, the strongest relation for that pair is used. The cohesion level of the module is the weakest (lowest level) of all of the pairs. That is, the output pair with the weakest cohesion determines the cohesion of the module.

We have shown that the DLC measure is on an ordinal scale as long as we accept the ordering implied by the association principles of SMC Cohesion. An IODG can be displayed visually in an IODG diagram. In its graphical form, the IODG visually displays the functional structure of the module. In such a diagram, the caller-callee relationship is represented by including the IODG of the callee in the IODG diagram of the caller. In an IODG diagram, an input is represented by a circle, and an output by a square. The texts in each circle and square are the names of input and output variables. Each arrow indicates the dependence between two components.

III. SLICE-BASED COHESION MEASURES

A program *slice* is the portion of the program that might affect the value of a particular identifier at a specified point in the program . In developing cohesion measures, slices can be used to represent the functional components of a module.

A. Functional Cohesion (FC) Measures

Bieman and Ott developed cohesion measures that indicate the extent to which a module approaches the ideal of functional cohesion. They introduced three measures of functional cohesion as the relative number of "glue" or “adhesive” data tokens based on “data slices” of a module (procedure). The *data slice* of a variable is the sequence of data tokens which have a dependence relationship with the variable. The three measures of functional cohesion are *Weak Functional Cohesion* (*WFC*), *Strong Functional Cohesion* (*SFC*),and *Adhesiveness* (*A*). WFC is the ratio of glue tokens to the total number of tokens in a procedure. SFC is the ratio of superglue tokens to the total number of data tokens in a procedure. Adhesiveness the ratio of the amount of adhesiveness to the total possible adhesiveness, which is the adhesiveness when all data tokens are superglue tokens

B. Design-Level Functional Cohesion (DFC)

i. Measures

We derive DFC measures following the approach used to develop the functional cohesion measures. Rather than Analyzing code details, we use a design level view modeled by the IODG to define the measure. The DFC measures use a ‘simplified’ IODG which includes only dependence relationships between input/output components, without classifying the dependencies. The DFC measures are analogous to the slice-based FC measures in that both are defined in terms of the connections between components and outputs. The components used to define the FC measures include all “data tokens,” while only input and output components are used to define the DFC measures. Inputs and outputs are the only externally visible components, and they represent the design-level information of the module. Fig. 3a shows a graphical (IODG), and Fig. 3b shows a tabular (IODT) representation of procedure *Sum_Max_Avg* of Fig. 2.In Fig. 3b, the names of the output are listed in the first row and the names of the components (inputs and outputs) are in the first column of the figure. The “1” in the figure indicates that the corresponding component has a dependence relation with the named output, and the “0” indicates no dependence relation. The IODG and IODT show the relationship between input/ output components of a module. The DFC measures are expressed in terms of the number of *isolated* and *essential* components, and the *connectedness* of components:

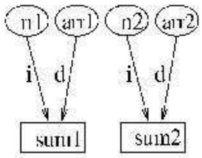
a). DEFINITION.

A component is isolated *if it affects only one local functionality, i.e., it has a dependence relationship with only one output.* For example, in Fig. 3, component ‘max’ is isolated since it has a dependence relationship with only one output, itself. The other components are not isolated.

b). DEFINITION.

A component is essential *if it affects (or is affected by) all functionalities of the module, i.e., it has dependence relationships with all outputs of the module.* If a module contains only one output, the output is the Only functionality of the module. Thus, every component in the module is not isolated and is essential. In Fig. 3, components ‘n’ and ‘arr’ are essential since they affect all outputs.

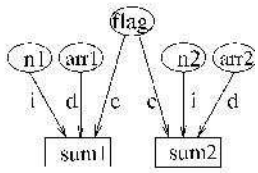
```
procedure Sum1_and_Sum2
  ( n1, n2 : integer;
    arr1, arr2 : int_array;
    var sum1,
        sum2 : integer );
var i : integer;
begin
   sum1 := 0;
   sum2 := 0;
    for i := 1 to n1 do
       sum1 := sum1 + arr1[i];
    for i := 1 to n2 do
       sum2 := sum2 + arr2[i];
end;
```



Coincidental cohesion

(A)

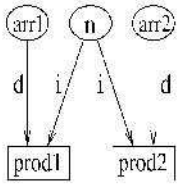
```
procedure Sum1_or_Sum2
  ( n1, n2, flag : integer;
    arr1, arr2 : int_array;
    var sum1,
        sum2 : integer );
var i : integer;
begin
   sum1 := 0;
   sum2 := 0;
   if flag = 1
     for i := 1 to n1 do
        sum1 := sum1 + arr1[i];
   else
     for i := 1 to n2 do
        sum2 := sum2 + arr2[i];
end;
```



Conditional cohesion

(B)

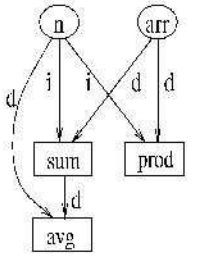
```
procedure Prod1_and_Prod2
  ( n : integer;
    arr1, arr2 : int_array;
    var prod1,
        prod2 : integer );
var i : integer;
begin
   prod1 := 1;
   prod2 := 1;
   for i := 1 to n do begin
      prod1 := prod1 * arr1[i];
      prod2 := prod2 * arr2[i];
   end;
end;
```



Iterative cohesion

(C)

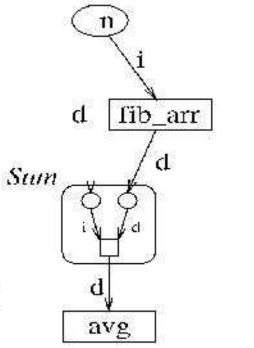
```
procedure Sum_and_Prod
  ( n : integer;
    arr : int_array;
    var sum,
        prod : integer;
    var avg : float );
var i : integer;
begin
   sum := 0;
   prod := 1;
   for i := 1 to n do begin
      sum := sum + arr[i];
      prod := prod * arr[i];
   end;
   avg := sum / n;
end;
```



Communicational cohesion

(D)

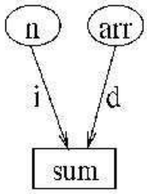
```
procedure Fibo_Avg
   ( n : integer;
     var fib_arr : int_array;
     var avg : float );
var sum : integer;
    i : integer;
begin
   fib_arr[1] := 1;
   fib_arr[2] := 2;
   for i := 3 to n
      fib_arr[i] := fib_arr[i-1]
             + fib_arr[i-2];
   Sum(n, fib_arr, sum);
   avg := sum / n;
end;
```



Sequential cohesion

(E).

```
procedure Sum
  ( n : integer;
    arr : int_array;
    var sum : integer );
var i : integer;
begin
   sum := 0;
   for i := 1 to n do
      sum := sum + arr[i];
end;
```



Functional cohesion

(F)

TABLE: 1. Shows statements of DFC

sum	max	avg	statement
			procedure Sum_Max_Avg
1	1	1	(n : integer; /* pre: n>0 */
1	1	1	var arr: int_array;
1		1	var sum,
	1		max : integer;
1		1	var avg : float);
1	1	1	i : integer;
			begin
2		2	sum := 0;
	3		max := arr[1];
3	3	3	for i := 1 to n do begin
4		4	sum := sum + arr[i];
	3		if arr[i] > max
	3		max := arr[i];
			end;
3		3	avg := sum / n;
			end;

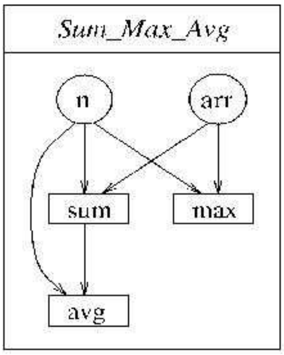
SMC Cohesion: Communicational cohesion

FC Measures:

WFC = 17/27 = 0.63

A = (11 * 2 + 6 * 3)/(27 * 3) = 0.49

SFC = 6/27 = 0.22



component \ output	sum	max	avg
n	1	1	1
arr	1	1	1
sum	1	0	1
max	0	1	0
avg	1	0	1

(A) *(B)*

Fig. 3. An example. (a) the IODG; (b) IODT of a procedure Sum_Max_Avg.

We define the *connectedness* of a component as the degree of "relatedness" of the component to the outputs. Connectedness provides more information than a simple classification of a component as isolated or essential. The connectedness of a component represents the relative number of outputs that the component relates together. We do not address the cases where an input does not contribute to the computation of any output, i.e., in our model, every component has a dependence relation with at least one output. Therefore, the connectedness of a component is the relative number of the other output(s) with which the component has a dependence relation. If a module contains only one output, the connectedness of every component in the module is 1 .

c). DEFINITION

For an arbitrary module, the connectedness of the i*th component is:*

$$Ci=\{ \frac{Ni-1}{O-1} \text{ if } o>1$$
1

where Ni is the number of outputs with which the i*th component has a dependence relation, and O is the total number of outputs in the IODG model of the module*

The connectedness of an isolated component is 0 and the connectedness of an essential one is 1. In Fig. 3b, the connectedness of *n* and *arr* is 1, the connectedness of *sum* and *avg* is 1/2, and the connectedness of *max* is 0. Three measures, *Loose Cohesiveness* (LC), *Tight Cohesiveness* (TC), and *Module Cohesiveness* (MC) are defined as the relative number of non isolated components, the relative number of essential components, and the average connectedness of the components of the model, respectively

C). DFC MEASURE DEFINITION

LC(m) = D/T

TC(m)= D/T

$$MC(m)=\frac{\sum_{i=1} Ci}{T}$$

where D, E, and Ci are the number of nonisolated components, the number of essential components, and the connectedness of i*th component, respectively, in the IODG of module m. T is the total number of components in m.*

Using the definition of component connectedness, module cohesiveness can be expressed as:

$$MC(m)=\frac{\sum_{i=1} (N-1)}{T*(o-1)}=\frac{\sum_{i=1} (N \quad -T)}{T*(0-T)}$$

The three measures for the procedure *Sum_Max_Avg* in: Fig. 2 and Fig. 3 are:

LC (*Sum_Max_Avg*) = 4/5 = 0.8

TC (*Sum_Max_Avg*) = 2/5 = 0.4

MC (*Sum_Max_Avg*) = (2 * 2 + 2 * 1)/(5 * 2)=0.6

An isolated component has zero connectedness, a nonisolated component has connectedness of greater than 0, and essential component has connectedness of one. Thus, for a given module *m*:

$$E<\sum_{I=1}^{T} ic<D$$

IV. CONCLUSION

I have formalized the concept of design cohesion based on a graph model of a procedure interface, the input/output dependence graph (IODG). I derived a design-level cohesion (DLC) measure using an association-based approach similar to that used by Stevens et al. , and design-level functional cohesion (DFC) measures using the slice-based approach used to derive code-level functional cohesion (FC) measures .

REFERENCES

[1] J. Bieman and B.-K. Kang, Measuring Design-Level Cohesion

[2] J. Bieman and B.-K. Kang, "Cohesion and Reuse in an Object-Oriented System," Proc. ACM Symp. Software Reusability, SSR'95, pp. 259–262, Apr. 1995. Reprinted in ACM Software Eng. Notes, Aug. 1995.

[3] J. Bieman and L. Ott, "Measuring Functional Cohesion," IEEE Trans. Software Eng., vol. 20, no. 8

[4] L. Briand, S. Morasca, and V. Basili, "Measuring and Assessing Maintainability at the End of High Level Design," Proc. Conf. Software Maintenance, pp. 88–95,.

[5] S. Chidamber and C. Kemerer, "A Metrics Suite for Object Oriented Design," IEEE Trans. Software Eng., vol. 20, no. 6

[6] T.J. Emerson, "A Discriminant Metric for Module Cohesion," Proc. Seventh Int'l Conf. Software Eng., ICSE-7

[7] N. Fenton, "Software Measurement: A Necessary Scientific Basis," IEEE Trans. Software Eng., vol. 20, no. 3

[8] B.-K. Kang and J. Bieman, "Using Design Cohesion to Visualize,Quantify, and Restructure Software," Eighth Int'l Conf. Software Eng. and Knowledge Eng., SEKE '

Evaluating the Performance of Diabetes using LAD Tree, J48 and Random Forest Data Mining Algorithms

Gaganjot Kaur
Department Of Computer Science and Engineering
GNDU, Amritsar (Pb.) India
gaganjot1409@gmail.com

Amit Chhabra
Department Of Computer Science and Engineering
GNDU, Amritsar (Pb.) India
amit.cse@gndu.ac.in

***Abstract*—This paper discusses the performance analysis of different data mining algorithms for efficiently mining the diabetes data set. Data mining algorithms are extremely popular in real time applications like intrusion detection system, diabetes mining, e-mail spam classification etc. The overall objective of this research work is to evaluate the performance of the J48, LAD tree and Random Forest data mining algorithms by considering the diabetes data set. WEKA machine learning tool is used to implement the proposed work. It has been found that that the LAD-Tree outperforms over the J48 and Random Forest based machine learning algorithms.**

Keywords- LAD tree, J48,Random Forest, Diabetes, WEKA

I. INTRODUCTION

Data mining is the conversion of huge volumes of data into expressive shapes and rules. It is about explaining the former and calculating the future by means of data analysis. Data mining is a multi-disciplinary field which combines, machine learning, statistics, database technology and artificial intelligence. Data Mining is accomplished in a number of phases: Business understanding, Data understanding, Data preparation, Modelling, Evaluation, and Deployment. There are several data mining techniques as Association, Classification, Clustering, Neural Network and Regression.

A. Diabetes

Diabetes is a common disease that costs a great deal of money and its incidence is rising. It is a disease in which the body does not yield or properly use insulin, the hormone that "unlocks" the cells of the body, allowing glucose to arrive and fuel them. Diabetes increases the risks of developing blood vessel harm, blindness, heart disease, nerve damage and kidney disease [3]. Diabetes occurs in two major: type 1(insulin dependent diabetes) and type 2(non-insulin-dependent diabetes).

Diabetes is a group of metabolic diseases categorized by high blood sugar (glucose) levels that outcomes from defects in insulin secretion, or its action, or both. Production of defective insulin, insufficient production of insulin or the inability of cells to use insulin properly and efficiently leads to hyper glycaemia and type-2 diabetes. The absolute lack of insulin, generally secondary to a destructive process distressing the insulin-producing beta cells in the pancreas, is the main syndrome in type 1 diabetes. In type 2 diabetes, there also is a solid decline of beta cells that enhances process of elevated blood sugars.

II. DATA MINING TECHNIQUES

Classification is the process of instinctively building a model of classes from a set of records that comprise class labels. A Decision Tree model contains rules to predict the target variable. The Tree Classification algorithm offers an easy-to-understand explanation of the essential distribution of the data [5].

A. J48

J48 is an extension of ID3 which accounts for missing values, pruning of decision trees, continuous attribute value ranges, derivation of rules, etc. J48 is an open source Java implementation of the C4.5 algorithm in the Weka data mining tool. It gives numerous options associated with tree pruning. Pruning can be used as a tool to precise for potential overfitting. The other algorithms recursively classifies till every single leaf is pure, that the data has been characterised as close to perfectly as possible. This process certifies maximum accuracy on the training data, but it may create extreme rules that only defines particular identity of that data.The aim is to progressively generalize a decision tree until it gains an equilibrium of flexibility and accuracy.

Features of algorithm:

(i) It handles both discrete and continuous attributes. For handling continuous attributes, C4.5 sets a threshold and then divides the list into those whose attribute value is below the threshold and those that are more than or equal to it.

(ii) It handles training data with missing attribute values.

(iii) It prunes trees after construction. C4.5 drives back through the tree once it's been formed and challenges to eradicate branches that do not help by exchanging them with leaf nodes.

B. Random Forest

It is a collaboration of un-pruned classification or regression trees, persuaded from samples of the training data, using random feature selection in the tree induction process. Prediction is made by collecting the predictions of the ensemble. Random forest usually reveals a considerable performance improvement over the single tree classifier.

On the other hand, alike most classifiers, RF can also suffer from the curse of learning from an extremely imbalanced training data set. As it is created to minimize the overall error rate, it will tend to concentrate more on the prediction accuracy of the majority class, which frequently results in poor accuracy for the minority class. Random Forests is best fit for the analysis of complex data structures implanted in small to moderate data sets containing less than 10,000 rows but potentially millions of columns.

C. LAD Tree

Logical Analysis of Data is a technique for classification suggested in optimization literature. It forms a classifier for binary target variable centred on learning a logical expression that can differentiate between positive and negative samples in a data set. The basic supposition of LAD model is that a binary point covered by some positive patterns, but not covered by any negative pattern is positive, and similarly, a binary point covered by some negative patterns, but not covered by positive pattern is negative. The building of LAD model for a given data set generally involves the generation of large set patterns and the collection of a subset of them that achieves the above supposition such that each pattern in the model satisfies certain requirements in terms of occurrence and similarity.

III. RELATED WORK

In [1] Al Jarullah A.A (2011) has studied decision tree method that was used to calculate patients with developing diabetes. The study contains two phases. The first phase is data pre-processing which includes attribute identification and selection, numerical discretization and handling missing values. The second phase is a diabetes prediction model structured using the decision tree method. WEKA software was used during all the phases.

In [2] Folorunsho Olaiya (2013) modelled data from diabetes patients and used it for calculating the diabetes probability of a patient. The performances of both Decision Tree Algorithms and Artificial Neural Network on medical data were computed based on kappa statistics, mean absolute error, relative squared error, time to model and mean-squared error. The method to collect the data about diabetes record could be utilized in the predicting the disease of the new patient. The disease can be predicted, with certain accuracy if the data related to some important aspects of the patient health record is present.

In [3] Huang Feixiang et al. (2012) discussed the data mining process for predicting hypertension from patient medical records having eight other diseases. WEKA has been used to generate the J-48 and Naive Bayesian classifiers with overall accuracy around 83%..An ensemble of five J-48 classifiers was formed trying to increase the prediction performance. The resultant classifier signifies a second-order approximation of the relation among the diseases. Experimental results exhibited a slight improvement of the ensemble approach over pure Naive Bayesian and J-48 in sensitivity, accuracy and F-measure using the ensemble, and rough set tools were able to decrease the ensemble of five members to three. However, there is substantial growth of sensitivity.

In [4] Marcano-Cedeno Alexis and Andina Diego Alexis (2012) have applied the artificial meta plasticity on multilayer perceptron as a data mining technique for the diabetes disease analysis. The results attained by artificial meta plasticity on multilayer perceptron were compared with Bayesian classifier (BC), decision tree (DT) using same database. For validation methods the algorithm artificial meta plasticity on decision tree, Bayesian classifier and multilayer perceptron were trained 50 times, respectively, to form classification model and compute classification efficiency. The decision tree was used J48 algorithm with different confidence factor: 0.25, 0.4 and 0.6.The Bayesian classifier was used with and without kernel estimation By equating standard deviation, with the more stable classification efficiency, decision tree is the best classification algorithm.

In [6] Nincevic et al. (2010) classifies the influence of various variables on life expectancy at birth. To define and present key variables, demographic data mining was implemented based on accessible worldwide data from the period from 1990 to 2006. Extraction –transformation – loading (ETL) processes that preceded data mining are also presented. It has been defined that two main methodologies govern data mining: Cross Industry Standard Process and Sample, Explore, Modify, Model, Assess (SEMMA).From other mining methods, decision tree method is chosen as the optimal for the problem.

In [7] Robu and Hora (2012) have discussed the examination of medical data through data mining techniques. The purpose is to improve disease diagnosis, to more efficiently stop them and treat them more easily. The construction of a classifier model is done with algorithms such as J48, Random Forest, Naive Bayes etc. As the models attained with these algorithms led to good accuracy, they were used to make predictions. The extended application was tested on four medical datasets. The constructed models were tested through cross validation. Later, the predictions were tested with the constructed models and with the help of the established dynamic interface. The application worked properly in each considered case, the constructed models could be successfully used by the dynamic interface for making predictions.

In [8] Salama et al. (2012) presents a comparison among the different classifiers Naive Bayes, decision tree (J48), Sequential Minimal Optimization (SMO), Instance Based for K-Nearest neighbor (IBK) and Multi-Layer Perception on three different databases of breast cancer (Wisconsin Breast Cancer, Wisconsin Diagnosis Breast Cancer and Wisconsin Prognosis Breast Cancer) by using confusion matrix and classification accuracy based on 10-fold cross validation method. A combination at classification level is accomplished between these classifiers to get the best multi-classifier approach and accuracy for each data set.

In [10] UM Ashwinkumar and KR Anandakumar (2012) finds the symptoms of diabetes and cardiac using data mining technology by using the method of Decision Tree and Incremental Learning at the early stage. A novel learning algorithm i+Learning as well as i+LRA has been defined, which ostensibly achieves the highest classification accuracy over ID3 algorithm. The solid evidence demonstrates that both i+Learning and i+LRA performs better than the other incremental learning algorithms not only on the classification accuracy, but also be capable to handle the incremental learning about the new incoming attribute other than the new instance only without sacrificing the learning performance. There is no perfect algorithm, which is also true to i+Learning. The main limitation of this method is the adoption of binary tree rather than multi-branch tree.

In [13] Hussein et al. (2012) proposed a Chronic Disease Diagnosis Recommender System approach based on a hybrid method using unified Collaborative Filtering and multiple classifications. The Unified CF method based on learning classification model using both historical binary ratings and external features will be used for attaining higher recommendation accuracy when recommending medical advices for patients. Describing the relevant features through Attribute Selection method is used to decrease data generation and increase the predictive model performance. Diabetes diagnosis case study is designed as experiments to show the feasibility of the model. Multiple decision tree classifiers have been applied through the first stage of CDD system to attain high accuracy predictive model.

In [12] Takada et al. (2013) have equated the performance of ADTree and MLR models in terms of robustness against missing values. Ensembled ADTree models are more robust against missing values than MLR. Sufficient robustness is achieved at low boosting and ensemble number, and is compromised as these numbers increase.

In [11] Ramani R et al. (2012) proposed a novel computational approach for automatic disease detection in retinal image analysis and data mining techniques to accurately categorize the retinal images as Normal, Diabetic Retinopathy and Glaucoma affected. Three feature relevance and sixteen classification Algorithms were evaluated and used to detect the contributing features that gave better prediction results. His results prove that random tree and C4.5 classification techniques generate the maximum multi-class categorization training accuracy of 100% in classifying 45 images from the Gold Standard Database.

IV. EXPERIMENT SETUP

WEKA is an innovatory system in the history of the data mining and machine learning research communities. The WEKA tool was developed by WEKA team by putting a remarkable effort since 1994.WEKA offers many different algorithms for data mining and machine learning. It is open source and freely available platform-independent software. It offers flexible facilities for scripting experiments and is easily useable by people who are not data mining specialist.

The steps performed for data mining in WEKA are:

(i) Data pre-processing and visualization

(ii) Attribute selection

(iii) Classification (Decision trees)

(iv) Prediction (Nearest neighbour)

(v) Model evaluation

(vi) Clustering (Cobweb, K-means)

(vii) Association rules

The data for the experiments is taken from the Pima Indians Diabetes dataset [14]. It consists of 768 instances and 8 attributes are chosen from huge number of attributes. The attributes are: Number of times pregnant, Plasma glucose concentration a 2 hours in an oral glucose tolerance test, Diastolic blood pressure, Triceps skin fold thickness ,2-Hour serum insulin, Body mass index, Diabetes pedigree function, Age, Class variable. The class variable has two possible values which are 0 and 1.0 indicates that the person is not having diabetes and 1 indicates that the person is diabetic.

Fig 1 has shown the various steps required to accomplish this research work. Each step has been designed in WEKA knowledge flow. The main objective behind this setup is to evaluate and compare the performance of the LAD tree, J48 and Random forest to find the best algorithm among these.

V. PERFORMANCE EVALUATION

This section has shown the comparison of the J48, LAD tree and random forest data mining algorithms.

The accuracy is calculated as:

(i) $$\mathrm{TA} = \frac{(\mathrm{TP+TN})}{\mathrm{TP+TN+FP+FN}} \quad (1)$$

(ii) $$\mathrm{RA} = \frac{(\mathrm{TP+FP})*(\mathrm{TN+FN})+(\mathrm{FN+TP})*(\mathrm{FP+TP})}{(\mathrm{Total*Total})} \quad (2)$$

The kappa statics is calculated as:

$$\text{(iii)} \qquad KS = \frac{TA-RA}{1-RA} \qquad (3)$$

In the equation (i) TA represents Total Accuracy, TP is True Positive, TN is True Negative, FP is False Positive and FN is False Negative. In equation (ii) RA represents Random Accuracy. In equation (iii) KS is Kappa Statistics

TABLE 1: COMPARISON BETWEEN LAD TREE, J48 AND RANDOM FOREST

Algorithms	Accuracy	Execution time	Kappa Statics
J48	73.8281	0.02	0.4164
Random Forest	73.9583	0.27	0.4052
LAD Tree	74.0885	0.28	0.415

Fig 2 has shown that the LAD tree has shown quite significant results than J48 and Random forest. As the line graph of the LAD tree is always more than the J48 and Random forest. In figure 2 X-axis has shown the false positive rate and y-axis has shown true positive rate. .

VI. CONCLUSION AND FUTURE WORK

This work has analysed different data mining algorithms namely J48, LAD tree and random forest tree for effectively mine the diabetes data set.

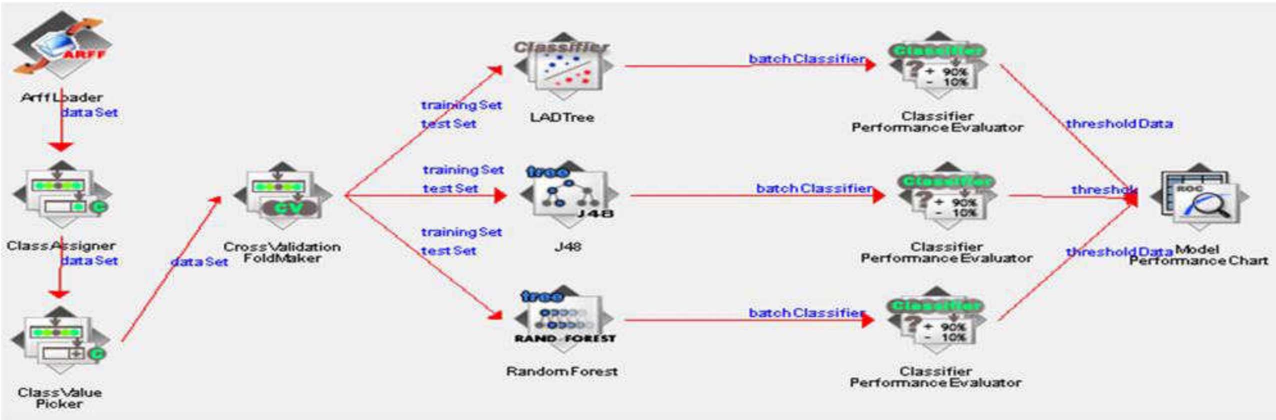


Fig. 1 Proposed Scenario

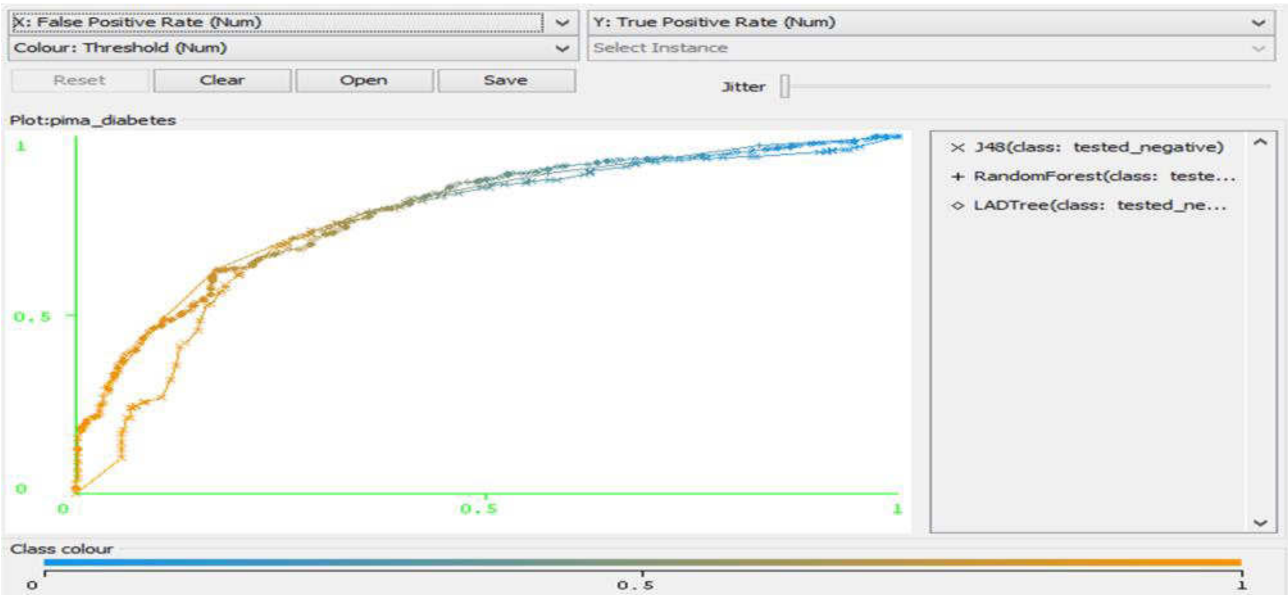


Fig. 2 Comparison of J48, LAD tree and Random Forest data mining algorithms

To accomplish the objectives of this work; WEKA machine learning tool is used. WEKA knowledge flow environment has been used to successfully implement the data mining learning tool. Prima Indians diabetes data set [14] is used to justify the proposed work. The experimental results have shown that the LAD-Tree performs better than the J48 and random forest based machine learning algorithms.

In future we will extend this work by modifying the J48 machine learning algorithm to enhance the decisions of J48.

REFERENCES

[1] Al Jarullah, A.A., "Decision tree discovery for the diagnosis of type II diabetes," Innovations in Information Technology (IIT), 2011 International Conference on , vol., no., pp.303,307, 25-27 April 2011

[2] Folorunsho, Olaiya. "Comparative Study of Different Data Mining Techniques Performance in knowledge Discovery from Medical Database." International Journal 3, no. 3 (2013).

[3] Huang, Feixiang; Wang, Shengyong; Chan, Chien-Chung, "Predicting disease by using data mining based on healthcare information system," Granular Computing (GrC), 2012 IEEE International Conference on , vol., no., pp.191,194, 11-13 Aug. 2012

[4] Marcano-Cedeno, Alexis; Andina, Diego, "Data mining for the diagnosis of type 2 diabetes," World Automation Congress (WAC), 2012 , vol., no., pp.1,6, 24-28 June 2012.

[5] Nadali, A; Kakhky, E.N.; Nosratabadi, H.E., "Evaluating the success level of data mining projects based on CRISP-DM methodology by a Fuzzy expert system," Electronics Computer Technology (ICECT), 2011 3rd International Conference on , vol.6, no., pp.161,165, 8-10 April 2011

[6] Nincevic, I.; Cukusic, M.; Garaca, Z., "Mining demographic data with decision trees," MIPRO, 2010 Proceedings of the 33rd International Convention , vol., no., pp.1288,1293, 24-28 May 2010

[7] Robu, R.; Hora, C., "Medical data mining with extended WEKA," Intelligent Engineering Systems (INES), 2012 IEEE 16th International Conference on , vol., no., pp.347,350, 13-15 June 2012

[8] Salama, G.I.; Abdelhalim, M.B.; Zeid, M.A., "Experimental comparison of classifiers for breast cancer diagnosis," Computer Engineering & Systems (ICCES), 2012 Seventh International Conference on , vol., no., pp.180,185, 27-29 Nov. 2012

[9] S. Moertini Veronica ,"Towards The Use Of C4.5 Algorithm For Classifying Banking Dataset",Integeral Vol 8 No 2,October 2013.

[10] UM, Ashwinkumar, and Anandakumar KR. "Predicting Early Detection of Cardiac and Diabetes Symptoms using Data Mining Techniques.",IEEE,pp:161-165,2011

[11] Geetha Ramani R, Lakshmi Balasubramanian, and Shomona Gracia Jacob. "Automatic prediction of Diabetic Retinopathy and Glaucoma through retinal image analysis and data mining techniques." In Machine Vision and Image Processing (MVIP), 2012 International Conference on, pp. 149-152. IEEE, 2012

[12] Sugimoto, Masahiro, Masahiro Takada and Masakazu Toi. "Comparison of robustness against missing values of alternative decision tree and multiple logistic regression for predicting clinical data in primary breast cancer." In Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE, pp. 3054-3057. IEEE, 2013.

[13] Hussein Asmaa S,Wail M. Omar, Xue Li, and Modafar Ati. "Efficient Chronic Disease Diagnosis prediction and recommendation system." In Biomedical Engineering and Sciences (IECBES), 2012 IEEE EMBS Conference on, pp. 209-214. IEEE, 2012.

[14] Bache, K. & Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

Analysis of Performance of Component Based System Architecture

Pahul Preet Kaur
Dept. of Computer Science and Engineering
Guru Nanak Dev University
Amritsar, Punjab
pahulsaini@gmail.com

Hardeep Singh
Dept. of Computer Science and Engineering
Guru Nanak Dev University
Amritsar, Punjab
hardeep_gndu@rediff.com

***Abstract*—Analysis of component system architecture is important to recognize potential risks of software design before implementation. With this we can achieve reduced costs and improved software quality. Performance analysis at this level can be useful for assessing whether a proposed architecture can meet the desired performance specifications and can help in making key architectural decisions. This paper presents various approaches for performance evaluation of component based systems. The analysis starts from the component base system, its architecture and then evaluating the performance. Approaches have been used to analyze performance depending upon the architecture (layered, distributed, client server), performance parameters (response time, throughput, utilization, scalability, etc) and the tool one can use to produce results of parameters over it.**

Keywords-component, component based architecture, architecture analysis, performance, parameters.

I. INTRODUCTION

In large, complex software systems the design of the overall system structure is a central problem. The software architecture of a system consists of software components, the externally visible properties of those components, and the interconnections among them. [8]

A. Component and its Interface

A Component is a software object, meant to interact with other components, encapsulating certain functionality or a set of functionalities. A component has a clearly defined interface and conforms to a prescribed behaviour common to all components within an architecture. [10] A connector/Interface encapsulates the interconnection protocol between two or more components. The description of an interface must contain information about all the viewpoints among, for example functionality, behaviour, protocols, safety, reliability, real-time, power, bandwidth, memory consumption and communication mechanisms, that are needed for composing the component in the given architecture for the application of the system.[8]

B. Parameters

The performance parameters for a component deal with two types of interface, the port and the entry. A port supports one-way communication. A message is sent on an output port and received on an input port. Message communication may be asynchronous or synchronous without response. An entry supports two-way message communication, i.e., synchronous message communication with response.

The type of connector used to interconnect components influences the performance of component based systems. The following performance parameters are important when describing connectors:

- Client and server stub times. This is the time needed by the client and server stubs to carry out tasks such as parameter marshalling, parameter conversion, and message generation
- Number of concurrent server threads
- Maximum number of connections to be accepted by the server

The following parameters are important when specifying a network resource:

- Bandwidth: defines the rate at which bits are transmitted over the network
- Latency: time it takes a bit to propagate across the network.

Fig. 1. Shows architecture of component based system. The user may interact with a component and mentioned component interacts with others if needed. Job can be completed in each component and the response is given back to user.

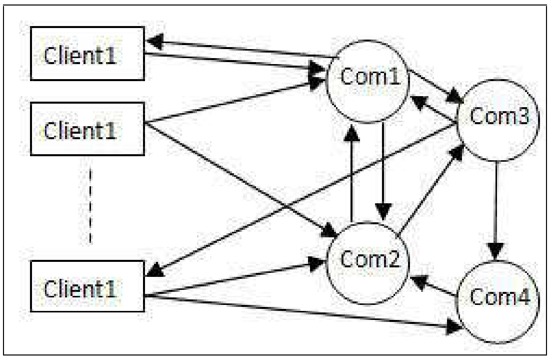


Fig. 1. Component-based System Architecture [8]

Software components and their interactions can be formally specified using an architecture description language to describe the software architecture. Informal and graphical notations, such as UML, are now also widely used by practical software developers for architecture specification. [8]

II. NEED TO ANALYZE ARCHITECTURE

Before analyzing architecture of a system one should ask and work on certain questions about the architecture before implementing the system over it. Questions like why? What? Who? When? How? Where? Earlier that one can detect a problem with architecture, the better of that the project will be because the longer that a fault goes undetected, the costlier that it will be to correct.

- First question can be arised that why an organization should review and evaluate software architecture? The bottom line is that architecture review produces better architectures resulting in the delivery of better systems. The most significant benefit of evaluation is to reassure stakeholders that the candidate architecture is capable of supporting the current and future business objectives; specifically, it can meet its functional and non-functional requirements. The quality attributes of a system such as performance, availability, extensibility, and security are direct results of its architecture; therefore quality cannot be introduced easily to your system late
- Evaluation also forces you to document the architecture clearly, so that it can be reviewed. Poorly specified requirements result in hit or miss architecture
- The architecture of the entire system can be evaluated, or only part of the system. A review can evaluate the architecture against all of the system's quality requirements, or only the most critical ones. Create a list of the specific criteria against which the architecture will be measured. The goal is to review and assess how each item on the list is affected by the architectural decisions that are made. Reliability, security, availability, extensibility, manageability, and portability are all quality attributes that can be considered in an architecture evaluation and review. While each evaluation produces different results, the goal is always the same to produce a better architecture
- When should an architecture evaluation take place? Generally, you want to conduct the evaluation when the architecture is specified but before anything has been implemented.[7]

III. PERFORMANCE IN ARCHITECTURE OF COMPONENT BASED SYSTEM

Performance prediction and measurement approaches for component-based software systems help software architects to evaluate their systems based on component performance specifications created by component developers. During the last ten years, researchers have proposed many approaches for evaluating the performance (i.e., response time, throughput, resource utilisation) of component-based software systems. These approaches deal with both performance prediction and performance measurement. Formal performance prediction methods, based on queuing network models, allow evaluating software architectural designs for performance. Existing methods provide prediction results such as response times and throughputs, but do not guide the software architect on how to improve the design. Classical performance models such as queuing networks, stochastic Petri nets, or stochastic process algebras can be used to model and analyse component-based systems, specialised component performance models are required to support the component-based software development process and to exploit the benefits of the component paradigm, such as reuse and division of work. [3]

Proposed methodology [9] for predicting the performance of component-based applications is depicted in *Fig. 2*

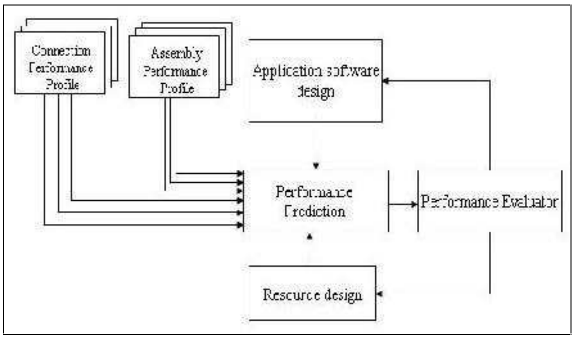


Fig. 2. Application performance prediction methodology [9]

- The Performance prediction module predicts the performance of an application taking into account the application software design, the resource design (i.e. hardware configuration), the assembly performance profiles of all contained assemblies in the application, and the connection performance profiles of all connection types used
- The Application software design comprises the design of the application, the use cases and scenarios that are important from performance point of view. In addition, it includes the workload estimation in terms of client requests, transaction mix, etc
- The Resource design defines the hardware resources (i.e. CPU, memory) used by the application
- Performance Evaluator analyses the predicted performance against the required performance. The feedback loop from the performance evaluator informs the application and system designer about the results of the design evaluation. If the performance goals are not met, either the application design or the resource design has to be changed and a new re-evaluation of the performance could be driven.[9]

Although many approaches have been proposed, they use different component notions (e.g., EJB, mathematical functions, etc.), aim at different life-cycle stages (e.g., design, maintenance, etc.), target different technical domains (e.g., embedded systems, distributed systems, etc.), and offer different degrees of tool support (e.g., textual modelling, graphical modelling, automated performance simulation). None of the approaches has gained widespread industrial use due to the still immature component performance models, limited tool support, and missing large-scale case-study reports. [3]

IV. RELATED WORK

Many researchers brought their work in this field, Dorina Petriu et al. (2000) [6] contributes toward bridging the gap between software architecture and performance analysis. They proposed a systematic approach to building Layered Queueing Network (LQN) performance models from a UML description of the high-level architecture of a system, and more exactly from the architectural patterns used for the system. The analysis of an LQN model produces results such as response time, throughput, queueing delays and utilization of different software and hardware components, and indicates which components are the system bottlenecks. In the second part of the paper, the proposed approach is applied to a telecommunication product for which a LQN model is built and analyzed. The analysis shows how the performance bottleneck is moving from component to component (hardware or software) under different loads and configurations, and exposes some weaknesses in the original software architecture, due to excessive serialization, which show up when more processing power is added to the system. Software components that do relatively little work on behalf of a system request can become the bottleneck in certain cases, whereas components that do most of the work do not. After removing the serialization constraints, a new software bottleneck emerges, which leads to the conclusion that the software architecture as it is does not scale up well.

Hassan Gomaa1 and Daniel A. Menascé (2001) [1] Worked in estimating the future performance of large and complex distributed software system at design time in order to reduce overall software cost and risk. It investigated into design and performance modeling of component interconnection patterns in client/server system defining the way client and server components communicate with each other. Approach they used is UML design models of the component interconnection patterns which are then performance annotated using an XML-based notation. The performance-annotated UML design model is mapped to a performance model, which can be used to analyze the performance of the software architecture on various configurations. In the UML notation, the class diagram is used to depict the static view of interconnection patterns and use stereotypes to differentiate among the different kinds of component and connector classes whereas the collaboration diagram depicts the dynamic view. This paper described component interconnection patterns for synchronous and asynchronous communication.

Vibhu Saujanya Sharma et al.(2005) [2] Presented an approach for performance evaluation of software systems following the layered architecture is proposed, which is a common architectural style for building software systems. The approach deals with first constructing a DTMC model of the software system, using the specifications user has provided and extracts parameters for constructing a closed Product Form Queueing Network model that is solved using the SHARPE software package. This PFQN model also takes care of limited software resources as threads on a particular machine .The tool built implements our approach and is very simple to use. The approach predicts the throughput and the average response time of the system under varying workloads and also identifies bottlenecks in the system, suggesting possibilities for their removal. The approach also allows for a prediction of the improvement in system performance if the scale up is actually done. There are two major applications of this approach: First is in architecting and deploying new layered systems and the second is in tweaking or upgrading or scaling up existing systems.

Heiko Koziolek (2009) [3] Conducted a comprehensive state-of-the-art survey of more than 20 of Performance prediction and measurement approaches assessing their applicability. It classified the approaches according to the expressiveness of their component performance modelling languages and critically evaluated the benefits and drawbacks. They claim there survey to help practitioners to select an appropriate approach. One can orient themselves with the proposed classification scheme and assess new approaches in the future. Practitioners can select methods according to their specific situation and thus increase the technology transfer from research to practise. A generic approach applicable on all kinds of component based systems may not be achievable. Instead, domain-specific approaches could decrease the effort for creating performance models. Many approaches have been proposed, they use different component notions (e.g., EJB, mathematical functions, etc.), aim at different life-cycle stages (e.g., design, maintenance, etc.), target different technical domains (e.g., embedded systems, distributed systems, etc.), and offer different degrees of tool support (e.g., textual modelling, graphical modelling, automated performance simulation).This paper presented a survey and critical evaluation of the proposed approaches to help selecting an appropriate approach.

Anne Martens et al. (2010) [4] Contributed a novel approach to find architectural design models with optimal performance, reliability, and cost properties. In this, an automated approach to search the design space for good solutions is presented. Starting with a given initial architectural model as an input, the approach iteratively modifies and evaluates architectural models.

Our approach applies a multi-criteria genetic algorithm to software architectures modelled with the Palladio Component Model. It supports quantitative performance, reliability, and cost prediction and can be extended to other quantitative quality criteria of software architectures. Quality property prediction is done using Layered Queueing Networks (LQN) for performance metrics; Markov models for reliability metrics; and a newly introduced PCM cost extension for cost. As metaheuristics are used, it is a best-effort approach. We validate the applicability of our approach by applying it to an architecture model of a component-based business information/reporting system (12 components, 40 tasks in LQN)and analyse its quality criteria trade-offs by automatically investigating more than 1200 alternative design candidates.

Seyed Saber Jalali et al. (2011) [5] Aims to recognize potential risks and investigating qualitative needs of software design before the process of production and implementation. Achievement to this goal reduces the costs and improves the software quality. In this paper, a new approach is presented to evaluate performance of component-based software architecture for software systems with distributed architecture. In this approach, first the system is modeled as a Discrete Time Markov Chain (DTMC), of which then the needed parameters for making a Product Form Queueing Network (PFQN) are extracted and the produced PFQN is solved by SHARPE software package. As the result of the solution of the produced model in this approach, throughput and the average response time and bottlenecks in different workloads of system are predicted and also suggested a required scale up to improve the performance of the system. Limitations of the source such as number of the threads are considered in a particular machine which allows the model to reach the real system and its limitations. This approach provides the possibility of the analysis of the software model on different hardwares with specifying the hardware rating factor by user.

V. CONCLUSION

In this paper, different approaches used for evaluating or analysing performance of component based architecture were reviewed. It was found to focus on the performance parameters of architecture. Currently Performance parameters like response time, throughput, and resource utilization can be evaluated over performance models of different architectures using various methods or tools like SHARPE, PCM, etc. We see how to optimise the expected performance of component-based early before implementing architecture to save cost of removing errors in late stage

REFERENCES

[1] Hassan Gomaa1 and Daniel A. Menascé, "Performance Engineering of Component-Based Distributed Software Systems," in Springer, R. Dumke et al. (Eds.): Performance Engineering, 2001, pp. 40–55.

[2] Vibhu Saujanya Sharma, Pankaj Jalote, Kishor S. Trivedi, "Evaluating Performance Attributes of Layered Software Architecture," in Springer, G.T. Heineman et al. (Eds.), 2005, pp. 66-81.

[3] Heiko Koziolek, "Performance Evaluation of Component-based Software Systems: A Survey," 2009.

[4] Anne Martens, Heiko Koziolek, Steffen Becker, Ralf Reussner, "Automatic, Model-Based Software Performance Improvement for Component-based Software Designs," Elsevier B.V., 2010.

[5] Seyed Saber Jalali, Hassan Rashidi, Eslam Nazemi, "A New Approach to Evaluate Performance of Component-based Software Architecture," UKSim 5th European Symposium on Computer Modeling and Simulation, 2011.

[6] Dorina Petriu, Christiane Shousha, and Anant Jalnapurkar, "Architecture-Based Performance Analysis Applied to a Telecommunication System," 2000.

[7] Denise cook, in skyscraper.net, 2007.

[8] He Jifeng, Xiaoshan Li and Zhiming Liu, "Component-Based Software Engineering-the Need to Link Methods and their Theories," UNU-IIST Report No.330, 2005.

[9] Nicolae Dumitrascu, Sean Murphy, Liam Murphy," A Methodology for Predicting the Performance of Component-Based Applications,"2003.

[10] Fred Broks," Component based Architecture", http://www.seas.gwu.edu/~bell/csci210/lectures/components.pdf

Recent Trends and Challenges in Augmented Reality

Saurabh Kapoor
Dept. of CSE
CTIEMT, Shahpur
Jalandhar, India
sauravkapoor04@gmail.com

Prashant Handa
Dept. of CSE
CTIEMT, Shahpur
Jalandhar, India
prashanthanda47@gmail.com

Ravinder Kaur
Dept. of ECE
CTIEMT, Shahpur
Jalandhar, India
er.rkaur@gmail.com

***Abstract*—Augmented Reality is a developing area in the field of virtual reality research. Similarly like Virtual Reality, Augmented Reality is becoming an emerging platform for numerous applications. The work done here reveals the current state-of-the-art in Augmented Reality. Moreover current issues, trends and challenges are analyzed here.**

***Keywords*—Augmented reality (AR), Heads-Up Display (HUD), Point of view (POV).**

I. Introduction

The first modern practice of AR was make known in 1993 [1] when a system, named KARMA, was established to assist maintenance by projecting wireframe schematics and maintenance directions on top of whatever was being repaired.

Augmented reality is a live sight of a physical, real-world environment which can be direct or indirect. AR is technology that associates virtual reality with the real world. The components of the environment are augmented. These components are sensory inputs which are created by the digital media (computer) such as sound, video, graphics or GPS data.

The virtual reality swaps the real world with a simulated one. Augmentation is convenient in real-time and in the semantic context. The advanced AR technology helps user to be interactive and scheming with the information about the real world. So the real world is superimposed by the artificial information about the environment and its objects. For example, a user may wear translucent goggles or see the screen of a camera equipped mobile device where they can see the real world plus the strategically placed computer generated images.

This work reveals the current state-of-the-art in AR. It also discusses the issues and problems faced in current era when building Augmented Reality Systems. It encapsulates the tradeoffs and approaches taken so far to overcome these problems and take a chance on future directions. The paper reviews few technologies, approaches and techniques that have been come across, like head-worn and hand-held devices, [2] [3]. Approaches related to vision based tracking are also analyzed.

Of course, this technology is gaining popularity due to its wide-range of applications. Still deemed as unfeasible and isolated just a few years back, augmented research and growing demand are pushing the barriers of AR further every day and consumer devices are progressively appearing in the market.

II. Types of AR

In light of the above criteria, two types of augmented reality recognized:

A. *Type I:* Augmented Reality (AR1), whereby the artificial imagery is layered over the background from the personal POV in real-time

B. *Type II:* Augmented Reality (AR2), whereby the artificial imagery is layered over the background from an impersonal POV or not in real-time

III. APPLICATIONS OF AR

A. Military

The Heads-Up Display is the emblematic example of augmented reality when it comes to military applications of the AR technology. A transparent display is placed directly in the fighter pilots view. Data usually displayed to the pilot includes altitude, airspeed and the horizon line in accumulation to other critical data. The term "heads-up" comes from the point that the pilot doesn't have to look down every time, at the aircraft's instrumentation to get the data they need. The HMD is also used by ground troops. Perilous data such as enemy location can be presented to the soldier within their line of sight. Moreover this technology is used in simulations for training purposes.

B. Medical

There have been really remarkable advances in medical application of augmented reality. Medical students use the technology to practice surgery in a controlled environment. Visualizations aid in clarifying complex medical situations to patients. Augmented reality can diminish the risk of an operation by giving the surgeon enhanced sensory perception. AR technology can be collaborated with MRI or X-ray systems and deliver everything into a single view for the surgeon.

Neurosurgery is at the pole position when it comes to surgical applications of AR. The ability to image the brain in 3D on top of the patient's actual structure is very powerful for the surgeon. As the brain is somewhat stationary matched to other parts of the body, the registration of exact coordinates can be accomplished. Apprehension still exists surrounding the movement of tissue during surgery. As this can disturb the actual positioning required for augmented reality to work.

C. Google's Project Glass

The Project Glass is an undertaking of the Google X Lab, which is already know for creating a driverless car shown in figure 1. Project Glass is a head mounted display that appears likes a pair of eyeglasses. It features an LED display over one eye, a touch interface, and voice-activated controls. It is another application for augmented reality.

Fig.1. Google's Project Glass

D. Mobile Applications

Today the one common way for everyone to interact with augmented reality is through mobile applications because modern smartphones are the perfect platform for an augmented reality interface. Most of the smartphones feature a video camera capable of generating a view of the world on which the phone hardware can overlay real-time information.

One such AR application is the Yelp app for iOS and Android. This app has a feature called "Monocle" that facilitates the camera to create a display that is superimposed with Yelp information. A user can then grip up the phone at a busy retail street and look at a real-time view of businesses with their Yelp information displayed on top.

E. Transportation Head-Up Displays

By projecting significant information onto the windshield of a vehicle, head-up displays can improve operator safety. Few cars currently feature simple head-up displays, like the Cadillac STS. Moreover all major automakers are looking to incorporate enhanced and sophisticated augmented reality into vehicles in the coming future. For example, General Motors is emerging technology to pool a head-up display with road sign recognition technology that will alert the driver to important landmarks and directions, shown in figure 2.

Fig.2. Head-Up Displays

F. Hand- Held Gaming

The Nintendo 3DS and PlayStations Vita hand-held game systems come with broad support for augmented reality. Both systems work in the same way by using cardboard "AR cards" to recognize the physical environment using a built-in camera. Playable game components are then overlaid on the camera image, giving the user a sensation of playing a video game in the physical world, shown in figure 3.

Fig.3. Hand- Held Gaming

IV. RESEARCH CHALLENGES

The AR system is basically affected by the display type, the system's sensing abilities, and the means for interaction. The display and sensing techniques regulate the efficiency and realism possible in the blending of the two realities, but may at the same time have ergonomic and social penalties. It may, in particular, be anticipated to achieve walk-up-and-use scenarios that support spontaneous interaction with marginal user preparation [4]. Unencumbering technology can also be accentuated, avoiding setups that rely on user-worn equipment [5], such as head-worn displays [6] or motion sensors [7]. It can also be advantageous to, to the greatest extent possible, preserve the qualities of the real space, while augmenting and supporting the user with unmediated view and control. The extreme use of artificial elements, such as visual reference patterns used for tracking, may, for example, have negative side-effects by cluttering or blocking the real environment that the system is meant to augment. It is also possible that some display technologies can result in significantly abridged visual quality due to optical properties, or the use of a down sampled view of the real environment.

V. VISUALIZATION ISSUES

Researchers have initiated to address issues in displaying information in augmented reality, resulted by the nature of AR technology or displays. Major issues are discussed below:

A. Visualization Errors

In fewAR systems, registration errors are noteworthy and inescapable. For example, the calculated location of an object in the environment might not be known adequately to avoid visible registration error. Under such circumstances, one approach for rendering an object is to visually display the area in screen space where the object could inherent, based upon expected tracking and measurement errors [8]. This assures that the virtual representation always contains the real counterpart. Second approach while rendering virtual objects should be obstructed by real objects is to use a probabilistic function that gradually fades out the hidden virtual object along the edges of the obstructed region, making registration errors less objectionable [9].

B. *Removing real objects from the environment*

The difficulty of removing real objects is more than simply take out depth information from a scene. The system must also be able to fragment individual objects in that environment.

A semi-automatic technique for identifying objects and their locations in the scene through silhouettes is found in [10]. It enables the enclosure of virtual objects and omission of real objects without an explicit 3D reconstruction of the environment.

C. *Photorealistic Rendering*

A main necessity for improving the rendering quality of virtual objects in AR applications is the capability to automatically capture the environmental illumination info [11-13, 15].For example, in [14] it presented a technique that, using only an uncalibrated camera, permits the capture of object geometry and appearance, and then, at a future stage, rendering and AR overlay into a new scene.

VI. Conclusion & Future Scope

The aim of Augmented Reality is to advance and enrich our perception of the surroundings by combining sensing, computing and display technologies. Although many recent advances are done in AR still much work remains to be done. Application developments can be facilitated by using the available libraries and one of them is ARToolkit [14], that helps with computer vision techniques to measure a camera's position and orientation relative to marked cards so as virtual 3D objects can be superimposed accurately on the markers. Still few areas require further research if AR is to become commonly deployed, such as Ubiquitous tracking and system portability, Photorealistic and advanced rendering and AR in all senses.

References

[1] Feiner, S., MacIntyre, B., and Seligmann, D. "Knowledge-based augmented reality Communications of the ACM". - 1993. - Vol. 7.

[2] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent advances in augmented reality," IEEE Computer Graphics and Applications, vol. 21, no. 6, pp. 34-47, Nov. 2001.

[3] O. Bimber and R. Raskar, Spatial Augmented Reality: Merging Real and Virtual Worlds. MA, USA: A. K. Peters, Ltd., 2005.

[4] Encarnacao, I. M., Barton, R. J., I., Bimber, O., and Schmalstieg, D. (2000). Walk-up VR: Virtual reality beyond projectionscreens. *IEEE Computer Graphics and Applications*, 20(6):19–23.

[5] Kaiser, E., Olwal, A., McGee, D., Benko, H., Corradini, A., Li, X., Cohen, P., and Feiner, S. (2003). Mutual disambiguation of 3Dmultimodal interaction in augmented and virtual reality. *Proc. ICMI '03 (International conference on Multimodalinterfaces)*, 12–19.

[6] Cakmakci, O. and Rolland, J. (2006). Head-worn displays: A review. Display Technology, 2(3):199–216.

[7] Welch, G. and Foxlin, E. (2002). Motion tracking: no silver bullet, but a respectable arsenal. IEEE Computer Graphics and

Applications, 22(6):24–38.

[8] B. MacIntyre and E. Coelho. Adapting to dynamic registration errors using level of error (loe) filtering. Proc. Int.lSymp. Augmented Reality, 85-88, 2000.

[9] A. Fuhrmann et. al. Occlusion in collaborative augmented environments. Computers Graphics, 23 (6): 809-819, 1999.

[10] V. Lepetit and M. Berger. Handling occlusions in augmented reality systems: A semi-automatic method. Proc. Int.l Symp. Augmented Reality, 137-146,2000.

[11] S. Boivin and A. Gagalowicz. Imagebased rendering for industrial applications. ERCIM News, 2001.

[12] Y. Mukaigawa, S. Mihashi, and T. Shakunaga. Photometric imagebased rendering for virtual lighting image synthesis. Proc. 2nd Int.l Workshop Augmented Reality (IWAR '99),115-124, 1999.

[13] J. Stauder. Augmented reality with automatic illumination control incorporating ellipsoidal models. IEEE Trans. Multimedia, 1 (2): 136-143, 1999.

[14] D. Cobzas, K. Yerex, and M. Jagersand. Editing real world scenes: Augmented reality with image-based rendering. Proc. of IEEE Virtual Reality, 291-292, 2003.

[15] M. Billinghurst, S. Baldis, E. Miller, and S. Weghorst. Shared space: Collaborative information spaces. Proc. of HCI International, 7-10, 1997.

Analysis of Security Processes in Software Development

Surkhab Shelly
Student, Department Of Computer Science
Guru Nanak Dev University
surkhab77@gmail.com

Anil Kumar
Department of Computer Science
Guru Nanak Dev University
anil.gndu@gmail.com

***Abstract*--Software Security is very critical issue for any Organization. Customers always need a software product without security flaws that assure information integrity, availability, and confidentiality. As soon as software products continue to present security flaws and be compromised by attacks, the Systems Security Engineering – Capability Maturity Model (SSE-CMM) becomes the de facto model to structure a software security approach. The focus is on establishment of security engineering approach to support software security through a specialized process, entitled Process to Support Software Security (PSSS) that helps to develop more secure software.**

Keywords- Security,PSSS,SDL,CLASP,OCTAVE

I. INTRODUCTION

Software security is the idea of engineering software so that it continues to function correctly under malicious attack. Security reflects the system's ability to protect itself from accidental or deliberate external attack. The primary goals of software security are the preservation of the confidentiality, integrity, and availability (CIA) of the information assets and resources.

ii. Security in SDLC

A software development life cycle (SDLC) [1] is a structure imposed on the development of a software product. The only reliable way to ensure that software is constructed secure and resilient is by integrating a security and resilience mindset and process throughout the entire software development life cycle (SDLC). Each phase of the SDLC is mapped with security activities, as demonstrated in the figure 1 and as explained below:

In Requirements phase, it needs Security Requirements, Setting up Phase Gates, Risk Assessment. Design Phase of SDLC involves some activities like Identify Design Requirements from security perspective, Architecture & Design Reviews, Threat Modelling, Security Design review. Coding Phase involves activities like Coding Best Practices, Perform Static Analysis and Peer Review. Testing Phase requires Vulnerability Assessment, Dynamic analysis and Fuzzing activities. Deployment Phase involves activities like Server Configuration Review, Finally Security Review, Application Security and monitoring and response plan.

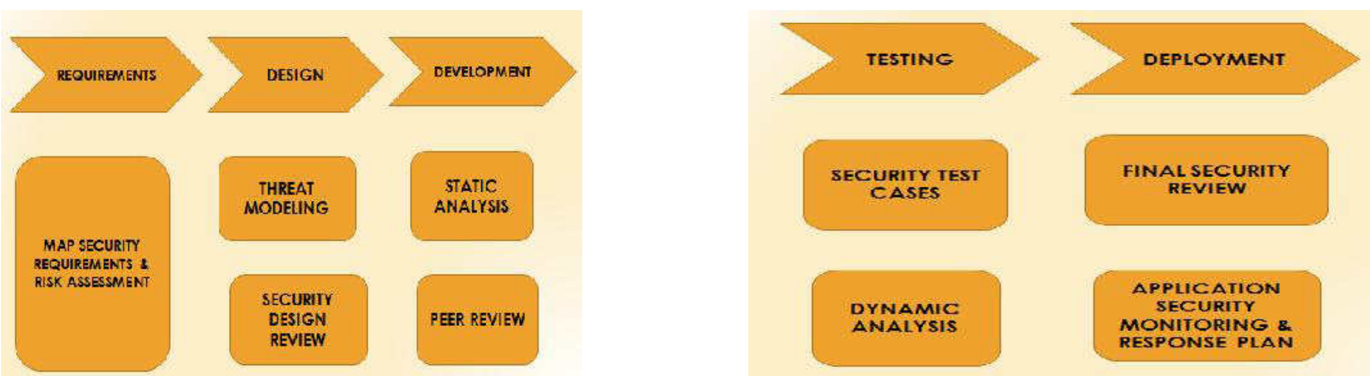


Fig. 1.Security in SDLC [1]

II. LITERATURE SURVEY

Literature is reviewed with respect to security standards as well as security processes.

A. Security Standards

i. Standards – Standards are established by some authority, custom, or by general consent as examples of best practices.

Standards provide material suitable for the definition of processes.

iii. ISO/IEC 15408

ISO/IEC 15408 [2] (Evaluation Criteria for Information Technology Security) presents a set of criteria to evaluate the security of products. This standard claims that a development process to produce secure software should include security in the development environment and security in the developed application. The secure development context of ISO/IEC 15408 is based on the execution of activities that describe how security requirements and specifications are derived when developing a software product.

iv. ISO/IEC 27002

ISO/IEC-27002 [3] (Code of Practice for Information Security Management) (2005) aims to preserve confidentiality, integrity and availability of information.

This is achieved through the implementation of security constraints, including policies and processes. These constraints ensure that defined security goals will be satisfied.

A restriction of ISO/IEC 27002 is that it contains a vast number of security controls to be applied among different processes in any kind of organization. This could be seen as a weakness as this could lead to wrong interpretations. Another issue is that the standard does not explain how to best implement each security control.

v. SSE-CMM

Noopur Davis (2005) has discussed that the SSE-CMM (System Security Engineering Capability Maturity Model) is a process model that can be used to improve and assess the security engineering capability of an organization. By defining such a framework, the SSE-CMM, provides a technique to improve the performance in the application of security engineering principles.

According to Anderson (2001) [4], Security engineering is about constructing systems to remain dependable in the face of risk, error, or mischance. It focuses on the tools, processes, and methods needed to design, implement, and test complete systems, and to adapt existing systems as their environment evolves. SSE-CMM ensures system security based on a framework that translates customer security needs into security products that satisfy the security requirements.

vi. OCTAVE

The Operationally Critical Threat, Asset, and Vulnerability Evaluation (OCTAVE [5] is a methodology for identifying and evaluating information security risks. It is intended to help an organization to

- Develop qualitative risk evaluation criteria that describe the organization's operational risk tolerances
- Identify assets that are important to the mission of the organization
- Identify vulnerabilities and threats to those assets
- Determine and evaluate the potential consequences to the organization if threats are realized

OCTAVE is a risk-based strategic assessment and planning technique for security. OCTAVE is self-directed, meaning that people from within the organization assume responsibility for setting the organization's security strategy

B. Security Processes

The IEEE defines a process as "a sequence of steps performed for a given purpose" [6]. A secure software process can be defined as the set of activities performed to develop, maintain, and deliver a secure software solution. Activities may not necessarily be sequential; they could be concurrent or iterative.

i. CLASP

According to Johan Gregoire [7] CLASP is a pre-defined set of documented processes and tools that can be integrated into any software development process. It is designed to be both easy to adopt and effective. The Open Web Application Security Project (OWASP) is an open community dedicated to enabling organizations to conceive development, acquire, operate, and maintain applications that can be trusted. One of their more prominent projects is called the Comprehensive, Lightweight Application Security Process, or CLASP.

The CLASP process includes seven best practices: "Institute awareness programs," "Perform application assessments," "Capture security requirements," "Implement secure development practice," "Build vulnerability remediation procedures," "Define and monitor metrics," and "Publish operational security guidelines." Under the best practices, 24 activities are organized. Furthermore, these activities consist of sub-activities. The primary goal of CLASP is to support the construction of software in which security takes a central role. Limitations are given below :

- CLASP does not build security team
- It does not determine whether the application is covered by methodologies
- It does not provide tools to track security issues

ii. SDL

Microsoft defined the SDL to address the security issues they frequently faced in many of their products. SDL comprises a set of activities, which complement Microsoft's development process and which are particularly aimed at addressing security issues [7]. Limitations are given below:

- It does not institute accountability for security issues
- While monitoring security metrics it does not identify metrics to collect and identify how they will be used.
- SDL also does not institute data collection and reporting strategy.

- It does not evaluate metrics during monitor security metrics.
- It does not build a global security policy, if necessary.

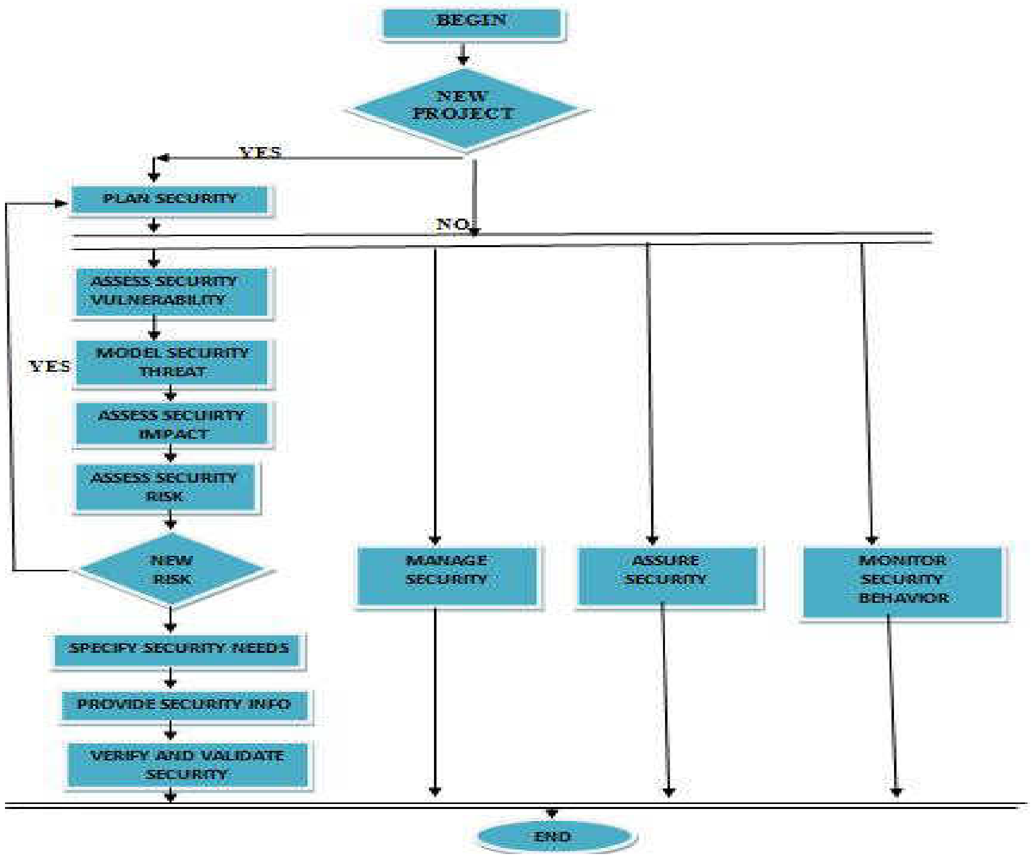


Fig. 2. Process to Support Software Security [8]

iii. PSSS

The Process to Support Software Security [8] was designed to follow the iterative and incremental life cycle approach which facilitates the coordination between the PSSS and any particular corporate development process. In order to use the PSSS with other life cycles, this would need validation. There is no need to use all the activities of the PSSS. It is an important aspect to have each activity as integrated as possible into the life cycle phases and one approach to reach this integration is to apply each activity in parallel with the phases. The PSSS identifies two important actors: the Security Engineer and the Security Auditor. The Security Engineer is responsible for the specialization of the PSSS based on the objectives of the software development project. The Security Auditor is responsible to evaluate whether the software development projects are done in compliance with the specialized PSSS. The main advantages gained by applying and following the PSSS are:

1) Assurance that security was considered during the system development through elaboration of security activities and artifacts.

2) Identification and definition of security requirements based on a set of security assessments and according to the major negative security impacts.

assurance that the limited project resources were effectively applied based on security

The PSSS considers 37 activities grouped in a set of 11 subprocesses (Figure 2). The subprocesses are described as follows:

- Plan security
- Assess security vulnerability
- Model security threat
- Assess Security Impact
- Assess Security Risk
- Specify Security Needs
- Provide Security Information
- Verify and Validate Security
- Manage Security
- Assure Security

III. CONCLUSION

After Analyzing above three security processes I have chosen PSSS as security approach because of its advantages over the other two security processes. The PSSS could be seen as an important tool to improve the effectiveness of software security projects. The definition of the PSSS tried to help in the production of more secure software as it protects the confidentiality, integrity, and availability of processed and stored information. The Process to Support Software Security was designed to follow the iterative and incremental life cycle approach which facilitates the coordination between the PSSS and any particular corporate development process. PSSS is designed for any iterative project and for any organization in need of software security solutions.

IV. FUTURE SCOPE

- To Implement the Security Engineering Approach in order to ensure software security throughout the entire software development life cycle.
- Design of an Iterative Project for an Organization in order to combine the Support Process and corporate software Development Methodology.
- Establishment of a security Engineering approach consisting of security activities forming a process to support the development of more secure software.

REFERENCES

[1] "SDLC Security – Security begins from within" available at http://dcvizcayno.wordpress.com/2012/03/28/sdlc-security-security-begins-from-within/

[2] "An overview of information security standards", february2008http://www.infosec.gov.hk/english/technical/files/overview.pdf

[3] Gefei Sun , Kenichi Yajima "A Supporting Tool for Creating and Maintaining Security Targets According to ISO/IEC 15408"

[4] Francisco José Barreto Nunes " PSSS - Process to Support Software Security" 2005.

[5] C. Alberts et al. (2001, December). OCTAVE - The Operationally Critical Threat, Asset, and Vulnerability Evaluation [Online]. Carnegie Mellon – Software Engineering Institute. Available: www.cert.org/octave.

[6] IEEE Standards Coordinating Committee. IEEE Standard Glossary of Software Engineering Terminology (IEEE Std 610.12-1990). Los Alamitos, CA: IEEE Computer Society, 1990 (ISBN 0738103918).

[7] Johan Gregoire, Koen Buyens, Bart De Win, Riccardo Scandariato, Wouter Joosen "On the Secure Software Development Process: CLASP and SDL Compared" 2007 IEEE.

[8] Francisco José Barreto Nunes , Arnaldo Dias "Security Engineering Approach to Support Software Security" IEEE 2010.

A Proposed Framework for Information Retrieval Model Based on Ontology

Javed Ahmad Khan
Dept. of computer science and engineering
AIACT&R, GGSIP University
New Delhi, India
javedahmad1989@gmail.com

Shipra Jain
Dept. of computer science and engineering
AIACT&R, GGSIP University
New Delhi, India
sjshiprajain@gmail.com

***Abstract*— In this paper we proposed a framework for information retrieval models based on ontology that show the domain knowledge representation and also improve the sharing and reusing of knowledge. By taking Course example, ontology based algorithm is designed and also describes the functions of the proposed framework. We also show that how we make the ontology concept more effective during the process of information retrieval. Ontology based information retrieval is a conventional idea that helps in information retrieval field and provides better quality of knowledge to user.**

***Keywords*—Semantic web, Ontology, OWL Language, Information retrieval, Protégé 4.3 alpha tool.**

I. INTRODUCTION

Ontology describes the essential links between the related components because it defines the better retrieval methods for the need of users and the facts of the existing of subjects. In artificial intelligence, Gruber proposed the definition of an ontology, which says that "ontology is a specification of a conceptualization "[1]. Ontology is a collection of related documents and these documents are related to specific domain and its main purpose is to define the ontology database. In this paper we will discuss the key technology of semantic retrieval based on domain ontology. In Traditional model the user query request compare with the word of the text one by one without considering the semantic matching concept between the user inquiries and domain information. If no element matched with the user keyword empty record will be returned for the user. Then the user modify the given query condition and again retrieve the information. But ontology based model give the user clear and formal description about the query and assures the accuracy for information retrieval. Information retrieval is the process that finds the user's information from the domain ontology based collection. For example, when user retrieves "Engineering Branch", retrieving transformer will make use of concepts of computer in ontology, to add "Engineering Branch", "Computer Science", "Information Technology" and other branch related to "Engineering " into retrieving data set. The enhanced retrieving condition will composed of all above data. Then, the retrieving coverage range will be dramatically improved. One word has two or more meanings and so there are difference between the user query request and the user real needs. So the precision of user query is low and the false rate is high. For improving the precision and recall we used the ontology based information retrieval model.

The paper is organized as follows: Section 2 contains Related Work, Section 3 Proposed information retrieval model framework; Section 4 proposed framework solve the problems of traditional models, Section 5 contains Ontology construction, Section 6 contains Query processing steps, and Section 7 contains Conclusion.

II. RELATED WORK

A. *Semantic Web*

Semantic Web is a concept proposed by Tim Berners-Lee who is the founder of the Internet in 1998. WWW is added to the semantic web that is the core concept of the semantic web and that is understood by the computer easily so that internet is used as a common medium for exchanging the information. Semantic web consist of seven layers as shown in figure 1 and its layer function increase gradually in bottom-up fashion [2] [3].

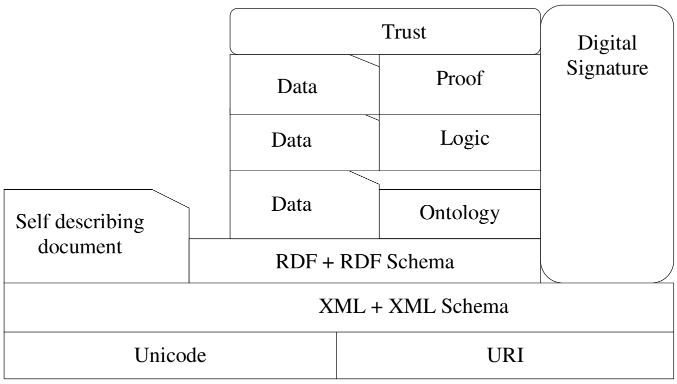


Fig.1. Semantic web layer architecture

1) *Unicode and URI Layer*

Unicode is responsible for resources encoding and URI is responsible for the identification of resources.

2) *XML+NS+XML Schema layer*

It's used to represent content and structure of data and information.

3) *RDF+RDF Schema layer*

This layer describes resources and their types on the Web.

4) *Ontology vocabulary layer*

Its describe resources and their link with each other.

5) *Logic layer*

This layer conducts logical reasoning operations on the basis of the following four levels.

6) *Proof layer*

This layer conducts verification in order to draw conclusions based on logical statements.

7) *Trust layer*

This is used to establish a trust relationship between users.

B. Ontology

Ontology provides a common understanding of a domain that can be exchange information across people and application systems. In 1993, Gruber said “Ontology is a formal explicit specification of a shared conceptualization” [1]. Ontology defines common words with which queries and assertions are exchanged for those who want to share information in a specific domain. It is a dynamic process and has the potential to capture modification of meaning (Fensel, 2001).Ontology aims to provide shared semantics for user communication. Ontology is used in the IR field to represent shared and more or less formal domain descriptions in order to add a semantic layer to the information retrieval system. Ontology established the concept, property, instance, as well as the relation between the information and the document. Concepts and relation can be regarded as basic constituent elements of ontology. Domain knowledge can be out clearly through the structure composed by concepts and the relationship of them. The content of information in each concept can be expressed by its attribute values [4]. Ontology concepts can be used to describes the information in the level of semantic and knowledge and provides the semantic foundations of information exchange and common understanding to the meaning of terms retrieve the document or information pages that have different appearance but have semantic similarity. Semantic web with the help of ontology concept focuses on the intelligent information retrieval. Ontology show different properties on the basis of these properties Gruber divide ontology into four categories as according to the level of precision and domain dependence as shown in figure 2.

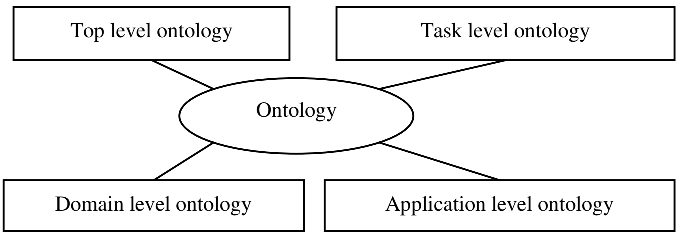


Fig.2. Categories of ontology

1) *Top level ontology*

Top level ontology or upper level ontology are the most general ontology describing the top-most level in ontology to which other ontology can be connected, directly or indirectly.

2) *Domain ontology*

Domain ontology describe a given domain, e.g. medicine, agriculture, politics; etc. Domain task ontology’s define domain level ontology on domain specific task and activities are primarily designed to fulfill the need for knowledge in a specific application.

3) *Task ontology*

Task ontology defines the top level ontology for generic tasks and activities.

4) *Application ontology*

Application ontology defines knowledge on the application-level. Evaluating an ontology language is a matter of determining what relationships are supported by the language and required by the ontology or application domain Ontology based domain knowledge representations and relationship among the ontology we takes the examples that show the relationship between these.

To show the concrete ontology-based domain knowledge representation and the relationships among them as shown in figure 3.

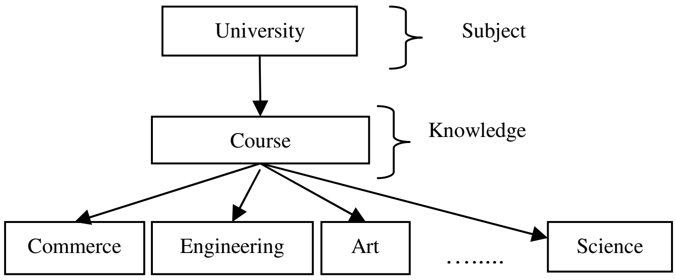


Fig.3. Domain knowledge representation of ontology

In above examples we can see that a subject corresponds to a set of knowledge concepts used to describe the subject; and a knowledge concept corresponds to plenty of multimedia materials. We extract part of the materials from the multimedia materials according to user’s cognitive style, etc. and organize these materials based on some specific question-answer mode to form the answer of user’s question.

C. Web Ontology Language

In present time there a lot of ontology description languages are presents they are based on the standard web ontology languages XML based , ontology exchange languages (XOL), ontology Inference layer (OIL), DARDPA agent markup languages (DAML), Resource description framework (RDF), RDF Schema (RDFS), DAML + OIL and Web ontology languages (OWL). The first thing is that it describes the content and deals with the other realized functions’ without any mechanism that is not precisely marked. The second thing is that coding languages such as the RDF that added the more features for designing of ontology [5] [6].OWL is an ontology description language standard recommended by World Wide Web consortium, OWL is semantic language and help for publishing in the WWW and sharing the ontology. OWL developed on the basis of DAML + OIL and also useful in the expansion of RDFS and also provide additional language to support semantic expression. Ontology represents the hierarchical relationship between the classes [7].

III. PROPOSED INFORMATION RETRIEVAL FRAMEWORK

A. *Ontology based information retrieval model*

With the help of our proposed framework we can extract the information from the ontology based data base such as files. We proposed framework for ontology based information retrieval model as shown in figure 4.

The functions and mechanism of each module are described as follows:

4) *Jena Inference Engine*

The function of the ontology inference engine is to improve the concepts of the inquiry sentences that are deal with the help of the query processing module. Inference engine used the reasoning set binding rules for extracting the related information from the data base and precise query statement is given as an output. Inference rules, including the attributes, relationship and so on. Jena is one of Hewlett-Packard (HP) laboratories for the development of a semantic Web application development package, the content is comprehensive, targeted and high efficiency; Jena is the most popular use inference engine for the OWL [8]. We take Jena as inference engine in our proposed framework.

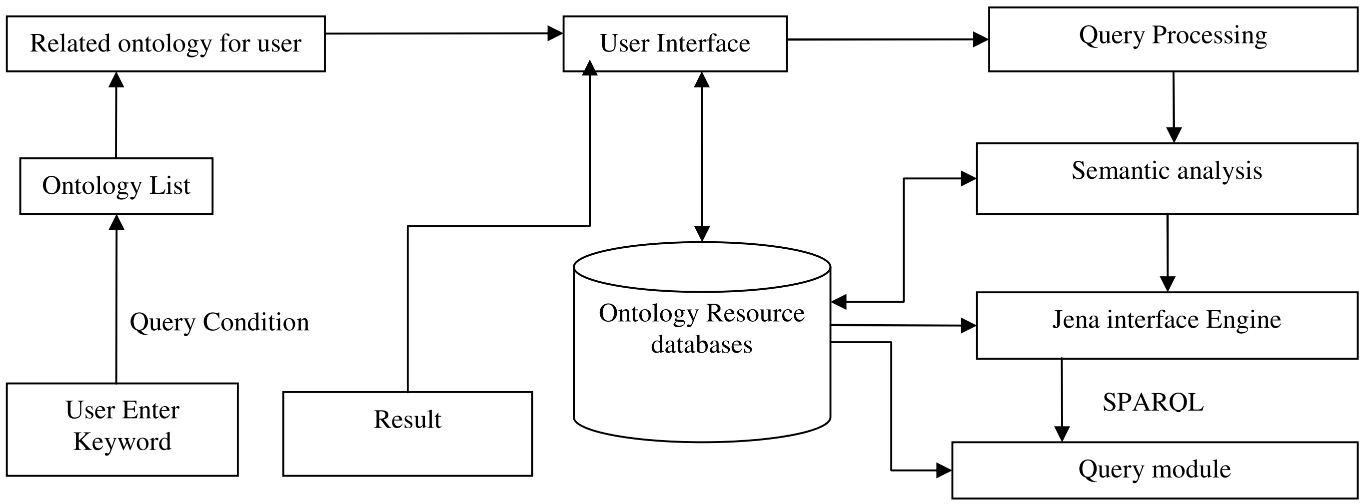


Fig.4. Framework for ontology based information based retrieval models

1) *Ontology List*

In this module ontology list are present and user wanted to retrieve according to their domain knowledge.

2) *User Interface Module*

User interface module aims to provide the visual interface to the user for inserting the query . User submits the query to the query processing modules for analyze and calculating the user request.

3) *Query Processing Module*

Query processing modules aim to managing the query submitted by the user. User query are presented in natural languages that contain the words or phrases. Query processing module split the user query and makes it a clear identifiable query set that the inference engine understands it.

5) *Ontology Resource Module*

Ontology resource module is the central part of the system. All types of information are described in this module by using the OWL language in the form of XML document that are linked with the information resources themselves whether the data are structured or unstructured. It will be loaded at startup. When the user put the query Jena will load OWL ontology to deal with the given query and also response to the user query.

6) *User interface module*

User interfaces interact with the system and accept the user query and after that this query are comes into the preprocessing modules and these query are submitted to the query semantic analyzer. The query request will be regulated with the help of the ontology and the after the knowledge base of domain ontology will be queried and the user query request are listed.

7) Semantic processing module

Semantic processing module executes the semantic query expansion to the query word or phrases. Semantic retrieval models are refers to the domain ontology and add the metadata in the semantic metadata and adds new phrases or words.

8) Query Preprocessing

Information retrieval help the user with the keyword based search interface but there are some problem to clearly express the user need and this is main drawback of the existing retrieval system can't meet the user needs. The second thing is that the user is not comfortable with the area from which he wants to retrieves the information and the user retrieved content can't express in accordance with the specification in the field. When the user input the query to the system in the form of natural languages the module executes the query preprocessing on the user request us in the domain ontology knowledge and some simple natural languages technology.

IV. PROPOSED FRAMEWORK SOLVE THE FOLLOWING PROBLEM

1) This proposed model add ontology knowledge data base between user and database which can solve the problem multiple meaning of one word. This model provides the facility to the user with in a specific domain.

2) In this proposed framework a system can infer set of words of same meaning from user input, retrieving word to the retrieving system because the similar meaning and abbreviations was added to ontology data added.

3) Ontology represents the powerful relationship in certain domain. It makes the concepts in a certain domain form a knowledge system which can represent the relationship in a logical way and can be used for inferring, which can efficiently feedback user most important information.

V. EXAMPLE OF ONTOLOGY CONSTRUCTION

In order to illustrate the ontology construction, in this paper, we choose one example to build product ontology. We combine the actual requirement of the retrieving model and the ontology theory. Protégé is developed by Stanford Medical Informatics that is an open source platform of ontology edition and knowledge management [9]. Taking a course example, using the ontology modeling tools protégé 4.3 alpha tool, the enterprise ontology is constructed. Protégé is a meta-tool for knowledge based systems which was discovered in 1987 [10]. An example is shown as figure 5.

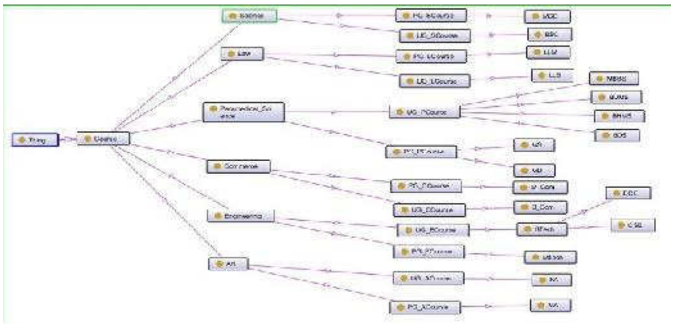


Fig.5. Construction of course ontology

VI. QUERY PROCESSING STEPS AND ALGORITHM

We show the query processing steps in figure 6.

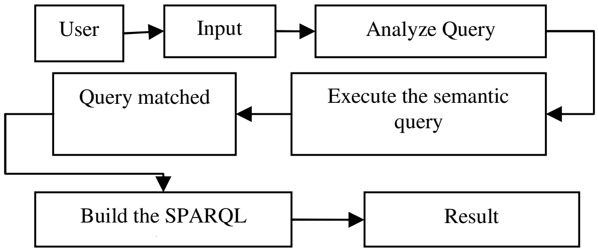


Fig.6. Query processing Steps

Step 1) User input the query

Step 2) Analyze sentence, when user submitted inquiries are received, combined with an analysis of information ontology, filter the user queries.

Step 3) executes the semantic quires for identification and judge the query which has been filtered from the number of query keyword in last steps combined with executes the semantic identifications. If the user not matches them return empty record if the matched go to next step.

Step 4) matches the query, the query that is given by user first save to query set the user requested. Matching can be divided into precise matching and fuzzy matching.

Step 5) build the SPARQL query sentence combined it with the query result.

Step 6) Result will be displayed data are extracted from the metadata according to the SPARQL query sentence and return the relevant information, the result or the output to the user interface and presented to the user.

VII. CONCLUSION

In this paper we propose ontology based information retrieval model that solve the problem which shows ambiguous relationship between words and ambiguity in simple keyword search and also show the domain knowledge ontology based relationship. Traditional keyword based retrieval model does not solve the ambiguous problem but ontology based information retrieval model solve this problem. Ontology based information retrieval models improve the completeness and accuracy of the users given query.

REFERENCES

[1] T.R. Gruber, A translation approach to portable ontology specifications, Knowledge Acquisition, 5 (2), pp 199-220, (1993).

[2] Berners- Lee, T.Hendler, J& Lassila "The Semantic Web,"Scientific American, May 2001.

[3] P. Lambrix, "Towards a Semantic Web for Bioinformatics using ontology-based annotation",in: Proceeding of the 14th IEEE International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises,2005,pp.3-7

[4] L. Minghong, A. Andreas, B. Ansger, H. Knut, S. Michael, "Ontologies for knowledge retrieval in organizational memories," International Workshop on Learning Software Organizations (LSO) at SEKE'99, 1999.

[5] M. C. Daconta, L. J. Orbst, K.T. Smith. The Semantic Web. Wiley Publishing, Inc., 2003.

[6] B.L. Li, Y.Z. Chen, "Information extraction research", Computer Engineering and Applications, vol.10, 2003, pp.1-5.

[7] P. Zhang, J.G.Guo, "Semantic web description language analysis,". Development and Application of Computer, Vol. 16, April 2002, pp. 31-33.

[8] F. Y. Zheng, L. H. Zhang, "Implementation of enterprise Web knowledge management system based on Ontology", Electronic Science and Technology, Vol.21, May 2008, pp. 57-60.

[9] X. H. Lv, B.Y. Chou, H. Qiao, " Ontology-based patent knowledge Discovery System," Computer and Information Technology, Vol.43, July 2008, pp. 43 49.

[10] John H. Gennari, Mark A. Musen etal, "The Evolution of Protégé: An Environment for Knowledge-Based Systems Development" International Journal of Human-Computer Studies , volume 58 issue1,January 2003, Page 89-123.

Comparison of Cost of Subquery Allocation for Different Allocation Plan

Palak Sharma
Dept. of CSE,
Guru Nanak Dev University, Amritsar.
palaksharma_1991@yahoo.com

Shelza
Dept. of CSE,
Guru Nanak Dev University, Amritsar.
shelzamahajan13@yahoo.in

Rajinder Singh Virk
Dept. of CSE,
Guru Nanak Dev University, Amritsar.
tovirk@yahoo.com

***Abstract*—The main purpose of this paper is to present a comparison of different dynamically generated allocation plans on the output of a simulator generating distributed subquery allocation. With the advancements in technologies need for distributed databases has also increased. The problem of data allocation is one of the major issues in distributed database design. It is an np hard problem and is very difficult to solve. Therefore the fragments/ tables accessed by queries to sites so as to reduce communication cost. Replicas of different tables can also be stored so as to reduce communication costs. We present a comparative study of effect of different allocations plans for replicas of tables on cost of subquery allocation.**

***Keywords*—Distributed databases, replication, comparison, sites, subquery allocation.**

I. INTRODUCTION

Distributed databases [7] have advanced a lot in the last few years, because of the advancement of technologies, networks and databases. A distributed database is basically a collection of sites, each site having different or same database, but belonging to same system. All these sites work together in collaboration and make it possible for a user to access data residing at any site from any site without any knowledge of location of database containing that data. Such a database can reside in same room, but all the fragments communicate with each other through network instead of shared memory. The main problem that arises in each and every distributed database is of fragmenting/ replicating the database and then allocating those fragments/ replicas to different sites. So, the main issues with which a distributed database management system deals is:

- How to fragment
- How to replicate
- How to allocate fragments to the sites.
  - Distributed query processor[2]

The main aim of query processor is to transform a high level query into low level query plan. Query processing becomes much more important in case of distributed databases. As in case of distributed systems[5], relations involved in query may be fragmented or replicated, and hence increase communication cost. In distributed databases, data is fragmented or replicated to increase locality of reference and parallel execution. So, the role of a distributed query processor can be defined as, mapping a query on a distributed databases into a sequence of operations on fragments of relations. The transformation must be correct and efficient.

II. PROBLEM DEFINITION

One of the major problems that arise in designing distributed database is of data allocation and fragmentation[3]. The data allocation problem is to fragment database and allocate those fragments to sites so as to minimize transfer costs incurred while executing queries which are posed over the distributed database environment. An inefficient fragmentation and allocation plan can increase query execution cost tremendously. The problem of fragmentation and allocation is NP hard in nature and is not easy to solve.

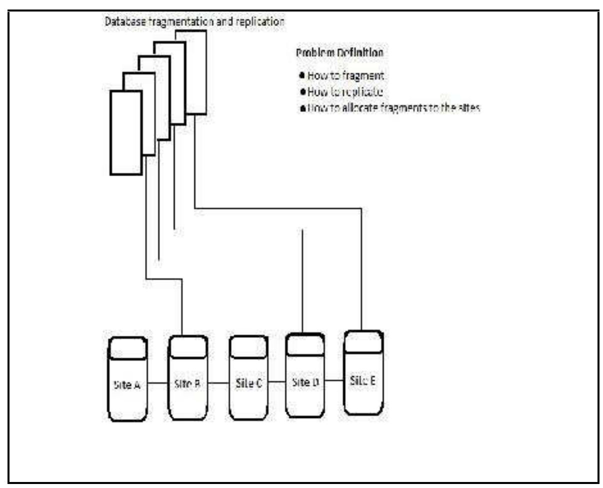


Fig. 1. Problem definition

The data allocation design problem is an essential issue, as a proper data allocation plan improves the performance of application processing. The fragments should be allocated to the sites, which access that fragment most frequently, so as to reduce communication costs and processing time We are working on a stochastic[6] simulator, which simulates distributed database environment. It performs task of sub query allocation i.e. allocates subqeries to those sites in such a way that total cost gets reduced. This simulator uses a static data fragmentation and allocation plan, which poses problems sometimes and hence, leads to increase in communication costs.

A formal description of problem can be given as:

A distributed database environment S consists of a number of sites:

$S= \{s_1, s_2 ,\ldots\ldots\ldots\ldots s_n\}$,

where $s_1,s_2\ldots\ldots$ etc. are sites, and site has some finite capacity.

A set of fragments F can be given as:

$F= \{f_1,f_2, \ldots\ldots\ldots\ldots\ldots f_m\}$,

where f_i represents i^{th} fragment and each fragment is required by at least one site.

A requirement matrix which shows which fragment is required by which site can be given as:

$$R = \begin{matrix} r_{1,1} & \cdots & r_{1,m} \\ \vdots & \vdots & \vdots \\ r_{n,1} & \cdots & r_{n,m} \end{matrix}$$

Where $r_{i,j}$ represents requirement for fragment j by site i , this is basically a weighted matrix.

The data allocation problem is to generate the set of fragments/replicas F and allocate the fragments to proper site on the basis of requirement matrix.

III. RELATED WORK

Static data allocation and fragmentation often results in inefficient query execution plan, which leads to high execution costs. Therefore a lot of methods have been devised to dynamically fragment and allocate databases to various sites in distributed database systems.

Bhuyar et al. (2012) have proposed a method for allocating fragments to a group of sites, instead of allocating fragments site by site. In this clustering method is used to reduce communication costs between sites during allocation process and fragment allocation is used to enhance system performance by increasing availability.

Rajesh John et al. (2008)[8] in his paper algorithm which uses intelligent agents to vertically fragment an object oriented databases.. He suggested that method based partitioning and attribute based partitioning can be used while vertical fragmentation. He gave a detailed method based partitioning algorithm.

Ishfaq Ahmad et al. (2002)[1] presented an evolutionary algorithm for data allocation. He used a site- independent fragment dependency graph to model the dependencies among fragments accessed by a query, and used it to formulate solution for data allocation problem.

Huang et al. (2001)[4] proposed two heuristic algorithms to find a near optimal allocation to minimize the total communication costs. He proposed solution for determining the replicated number of each fragments and then of finding a near optimal allocation for all those replicated fragments.

Horea Grebla[9] et al. presented a region based fragment allocation for non-replicated DDBMS. This algorithm moves the fragment from source node to target node by considering frequency of fragment access by region as well as individual nodes.

IV. METHODOLOGY

This section describes a simple approach to dynamically replicate database tables on different sites so that locality of table access can be improved and hence achieve reduction in communication cost. This strategy generates dynamic allocation plan using some replication[11] factor, which is used by a stochastic [6] simulator which generates subquery allocation plan for a distributed database using a genetic algorithm[10]. This Distributed subquery allocation simulator takes a query plan as input and generates a minimum cost allocation plan for its subqueries. But one of the drawbacks of this simulator is that it uses a static data allocation plan.

The effect on minimum cost generated by this simulator for different allocation plans can be observed. Following pseudo code has been used to generate dynamic allocation plans for this simulator. The allocation plans are used to generate subquery allocation plan for same query, to compare minimum cost generated.

```
Allocation[table][site] :- matrix represting allocation plan

For i=1 to no_of_tables

  For j=1 to no_of_sites

      Generate a random no (rand) between 0 and 1

        If(rand <replication factor)

                      Allocation[i] [j]=1;
```

Fig. 2. Allocation Plans

This code generates different allocation plans depending on the number of tables, number of sites and replication factor.

For a system with 7 tables (rows) and 10 sites(columns), and replication factor to be 0.2, an allocation plan as shown in figure will be generated.

0	1	0	0	0	0	1	0	1	1
1	0	0	0	0	0	1	0	0	1
0	0	0	1	0	1	0	0	0	1
0	1	0	0	1	1	0	0	0	1
0	0	0	0	0	0	0	0	1	0
0	0	0	0	1	0	0	0	0	0
1	0	0	0	0	0	1	0	1	0

Fig. 3. Allocation plan for 7 tables and 10 sites

V. COMPARISON OF COST FOR DIFFERENT ALLOCATION PLANS

This section presents different allocation plans generated as output by the above given code with different replication factor. Along with that it presents output of simulator corresponding to particular allocation plans and will show how it effects the cost of subquery allocation plan generation.

- For a replication factor : 0.2 ,One of the possible Allocation plans is:

0	0	1	1	1	1	0	1	1	0
1	0	0	0	1	0	0	0	0	0
1	0	0	0	1	0	0	1	1	1
1	0	0	0	1	1	1	0	1	1
0	0	0	0	0	0	1	0	1	1
1	1	1	0	0	1	0	1	0	0
0	1	0	0	1	1	0	1	0	0

And output of simulator for this allocation plan is:

10 10 4 5 9 5 7 2 1 4 2 9 5 7 1 2 9 2 7 5 2	19362.0	289.0	15319.0
10 10 4 5 9 5 1 2 1 4 2 9 5 7 1 2 9 2 7 5 2	19342.0	289.0	15799.0
10 10 4 6 9 5 7 10 1 4 5 9 5 7 4 8 4 6 8 5 6	21143.5	289.0	14384.5
10 10 4 5 9 5 7 9 1 4 5 9 5 7 4 8 4 6 8 5 2	21102.5	289.0	14371.5
10 10 4 5 9 5 1 2 1 4 5 9 5 7 4 8 4 6 8 5 2	21145.5	289.0	14361.5
10 10 4 5 9 5 1 10 1 4 2 9 5 7 1 2 5 2 7 5 10	19359.0	289.0	15792.0
10 10 4 5 9 5 7 2 7 6 2 9 5 7 8 3 6 6 5 8 2	22309.5	289.0	15704.5
10 10 4 5 9 5 1 10 1 4 2 9 5 7 8 3 6 6 5 8 2	22763.5	289.0	15663.5
10 10 4 5 9 5 7 2 1 4 5 9 5 7 4 8 4 6 8 5 2	21165.5	289.0	14371.5
10 10 4 6 9 5 1 2 1 4 2 9 5 7 1 2 5 2 7 5 10	19332.0	289.0	15789.0
10 10 4 5 9 5 7 9 1 4 5 9 5 7 4 8 4 6 8 5 2	21102.5	289.0	14371.5
10 10 4 5 9 5 1 2 1 4 2 9 5 7 1 2 5 2 7 5 10	19342.0	289.0	15799.0
10 10 4 5 9 5 7 2 1 4 5 9 5 7 4 8 4 6 8 5 2	21165.5	289.0	14371.5
2 10 4 5 9 5 1 2 7 6 5 9 5 7 4 8 4 6 8 5 2	21043.5	289.0	14375.5
10 10 4 5 9 5 7 9 1 4 5 9 5 7 4 8 4 6 8 5 2	21102.5	289.0	14371.5
10 10 4 5 9 5 7 2 1 6 2 9 5 9 7 6 9 4 2 2 10	20931.5	289.0	16112.5
10 10 4 5 9 5 1 9 1 4 5 9 5 7 4 8 4 6 8 5 2	21062.5	289.0	14361.5
10 10 4 5 9 5 7 9 1 4 5 9 5 7 4 8 4 6 8 5 2	21102.5	289.0	14371.5
10 10 4 6 9 5 7 9 1 4 5 9 5 7 4 8 4 6 8 5 2	21092.5	289.0	14361.5
10 10 4 5 9 5 7 10 7 6 2 9 5 7 8 3 6 6 5 8 6	22302.5	289.0	15697.5
10 10 4 5 9 5 1 10 1 4 2 9 5 7 8 3 6 6 5 8 2	22763.5	289.0	15663.5
10 10 4 6 9 5 7 10 7 6 2 9 5 7 8 3 6 6 5 8 2	22792.5	289.0	15687.5
10 10 4 5 9 5 1 2 1 6 5 9 5 7 4 8 4 6 8 5 10	22019.5	289.0	14381.5
2 10 4 6 9 5 7 2 7 4 2 9 5 7 1 2 5 2 7 5 10	20632.0	289.0	15333.0
10 10 4 5 9 5 1 10 7 6 2 9 5 7 8 3 6 6 5 8 2	22782.5	289.0	15677.5
10 10 4 6 9 5 7 9 1 4 5 9 5 7 4 8 4 6 8 5 2	21092.5	289.0	14361.5
10 10 4 5 9 5 7 2 1 4 2 9 5 7 1 2 9 2 7 5 2	19362.0	289.0	15319.0
10 10 4 5 9 5 7 2 1 6 2 9 5 7 1 2 5 2 7 5 10	19983.0	289.0	15319.0
2 10 4 6 9 5 7 2 7 4 2 9 5 9 7 6 9 4 2 2 10	20945.5	289.0	16126.5
10 10 4 6 9 5 7 9 1 4 5 9 5 7 4 8 4 6 8 5 2	21092.5	289.0	14361.5
10 10 4 5 9 5 7 2 1 4 2 9 5 7 1 2 5 2 7 5 10	19362.0	289.0	15319.0
10 10 4 5 9 5 7 2 1 6 2 9 5 7 8 3 6 6 5 8 6	22921.5	289.0	15650.5
10 10 4 5 9 5 1 10 7 4 2 9 5 7 3 3 6 6 5 8 10	22656.5	289.0	15677.5

total cost for generation is:1043650.5

Average cost for generation is:20873.81

Minimum cost chromosome is : 10 10 4 6 9 5 1 2 1 4 2 9 5 7 1 2 9 2 7 5 10

Minimum cost is :19832.0

Minimum cost for this allocation plan is 19832

- For a replication factor : 0.5, One of the possible Allocation plans is:

0	0	1	1	1	1	0	1	1	0
1	0	0	0	1	0	0	0	0	0
1	0	0	0	1	0	0	1	1	1
1	0	0	0	1	1	1	0	1	1
0	0	0	0	0	0	1	0	1	1
1	1	1	0	0	1	0	1	0	0
0	1	0	0	1	1	0	1	0	0

And output of simulator for this allocation plan is:

4 1 9 7 7 3 6 4 5 1 10 7 3 6 4 1 5 7 1 7 5	18568.0	279.0	15049.0	3240.0
4 1 9 7 7 3 6 4 5 1 10 7 3 6 4 1 5 7 1 7 5	18568.0	279.0	15049.0	3240.0
4 1 5 7 7 2 6 4 5 5 10 5 3 2 7 5 6 10 2 3 5	22022.0	279.0	16149.0	5594.0
4 1 9 1 7 2 6 4 5 1 10 7 3 6 4 1 5 7 1 7 5	18558.0	279.0	15039.0	3240.0
4 1 9 1 7 3 6 4 5 5 10 7 3 6 4 1 5 7 1 7 5	19457.0	279.0	15056.0	4122.0
4 1 5 7 7 2 6 4 5 5 10 5 3 2 7 5 6 10 2 3 5	22022.0	279.0	16149.0	5594.0
4 1 9 7 7 3 6 4 5 1 10 7 3 6 4 1 5 7 1 7 5	18568.0	279.0	15049.0	3240.0
4 1 9 7 7 2 6 4 5 1 10 7 3 6 4 1 5 7 1 7 5	18578.0	279.0	15059.0	3240.0
4 1 9 1 7 3 6 4 5 1 10 7 3 6 4 1 5 7 1 7 5	18548.0	279.0	15029.0	3240.0
4 1 9 1 7 3 6 4 5 1 10 7 3 6 4 1 5 7 1 7 5	18548.0	279.0	15029.0	3240.0
4 1 10 1 7 3 6 4 5 5 10 5 3 2 7 5 6 10 2 3 5	21952.0	279.0	16079.0	5594.0
3 1 9 1 7 2 6 4 5 1 10 7 3 6 4 1 5 7 1 7 5	18578.0	279.0	15059.0	3240.0
4 1 9 7 7 2 6 4 5 1 10 7 3 6 4 1 5 7 1 7 5	18578.0	279.0	15059.0	3240.0
4 1 9 7 7 2 6 4 5 1 10 7 3 6 4 1 5 7 1 7 5	18578.0	279.0	15059.0	3240.0
4 1 9 7 7 3 6 4 5 1 10 7 3 6 4 1 5 7 1 7 5	18568.0	279.0	15049.0	3240.0
4 1 9 7 7 3 6 4 5 1 10 7 3 6 4 1 5 7 1 7 5	18568.0	279.0	15049.0	3240.0
4 1 5 1 7 3 6 4 5 5 10 5 3 2 7 5 6 10 2 3 5	21952.0	279.0	16079.0	5594.0
4 1 9 1 7 2 6 4 5 1 10 7 3 6 4 1 5 7 1 7 5	18558.0	279.0	15039.0	3240.0
4 1 9 7 7 3 6 4 5 5 10 7 3 6 4 1 5 7 1 7 5	19457.0	279.0	15056.0	4122.0
4 1 9 7 7 2 6 4 5 1 10 7 3 6 4 1 5 7 1 7 5	18578.0	279.0	15059.0	3240.0
4 1 9 7 7 3 6 4 5 1 10 7 3 6 4 1 5 7 1 7 5	18568.0	279.0	15049.0	3240.0
4 1 10 1 7 3 6 4 5 1 10 7 3 6 4 1 5 7 1 7 5	18538.0	279.0	15019.0	3240.0
4 1 9 1 7 3 6 4 5 1 10 7 3 6 4 1 5 7 1 7 5	18548.0	279.0	15029.0	3240.0
4 1 9 7 7 2 6 4 5 1 10 7 3 6 4 1 5 7 1 7 5	18578.0	279.0	15059.0	3240.0
4 1 9 1 7 3 6 4 5 5 10 7 3 6 4 1 5 7 1 7 5	19457.0	279.0	15056.0	4122.0
3 1 9 7 7 3 6 4 5 1 10 7 3 6 4 1 5 7 1 7 5	18558.0	279.0	15069.0	3240.0
3 1 9 7 7 3 6 4 5 5 10 7 3 6 4 1 5 7 1 7 5	19477.0	279.0	15076.0	4122.0
4 1 9 7 7 2 6 4 5 1 10 7 3 6 4 1 5 7 1 7 5	18578.0	279.0	15059.0	3240.0
4 1 9 1 7 2 6 4 5 1 10 7 3 6 4 1 5 7 1 7 5	18558.0	279.0	15039.0	3240.0
4 1 9 1 7 3 6 4 5 1 10 7 3 6 4 1 5 7 1 7 5	18548.0	279.0	15029.0	3240.0

total cost for generation is:973192.0

Average cost for generation is:19463.84

Minimum cost chromosome is : 4 1 5 1 10 1 6 4 5 1 10 7 3 6 4 1 3 7 1 7 5

Minimum cost is :18528.0

Minimum cost for this allocation plan is 18528

- For a replication factor 0.7, One of the possible Allocation plans is:

1	1	1	1	1	0	0	1	1	0
0	0	1	1	1	0	1	1	1	1
1	0	0	1	1	1	1	1	0	1
1	1	1	1	1	0	1	1	1	1
1	0	1	1	0	1	0	1	1	1
0	1	0	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	0

And output of simulator for this allocation plan is:

3 10 10 10 6 2 2 4 3 5 4 10 7 7 10 8 2 3 2 8 7	21674.0	295.0	15244.0	6135.0
1 10 10 10 6 2 2 2 3 5 2 4 7 1 3 4 5 7 10 3 10	22331.5	285.0	16222.5	5824.0
1 5 10 10 6 2 2 2 3 5 2 4 7 1 3 4 5 7 10 3 10	22341.5	285.0	16232.5	5824.0
1 10 10 10 9 2 2 4 3 5 2 4 7 1 3 4 5 7 10 3 10	22334.5	285.0	16225.5	5824.0
1 10 10 10 6 2 2 2 3 5 2 4 7 1 3 4 5 7 10 3 10	22331.5	285.0	16222.5	5824.0
3 10 10 10 6 2 2 4 3 5 4 10 7 7 10 8 2 3 2 8 7	21674.0	295.0	15244.0	6135.0
3 10 10 10 9 2 2 4 3 5 2 4 7 1 3 4 5 7 10 3 10	22334.5	285.0	16225.5	5824.0
1 10 10 10 6 2 2 4 3 5 4 10 7 7 10 8 2 3 2 8 7	21674.0	295.0	15244.0	6135.0
1 10 10 10 6 2 2 4 3 5 2 4 7 1 3 4 5 7 10 3 10	22324.5	285.0	16215.5	5824.0
1 5 10 10 6 6 2 4 3 5 4 10 7 7 10 8 2 3 2 8 7	21674.0	295.0	15244.0	6135.0
1 5 10 10 6 2 4 4 3 5 2 4 7 1 3 4 5 7 10 3 10	22324.5	285.0	16215.5	5824.0
1 5 10 10 6 2 2 4 3 5 2 4 7 1 3 4 5 7 10 3 10	22334.5	285.0	16225.5	5824.0
1 10 10 10 6 2 2 4 3 5 4 10 7 7 10 8 2 3 2 8 7	21674.0	295.0	15244.0	6135.0
1 10 10 10 9 2 2 2 3 5 2 4 7 1 3 4 5 7 10 3 10	22341.5	285.0	16232.5	5824.0
3 10 10 10 6 2 2 4 3 5 2 4 7 1 3 4 5 7 10 3 10	22324.5	285.0	16215.5	5824.0
1 5 10 10 6 6 2 4 3 5 2 4 7 1 3 4 5 7 10 3 10	22324.5	285.0	16215.5	5824.0
3 10 10 10 9 2 2 4 3 5 4 10 7 7 10 8 2 3 2 8 7	21684.0	295.0	15254.0	6135.0
1 10 10 10 6 2 2 4 3 5 2 4 7 1 3 4 5 7 10 3 10	22324.5	285.0	16215.5	5824.0
1 10 10 10 9 2 2 4 3 5 4 10 7 7 10 8 2 3 2 8 7	21684.0	295.0	15254.0	6135.0
1 5 10 10 6 2 2 4 3 5 4 10 7 7 10 8 2 3 2 8 7	21684.0	295.0	15254.0	6135.0
3 10 10 10 6 2 2 2 3 5 2 4 7 1 3 4 5 7 10 3 10	22331.5	285.0	16222.5	5824.0
1 10 10 10 9 2 4 4 3 5 2 4 7 1 3 4 5 7 10 3 10	22324.5	285.0	16215.5	5824.0
1 10 10 10 9 2 2 2 3 5 2 4 7 1 3 4 5 7 10 3 10	22341.5	285.0	16232.5	5824.0
1 10 10 10 6 2 2 4 3 5 4 10 7 7 10 8 2 3 2 8 7	21674.0	295.0	15244.0	6135.0
3 10 10 10 6 2 2 4 3 5 2 4 7 1 3 4 5 7 10 3 10	22324.5	285.0	16215.5	5824.0
3 10 10 10 9 2 2 4 3 5 2 4 7 1 3 4 5 7 10 3 10	22334.5	285.0	16225.5	5824.0
1 10 10 10 6 2 2 4 3 5 4 10 7 7 10 8 2 3 2 8 7	21674.0	295.0	15244.0	6135.0
1 5 10 10 9 2 4 4 3 5 2 4 7 1 3 4 5 7 10 3 10	22334.5	285.0	16225.5	5824.0
1 10 10 10 6 2 2 4 3 5 4 10 7 7 10 8 2 3 2 8 7	21674.0	295.0	15244.0	6135.0

total cost for generation is:1104244.5

Average cost for generation is:22084.89

Minimum cost chromosome is : 1 10 10 10 6 2 4 4 8 5 4 10 7 7 10 8 2 8 2 3 7

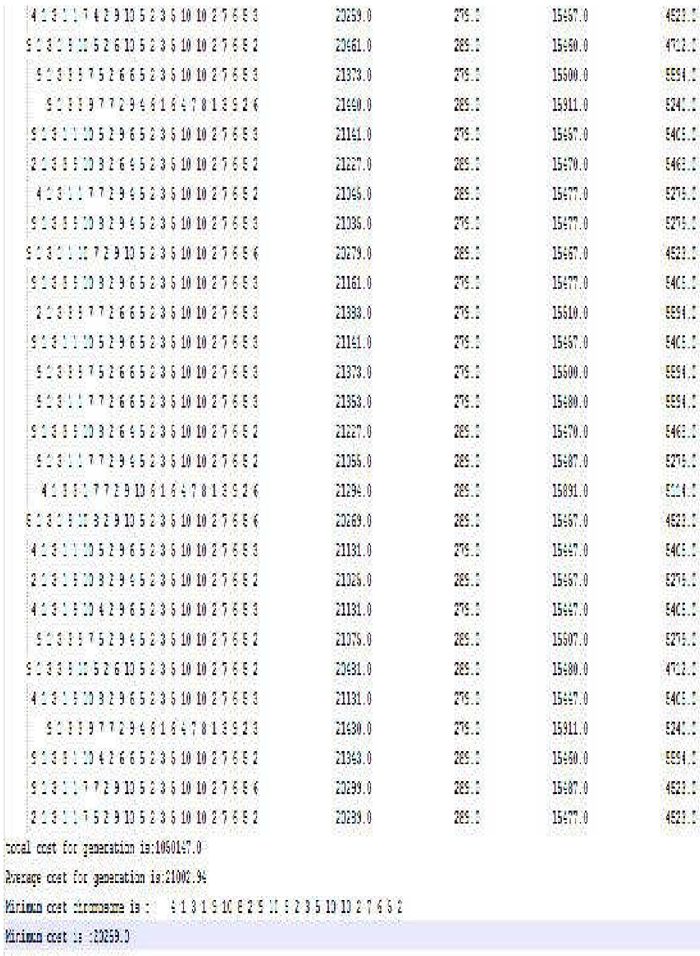
Minimum cost is :21664.0

Minimum cost for this allocation plan is 21664.

- For a replication factor 0.9 , One of the possible allocation plan is:

1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	0	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	0	1	1	1	1

And output of simulator for this allocation plan is:



Minimum cost for this allocation plan is 20599

VI. CONCLUSION

We presented a comparison of different dynamically generated allocation plan, having varying degree of replication and their effect on output of a stochastic simulator. Distributed design deals with the allocation of different replicas/fragments of a distributed database. This paper has presented a pseudo code for generating dynamic allocation with replication on the basis of a replication factor, which decides degree of replication. Different allocation plan have been generated and used as input for distributed subquery allocation generator and effect on output has been presented. It is clear from the comparative study that minimum cost is achieved with 50% replication. Small degree of replication or a large degree of replication increases the cost, because with a small degree of replication communication cost increases and with large degree of replication storage cost increases.

REFERENCES

[1]. Ahmad, Ishfaq, et al. "Evolutionary algorithms for allocating data in distributed database systems." Distributed and Parallel Databases 11.1 (2002): 5-32

[2]. Bamnote, G. R., and S. S. Agrawal. "Introduction to Query Processing and Optimization." International Journal 3.7 (2013).

[3]. Connolly, Thomas M. Database systems: a practical approach to design, implementation, and management. Pearson Education, 2005.

[4]. Huang, Yin-Fu, and Jyh-Her Chen. "Fragment allocation in distributed database design." J. Inf. Sci. Eng. 17.3 (2001): 491-506.

[5]. Korth, Henry F., Abraham Silberschatz, and Henry F. Korth. Database system concepts. Vol. 582. New York: McGraw-Hill, 1986.

[6]. Manik, R.virk, et al. "Stochastic Analysis of DSS Queries for a Distributed Database Design." International Journal of Computer Applications 83 (2013).

[7]. Özsu, M. Tamer, and Patrick Valduriez. Principles of distributed database systems. Springer, 2011.

[8]. John, Rajan, and V. Saravanan. "Vertical Partitioning in Object Oriented Databases Using Intelligent Agents." IJCSNS 8.10 (2008): 205.

[9]. Varghese, Pallath Paul, and Tarun Gulyani. "Region based Fragment Allocation in Non-Replicated Distributed Database System."

[10]. Whitley, Darrell. "A genetic algorithm tutorial." Statistics and computing 4.2 (1994): 65-85.

[11]. Wiesmann, Matthias, et al. "Understanding replication in databases and distributed systems." Distributed Computing Systems, 2000. Proceedings. 20th International Conference on. IEEE, 2000.

Analysis of Various Feature Selection Techniques for Network Intrusion Detection Dataset in WEKA

Tanya Garg
Centre for Computer Science & Technology
Central University of Punjab, Bathinda, India
Tanyagarg5023@gmail.com

***Abstract*- As the data on the network is growing day by day, there is requirement of detecting that data for intrusions with high speed and accuracy. These growing Network data is causing a serious problem of detecting intrusions to protect the useful information on the network. There are numerous network security tools to protect network from intrusions but still the fast growth of intrusive activities is a serious issue. Intrusion detection systems (IDSs) are used to detect intrusive activities on the network. Machine learning and Feature Selection Techniques help to design "Intrusion Detection Models" which can classify the network traffic into intrusive or normal traffic. Generally the intrusions are detected by analyzing 41 attributes from the intrusion detection dataset. In this work we tried to reduce the number of attributes by using various ranking based feature selection techniques and evaluation has been done using ten classification algorithms that I have evaluated most important. So that the intrusions can be detected accurately in short period of time. Finally the performance of the six reduced feature sets has been analyzed.**

***Index terms*—Network Intrusion Detection, NSL-KDD Dataset, WEKA, Data Mining, Feature Selection Techniques, Classification**

I. INTRODUCTION

Due to large amount of data on network the data content is vulnerable to various kinds of attacks i.e. increase in intrusions is growing day by day. Intrusion detection is very important to prevent the intruders to break into or misuse your system. To defend against various network attacks and computer viruses a lot of methods have been developed. Among these methods NID (Network Intrusion Detection) has been considered as the most promising method to protect from complex and dynamic intrusion behaviors [1]. Intrusion Detection Systems classifies data into various categories for example normal and intrusive. Various classification algorithms have been proposed to design an effective Intrusion Detection Model. The performance of a classifier is an important factor affecting the performance of Intrusion Detection Model. Hence the selection of accurate classifier helps to improve the performance of intrusion detection system. Feature Selection is an important data pre-processing tool in the field of Data Mining. The research is in the growing stage in this field for the past three decades [2]. As the data is growing day by day on network in terms of features and instances, it is necessary to reduce that data to reduce its processing time and to achieve higher accuracy in results. This process of removing irrelevant data is called Data Mining. In our research work we have to reduce the features for the Network Intrusion Detection Dataset, for which we used the ranking based feature selection techniques along with ranker to get the features ranked according to their importance. To analyze intrusions 41 attributes are considered. Our aim is to reduce number of those attributes so that intrusions can be detected in short period of time with higher accuracy. The use of feature selection techniques is to remove the irrelevant or useless features that are not contributing any useful information but are wasting time of intrusion detection model. There are several performance metrics to evaluate performance of Feature Selection Techniques. In this work, NSL-KDD (Network Security Layer-Knowledge Discovery in Database) compatible classification algorithms have been evaluated using WEKA (Waikato Environment for Knowledge Analysis) tool. Then Top ten classification algorithms are used to evaluate the Six Ranking Based Feature Selection Techniques. The performance of the Feature selection Techniques have been measured by considering Accuracy, Training time and FPR (False Positive Rate) performance metrics.

In this paper, initially, WEKA tool and ten classification algorithms to be used have been discussed in section II and III respectively. Various Feature selection techniques have been discussed in section IV. Chosen dataset has been introduced in section V. Results have been discussed in Section VI and conclusions are discussed in Section VII.

II. WAIKATO ENVIRONMENT FOR KNOWLEDGE ANALYSIS (WEKA)

WEKA is a Data Mining, Machine Learning Tool [1]. It is a collection of number of Machine Learning and Data Mining algorithms. This software is written in Java language and contains a GUI Interface to interact with data Files. It contains 49 data pre-processing tools, 76 classification algorithms, 15 attribute evaluators and ten search algorithms for feature selection [2]. It contains three algorithms to find association rules. It has three Graphical User Interfaces: "The Explorer", "The Experimenter" and "The Knowledge Flow." The WEKA supports data stored in ARFF (Attribute Relation File Format) file format. WEKA provides the capability to develop and include the new Machine Learning algorithm in it. The algorithms can be directly applied to dataset [7].

III. CLASSIFICATION ALGORITHMS

Classifiers are also called as machine learning algorithms or classification algorithms. They are used to classify the network traffic when used on dataset. They are responsible for classification of network traffic as normal or intrusive, if intrusive then to which category it belongs.

In our work there are 72 classifiers that are compatible on our chosen dataset and on the basis of their performance we have evaluated best ten classifiers and then these classifiers have been used for feature selection process. In this section the ten classifiers we are using for our work are discussed as:

Lazy classifiers: These learners are called as lazy learners also because they are computationally expensive i.e. they take a lot of time for computation for classification as their most of the power resides in the matching scheme. It takes large amount for classification due to the reason that each example to be classified must be compared to the each of the example in training dataset. The lazy classifiers we used are lazy IB1 and IBk [4].

Random Forest: It is an ensembling classification algorithm and is based upon decision tree algorithm and produces output in the form of individual trees. This algorithm is a combination of bagging idea and random selection of features to construct a collection of decision trees with controlled variation. It is one of the highest accurate classifier for many datasets. A large number of variables can be handled by this algorithm without ignoring any variable [22].

Random Tree: Random Tree as its name indicating it's a tree build by picking random branches from a possible set of trees. Each tree has an equal probability of being get sampled in this algorithm or we can say the trees are distributed in a uniform way. Random Trees can be generated easily and efficiently. Combination of large sets of Random trees mostly designs accurate models [17].

JRip (Extended Repeated Incremental Pruning) is that type of rule based classifier that implements a propositional rule that performs repeated pruning to reduce errors and also called as RIPPER (Repeated Incremental Pruning to produce Error Reduction). JRip is a rule learner that exactly works like commercial rule learner RIPPER [4].

NB Tree: NB Tree (Naïve Bayes) is used and applicable to scale large databases and used to improve the performance of decision trees and Naïve Bayesian Classifiers. Attribute need not to be independent for this classification algorithm [12].

Rotation Forest: Rotation Forest is also an ensemble classifier that transforms the data, subset of instances, subsets of classes and subsets of features using Principal Component Analysis because this method of transforming data requires less storage space and has low computation time. Rotation Forest is based upon two base classifiers: decision tree and Forest. It works on two key components: diversity and accuracy [5].

IV. FEATURE SELECTION

Feature selection is the process of choosing relevant features from large number of features for a particular dataset by applying particular evaluation criteria as desired by user. Generally Feature selection process involves three phases. It starts with selecting subset of original features and then it evaluates the worth of each feature for that particular dataset as we are using NSL-KDD Dataset for Intrusion Detection. Then after evaluation of worth of each feature by learning algorithms, the features having lower rank or lower value can be eliminated from dataset. Finally the third step is to evaluate the reduced feature set with classification algorithms to check whether it gives better result than previous one or not. If yes then that will be the final efficient reduced dataset [11].

Feature selection is broadly classified in two categories: one is Feature subset selection and other is feature ranking. The techniques used for Feature subset selection are different from that of the feature ranking techniques. Feature subset selection techniques gives the best reduced feature set according to algorithms used for searching best feature subset [21]. They do not rank the features. Feature ranking techniques calculates the score of each feature and rank them accordingly and we can pick the top k (as required by user) features according to higher rank and ignoring those having lower rank. We here are using Feature ranking for feature selection in our work Now the Feature selection can be done using three models i.e. one of these models can be used for feature selection. Feature selection can be done using Filter Model, Wrapper Model or Hybrid Model. Filter model performs feature selection technique without use of learning algorithm while Wrapper model involves use of learning algorithm first and then performs feature selection. Hybrid model involves use of both of these models together for feature selection.

Our work is focused and limited to filter model using seven ranking based feature selection techniques. The main advantage of using filter model is that it work without using learning algorithm, So it is unbiased. Secondly it is very simple and easy to use. It is very easy to design algorithms using his model due to its simple structure.

Information Gain (IG): Information Gain feature selection technique is based on the concept of entropy. Entropy is a measure of how pure or impure a variable is (Moore. W.A.). The expected resultant value for the feature by using information gain method is the mutual information of target variable say X and independent variable say Y. It is the reduction in the entropy of a target variable (X) achieved by using Learning algorithm (Y). Information gain method has a drawback that it chooses attributes having large distinct values over the attributes having fewer distinct values even though they are more informative [19].

Gain Ratio (GR): The Information Gain is biased towards tests with many outcomes [22]. Gain Ratio is a modification in Information gain to reduce its biasness. It takes into account the number of branches while choosing an appropriate attribute. It is based on the information given by intrinsic attributes. In this method the value of attribute decreases as intrinsic information gets larger. Gain ratio chooses an attribute only when its intrinsic information is very low. Intrinsic information means the amount of information which is needed to decide which branch belongs to which instance (Frank & Witten, 2011). The attribute having maximum value of gain ratio will be selected as the splitting attribute.

ReliefF: The Feature Selection Technique ReliefF was proposed by Kira and Rendell in 1994. It is very easy to use and is fast and accurate technique [23]. Relief works by measuring the ability of an attribute in separating similar instances. It can also be used to deal with noisy data and regression problems.

The process of selecting and ranking the features by using this technique involves basically three steps:

1. Calculate the nearest miss and nearest hit.
2. Calculate the weight of a feature.
3. Return a ranked list of features or the top k features according to a given threshold.

One R: It is a rule based algorithm as it generates rules and on the basis of those rules it selects features and rank them accordingly. OneR constructs rules and tests a single attribute at a time and branch for every value of that attribute.

Symmetrical Uncertainty: It is a correlation based Feature Selection approach. Correlation based feature selection evaluates the merit of a feature in a subset using a hypothesis – "Good feature subsets contain features highly correlated with the class, yet uncorrelated to each other" [22]. This Feature selection method is used to measure the degree of association between the discrete attributes. It is also based on the concept of entropy and is a symmetric measure to measure correlation between set of features.

Filtered Attribute Evaluator: It is based on the same principle of Information Gain Feature Selection Technique. In our work we used this technique and found it gives the attributes in the same order as selected by Information Gain Method.

Chi-Square: It is based on the statistical theory. It measures the lack of independence between the attributes [22]. It can test strength of relationship between two variables. It predicts the value of an attribute using observed and expected values of attributes. The value of an attribute based upon this feature selection method can be calculated using the formula:

$$\chi^2 = \sum_{ij} (O_{ij} - E_{ij})^2 / E_{ij} \qquad (1)$$

Here i and j are two attributes and O means observed value and E means expected value and means value of chi-square. Higher the value of chi-square of an attribute higher will be its rank.

V. DATA SET DESCRIPTION

The dataset used for experimental evaluation consists of randomly selected instances from NSL-KDD Dataset [5]. It classifies network traffic in five categories. The number of each class instances included in training and testing dataset are mentioned in table-I.

Table I. Description of dataset

Class Type	Instances in Training Dataset	Instances in Testing Dataset
Normal	48522	1478
Dos	41699	37490
Probe	3608	8301
U2R	21	28
R2L	50	1075

VI. DISCUSSION AND RESULTS

All the experiments have been evaluated on selected instances from NSL-KDD Dataset using ten classification algorithms. These classifiers have been evaluated using WEKA Data mining tool Version 3.6.10. Six Ranking based Feature Selection Techniques have been evaluated using ten classification algorithms. Top 15 features from ranked features by feature selection techniques have been chosen for evaluation of one feature classification technique with below mentioned 10 classification algorithms.

Table II. Performance with 41 features

Name of Classifier	Performance Metrics		
	Training Time	**Accuracy**	**FPR**
1. Rotation Forest	342.81 sec.	96.4	0.001
2. Random Tree	0.57 sec.	96.14	0.002
3. Random Committee	6.53 sec.	96.11	0.002
4. Random Forest	6.29 sec.	96.12	0.002
5. IBk	0.04 sec.	96	0.002
6. Random Subspace	18.71 sec.	96.07	0.002
7. IB1	0.07 sec.	96.08	0.002
8. Part	20.16 sec.	95.4	0.001
9. JRip	69.25 sec.	94	0.003
10. NB Tree	195 sec.	92.25	0.005

Table III. Performance of Feature Selection Techniques With Top 15 Features

Feature Selection Technique	Performance Metrics	Rotation Forest	Random Tree	Random Committee	Random Forest	IBk	Random Subspace	IB1	PART	JRip	NB Tree
Chi-square	**Accuracy**	95.81	95.5	95.82	96.73	96.43	94.31	96.43	95.32	95	94.77
	FPR	0.002	0.007	0.003	0.002	0.005	0.005	0.005	0.003	0.002	0.006
	Training Time	227.75 seconds	0.63 seconds	5.11 seconds	5.21 seconds	0.04 seconds	8.91 seconds	0.03 seconds	6.12 seconds	51.73 seconds	47.85 seconds
Gain Ratio	**Accuracy**	97.8	96.65	96.92	96.5	97.22	95.65	97.22	96.22	93	96.34
	FPR	0.001	0.003	0.001	0.001	0.003	0.002	0.003	0.003	0.006	0.003
	Training Time	277.61 seconds	0.41 seconds	4.81 seconds	3.74 seconds	0.02 seconds	9.88 seconds	0.02 seconds	6.21 seconds	68 seconds	30.28 seconds
One R	**Accuracy**	96.27	96.87	96.69	95.68	97.21	88.78	97.22	95.61	94	94.11
	FPR	0.002	0.003	0.003	0.003	0.003	0.005	0.003	0.004	0.003	0.006
	Training Time	290.99 seconds	0.26 seconds	3.15 seconds	3.27 seconds	0.03 seconds	4.56 seconds	0.02 seconds	2.4 seconds	51.44 seconds	14.35 seconds
Info Gain	**Accuracy**	96.16	95.45	96.22	96.52	97.26	95	97.3	96	95.22	95.44
	FPR	0.002	0.003	0.004	0.002	0.003	0.002	0.003	0.002	0.002	0.003
	Training Time	267.8 seconds	0.42 seconds	3.64 seconds	4.4 seconds	0.04 seconds	7.91 seconds	0.04 seconds	6.05 seconds	31.2 seconds	43 seconds
Relief F	**Accuracy**	96.95	97	95.58	95.61	95.94	95	95.94	94.83	94	95.56
	FPR	0.001	0.002	0.003	0.002	0.002	0.003	0.002	0.002	0.003	0.004
	Training Time	402.55 seconds	0.54 seconds	5.17 seconds	4.65 seconds	0.03 seconds	10.1 seconds	0.03 seconds	5.83 seconds	54.35 seconds	53.4 seconds
Symmetrical Uncertain	**Accuracy**	98	96.87	96.69	95.68	97.22	88.78	97.22	95.61	94	94.11
	FPR	0.001	0.003	0.003	0.003	0.003	0.005	0.003	0.004	0.003	0.006
	Training Time	220 seconds	0.23 seconds	3.67 seconds	3.69 seconds	0.02 seconds	5.33 seconds	0.02 seconds	2.21 seconds	46.75 seconds	13.13 seconds

VII. CONCLUSION

In this work, the performance of NSL-KDD Dataset having 41 features has been evaluated on ten classifiers. Then the Feature set was reduced using six Ranking based Feature selection techniques. Top 15 features from each feature set produced by six feature selection techniques having higher rank are chosen and evaluated using ten classification algorithms. Finally the performance of the dataset having 41 features and the six reduced feature sets consisting of 15 features in each set was evaluated. It can be observed rom results that with reduced number of features, there is increase in performance of the classification algorithm. Further this work can be extended by evaluating performance of various combinations of Feature selection techniques using Filter or Hybrid Model.

REFERENCES

[1] R. Dash, Selection of the Best Classifier from Different Datasets Using WEKA, IJERT, Vol.2 Issue 3, March 2013.

[2] H. Nguyen and D. Choi, Application of Data Mining to Network Intrusion Detection: Classifier Selection Model, @Springer Verlag Berlin Heidelberg, 2008.

[3] M. Panda and M. Patra, A Comparative Study of Data Mining Algorithms For Network Intrusion Detection, IEEE, First International Conference on Emerging Trends in Engineering and Technology." 2008.

[4] M. Panda and M. Patra, Ensembling Rule Based Classifiers for Detecting Network Intrusions, IEEE Conference on Advances in Recent Technologies in Communication and Computing, 2009.

[5] B. Neethu,, Classification of Intrusion Detection Dataset using machine learning Approaches, IJECSE,2013.

[6] S. Garcia and F. Herrera, An Extension on "Statistical Comparisons of Classifiers over Multiple Data Sets" for all Pairwise Comparisons", Journal of Machine Learning Research 9, 2008.

[7] M. Othman and T. Yau, Comparison of Different Classification Techniques using WEKA for Breast Cancer, 2012.

[8] Kdd cup 99 intrusion detection data set..Online Available:http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

[9] L. Kuncheva, "Combining Pattern Classifiers: Methods and Algorithms (Kuncheva, LI; 2004) [book review]." Neural Networks, IEEE Transactions on18.3 (2007): 964-964.

[10] P. Sangkatsanee, N. Wattanapongsakorn and C. Charnsripinyo. "Practicle Real- time intrusion detection using Machine learning approaches." Computer Communications (2011): 2227–2235.

[11] S. Yang, K. Chi Chang , H. Wei and C. Lin. "Feature weighting and selection for a real-time network intrusion detection system based on GA with KNN." Intelligence and Security Informatics (2008): 195-204.

[12] S. Mukherjee and N. Sharma "Intrusion Detection using Naive Bayes Classifier with Feature Reduction" Procedia Technology 4 (2012) 119 – 128.

[13] Sajja, A. A, Knowledge Based Systems. Jones and Bartlett, 2012.

[14] Rich, E., Artificial Intelligence (3rd Ed.). Tata Mc Graw Hill, 2013.

[15] Pang-Ning, T. V., Introduction to Data Mining. Pearson. 2013.

[16] Alpaydin, E., Introduction to Machine Learning (2nd Ed.). PHI, 2010.

[17] Flach, P., Machine Learning the Art and Science of Algorithms that Make sense of Data. Cambridge, 2012.

[18] J iawei Han, M. K., Data Mining Techniques and Concepts (3rd Ed.). Morgan Kauffman, 2013.

[19] Ian H. Witten, F. E., Practical Machine Learning Tools and Techniques (2nd ed.), 2012.

[20] V. Labatut and H. Cherifi, Evaluation of Performance Measurse for Classifiers Performance.

[21] Liu H, Motoda H, Setiono R. & Zhao Z., Feature Selection: An Eve Evolving Frontier in Data Mining", JMLR: Workshop and Conference Proceedings Vol.4, *Publisher:* Citeseer, pages 4-13, 2010.

[22] Vege, Sri, H., Ensemble of Feature Selection Techniques for High Dimensional Data (Published Master's Thesis). Western Kentucky University, 2010.

[23] Y. Wang, F. Makedon, Application of ReliefF feature filtering algorithm to selecting informative genes for cancer classification using microarray data, Computational systems bioinformatics conference, 2004 IEEE, pages 497 – 498.

[24] S. Pulatova, Covering (rule-based) algorithms Lecture Notes in Data Mining., World Scientific publishing Co, pages 87-97, 2006.

[25] D. Ienco, R. G. Pensa, R. Meo, Context-based Distance Learning for Categorical Data clustering, LNCS 5772, Springer, Berlin, pages 83 – 94, 2009.

An Ontology Driven Multi-Agent Approach For Diabetes Management

Aditya Khamparia
Department of Computer Science and Engineering
Lovely Professional University
Phagwara, India
khamparia.aditya@gmail.com

Babita Pandey
Department of Computer Application
Lovely Professional University
Phagwara, India
shukla_babita@yahoo.co.in

***Abstract*—Agents can be considered as self-contained autonomous software component used to fulfill user request in any domain. In this paper collaboration among several agents and problem of inconsistency among them is to tackle with help of ontological descriptions for diabetes management. This paper also outlines new approaches to dealing with issues faced by individual agents through development of an ontology driven multi-agent system.**

Keywords—Ontolog, Agent,SPARQL, Onto grap, OWL.

I. INTRODUCTION

Semantic web is an extension of current web which creates intelligent based applications that will make web content meaningful to computers to support machine processing capabilities that automates web application and services [1]. Semantic web is a vision the idea of having data on the web defined and linked in a way that it can be used by machine not just for display purpose, but for automation, integration and reuse of data across various applications [W3C 2003]. It provides a platform for data integration, sharing and reusing from various sources. The domain knowledge is represented in terms of ontology's which have complete information about that particular domain. Recent years have seen a remarkable growing interest in agent based technology and application. An agent is self-defined autonomous software component capable of selection, discovery, composition and monitoring of web services [4]. Ontology can play an essential role in design of multi agent systems. This paper examines the use of ontology-driven multi-agent systems in diabetes management. Diabetes treatment entails limited contact of patients with multiple health care professionals such as Regular Practitioners (RP), Medical Specialist (MS) and Clinician. Interaction among them is achieved by patient frequent visit to RP who collects blood samples to send to laboratory for analysis. After getting results from testing, it is reviewed by MS to determine patient condition to prescribe medication. Various problems can be identified from current system like results achieved by practitioners and specialist from laboratory performs malfunctioning, relationship among health entities become inconsistent due to lack of collaboration, unawareness of patient regarding disease symptoms due to lack of speed communication from practitioners. This paper aims to overcome the above problems by developing ontology based system to promote interaction among all health care entities like practitioners, specialist, reports outcome to provide efficient treatment to patient and enable consistency for getting updates regarding symptoms, medication to patient. Ontology based graphical tool Protégé [10] and SPARQL [11] used to retrieve information from database.

II. MULTI-AGENT SYSTEMS AND EXISTING TECHNOLOGIES

A multi-agent system is an organization of software agents, collectively performing a task which cannot be performed by any individual agent [6]. Multi-agent systems facilitate interoperability of heterogeneous systems. The idea is to agentify the heterogeneous components to wrap these components with an agent layer that enables them to interoperate with each other via uniform agent communication language [8]. Various agent based languages like AgentTalk, Java, KQML used to work on toolkits and platforms like JADE, JACK, Aglets, MadKit etc.

A. Role of Ontology in Multi-Agent Systems

Ontology provides a formal semantics that can be employed to process and integrate information on web [7]. Gruber [3] describes ontology as an explicit specification of conceptualization. Ontologies play an important role in providing vocabulary which provides formal support for communication between software agents with declarative knowledge representation approach. It is essential to agent communication and coordination with help of ontology relationships.

B. Ontological Description of Multi-Agent System for Diabetes Management

Multi-agent systems provide a framework to help practitioners, specialists, clinicians and patients to interact with each other effectively. Interaction between agents involve as patient consults doctor or practitioner as symptoms found, then specialist formed treatment plan based on patient information, practitioners suggest some treatment plan and finally both come to one common treatment i.e. negotiating ideal treatment plan for patient. Each agent engaged in flexible and robust communication among systems across different platforms through wrappers and communicates around them with help of agent communication language. Consistency within heterogeneous environment can be maintained by using ontology. Ontology is pictorial representation of domain knowledge. Intelligent web tools like Protégé, Jena, Jade used for development of agent profiles in languages like RDF and OWL [9]. Protégé is used to describe conceptual definition for diabetes domain. Classes are used to describe concepts in domain. In Fig. 1 Diabetes represents main class concept containing several subclasses such as person, symptoms, assessment, treatments to create class hierarchy. After knowledge base is developed, agent can access attributes associated with class members and relationship to reason and derive answer to queries. The ontological graph for diabetes management is shown in Fig. 2 That graph demonstrates relationship between subject and object. Onto graph contains complete information about diabetes domain in which isMedicatedWith, isMedicatedBy, isTreatedWith, hasSymptom are various object properties which are created between subjects and objects. Various Data Properties like first_name and Tablet_dosage are associated with their domain and ranges.

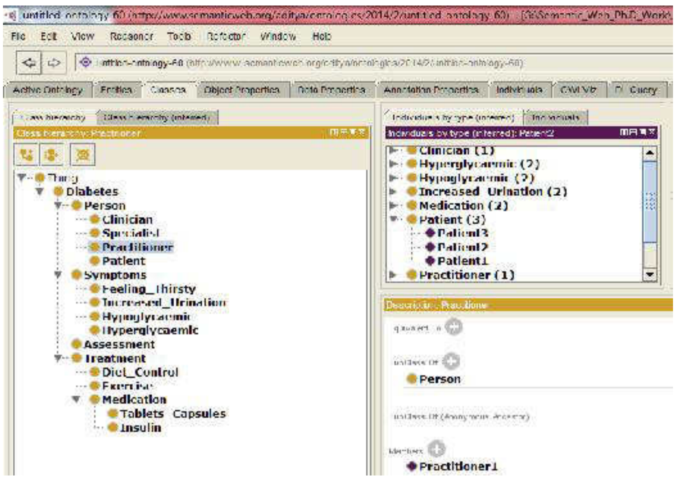


Fig.1. Class Hierarchy of Diabetes Management

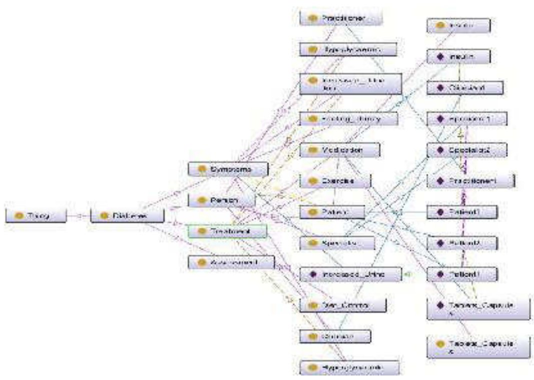


Fig.2. Onto-graph of Diabetes Domain

III ARCHITECTURE OF ONTOLOGY DRIVEN AGENT SYSTEM

The model of agent based diabetes system consists of following components like Diabetes agent system, Ontology agent, and Patient and Specialist agents.

Diabetes agent system provides an interface for patient agent, practitioners and specialist to send request from user to ontology agent and responds message from ontology agent. Ontology agent receives message from other agents in system and translates them into predefined ontology and performs query operation among all agent systems to determine which agent can satisfy message.

Specialist agent receives suggestions from RP's regarding treatment plan for patient and provides best possible treatment within low cost.

Patient agent responds to request for patient information and stored all information regarding patient behavior.

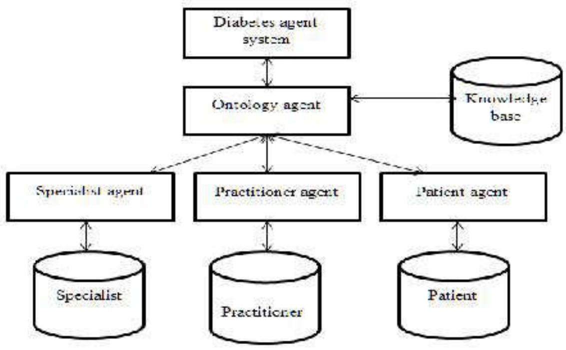


Fig.3. Ontology driven Multi-agent system

IV. QUERY ANSWERING USING SPARQL

SPARQL is an RDF query language that is able to retrieve and manipulate query data of diabetes domain based on querying of various agents. Various object and data properties with query editor can be expressed from diabetes ontology. Fig. 4, demonstrates how query retrieval occurs either it may related to symptoms, medication and treatment related with patient and specialists.

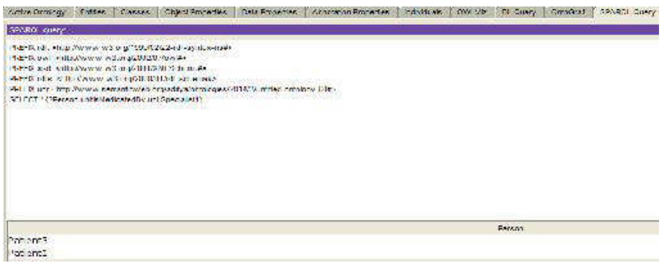


Fig.4. Query retrieval using SPARQL

V. CONCLUSION

This paper described an ontology driven multi-agent approach for diabetes management. Ontology editor, protégé is used to provide construction of domain ontologies and provides better flexibility for Meta modeling. The approach used multi-agent framework to increase collaboration among various individual agents and maintain their interoperability and completeness to facilitate knowledge discovery using SPARQL as query retriever. User friendly knowledge modeling application developed in protégé to determine effectiveness of individual agents in diabetes domain.

REFERENCES

[1] Berners-Lee, T., Hendler, J., and Lassila, O., 2001 the Semantic Web. Scientific American, 284:pp. 34-43.

[2] Falasconi, S., Lanzola, G. & Stefanelli, M, "Using Ontologies in Multi-Agent Systems". Proceedings of the 10th Knowledge Acquisition for Knowledge-Based Systems Workshop (KAW'96), 1996.

[3] T. R. Gruber. "A translation approach to portable ontology specifications", Knowledge Acquisition, 5:199–220, 1993.

[4] Khamparia Aditya, Pandey Babita. "Review on Semantic web services", WILKES- Proceedings of Elsevier conference, pp. 387-392, 2013.

[5] Wooldridge, M. "An introduction to Multi agent system", Wiley Ed. 2002.

[6] Sandeep Kumar, and R. B. Mishra, "Semantic web service composition" IETE Technical Review vol 25. No.3 May-June, pp. 105-121, 2008.

[7] Chandrabhan Singh, Sanjay Sachan and Mohit Gangwar, "MAS based Selection and Composition Process of SWS's for Medical Health Care Planning System" International Journal of Computer Applications 60(16):4044,December 2012. Foundation of Computer Science, New York, USA.

[8] Protégé 2000 – Integrated Ontology Development Tool. http://protege.stanford.edu.

[9] Sandeep Kumar and Mishra. "Multi-agent based semantic web service composition models", 2008.

[10] Stanford Center for Biomedical Informatics Research. Protégé, http://protege.stanford.edu/, Website accessed March 01, 2008.

[11] Stollberg, M., Haller, "A. Semantic Web Services Tutorial", 3rd International Conference on Web Services (ICWS 2005), 2005.

[12] Prudhommeaux, E., Seaborne, "A. SPARQL Query Language for RDF", http://www.w3.org/TR/2008/REC-rdf sparqlquery-20080115/, 2008.

[13] McGuinness, D.L, Harmelen, F.V. "OWL Web Ontology Language Overview", http://www.w3.org/TR/owl-features/, Website accessed on Feb 13, 2008.

[14] Shrivastava A et al. "Ontological Description and MAS based Composition Process of SWS's for Recruitment Domain", IJCA, 2013.

Modified Max Min Approach for Load Balancing In Cloud System

Rajwinder Kaur
CSE Deppartment
SBS State Technical Campus
Ferozepur, India
rajwindersandhu40@gmail.com

Navtej Ghumman
CSE Deppartment
SBS State Technical Campus
Ferozepur, India
navtejghumman@yahoo.com

***Abstract*— Cloud computing provide unlimited resources and services to the customers. Users can access these resources through internet directly. They need to pay only for those resources which they use. In cloud computing load balancing is very important technique. Max-min algorithm Provide load balancing in cloud computing. Modified max –min used the concept of original max - min. Modified Max-min is based on the execution time not on complete time as a selection basis. Load balancing algorithm provides better resource utilization and response time. New Max-min helps to achieving load balancing with lower makespan rather than original Max-min.**

Keywords— Cloud computing, load balancing algorithms, Max-min, Modified Max- Min algorithm

I. INTRODUCTION

Cloud service providers provide many services to end user. It provides all the services on internet with lower cost. It also provides some applications which users can use through internet. In this Applications are provide on lower cost. Users pay only for those resources which they used.[1].In this pay per use concept is used. In cloud environment resources are increase quickly. So load balancing is a major issue in cloud computing. In cloud many tasks are executed on available resources at same time. Load balancing is required for proper utilization of all the resources and for better response. Many algorithms are implemented for load balancing in cloud computing. All the algorithms work in different ways. In Cloud computing datacenters are used to collect all the resources; that all resources are shared by many users through the internet. Cloud computing is an internet based computing service that is provided by infrastructure providers, they provide the services on the users demand. In cloud computing Quality of Service and Load Balance (LB) is very important terms. So, to improve the performance of resources and for better results many effective task scheduling algorithm are used in cloud computing. Load balancing algorithm is responsible for managing all the jobs that are submitted to cloud environment. All the jobs are assigned onto available resources in such a way that the total response time minimized. The load balancing algorithms also help in manage the makespan. Many of these algorithms try to minimize the total completion time of tasks.. [2][3] Three well known algorithms which are used for load controlling in cloud computing environment are Max-min, Min-min and RASA In Max-min algorithm large tasks to be executed firstly, which means all small task delays for long time because they executed after the completion of long tasks. On the other hand, Min-min is executing smaller tasks firstly then large ones that mean long time tasks deal with the problem of delays. RASA algorithm provide solution for small and large tasks is reason for executing small tasks before large and it avoids delays of executing large tasks .It also support in execution of large and small tasks concurrently. Max-min strategy solves all the problems of Min-min, because it giving priority to large tasks as compared to small tasks. The Max-min algorithm selects the task with the maximum completion time and task with maximum completion time assigns it to the resource on which they take minimum execution time. It means the Max-min is very good choice when the numbers of small tasks are much more than large tasks. When numbers of small tasks are more than large ones then Min Min algorithm is a better choice. [4][5][6][7] RASA offers an improved load balancing algorithm based on Max-min to solve the mentioned above problems with both Max-min and Min-min algorithms.

The main idea of an improved Max-min algorithm is that assign task with maximum execution time to resource with minimum complete time at place of original Max-min assign task with maximum completion time to resource with minimum execution time.

II. CLOUD ARCHITECTURE

Cloud computing architecture consist cloud deployment model and different services .All the services have different layers. [7][8][9]

A. *Cloud Deployment Model*

There are four cloud deployment models as public, private, hybrid and community cloud. [10][11]

a) Public Cloud : Public cloud helps to users for access

Cloud publicly through interfaces. Web browser is a main interface which uses to access the cloud. This cloud computing infrastructure is made available to the general public.

b) Private Cloud : Private cloud is mainly used by a organization. Private cloud provides high level security. In provide more control for deployment of applications. Private cloud provides high level security. Intranet is a type Private cloud. In public cloud all the resources and applications were managed by the provider of services but in private cloud applications and services are combined together and made available for the users.

c) Hybrid Cloud: Hybrid cloud is a combination of public cloud and private cloud. In private and public work together. Hybrid cloud provides more security. It combines different services and applications and provide these services and resources to the customers through the internet. It allowed the organization to provide its needs in the private cloud and if some common requirements occur it provide the public cloud for computing resources and services.

B. Cloud Delivery Model

Cloud models provide three types of services Software as a Service (SAAS) and (PAAS) Platform as a Service and (IAAS) Infrastructure as a Service. All these services have different layers. All public, private, community, and hybrid clouds are implemented with the help of these services.[12][13][14]

a) Software As A Service : It provide many services to the users but user do not need to install the software. They directly access these services with interfaces like internet. All the installation of software and configuration of services are managed by cloud provider.

There are many advantages of SAAS but the main advantages are cost reduction. It reduce the cost because in this no need to buy the software. User directly used these software from internet on pay per used concept.. This is also removing licensing risk. SAAS have many advantages but there are some disadvantages also like billing management, synchronization of client requests.

b) Platform A Service: PAAS provide the platform to users to implement application according to their use. Cloud provider manages platform and servers, users only manage their applications. So applications can be created quickly without problems of buying and managing the server and software and hardware's. Google App Engine is a main example of platform of service.

c) Infrastructure As A Service: This service provide the Infrastructure of servers, many software, and network equipment to the users on their demand .These all the services are provided by the cloud provider. It is quickly fulfill the users demand.

III. LOAD BALANCING IN CLOUD COMPUTING

In cloud environment load balancing is required distribute the dynamic local workload evenly between all the nodes. Load balancing helps in better allocation of resources to achieve a high user satisfaction .It helps in implementing fail over and avoiding bottlenecks.

With the help of load balancing techniques data can be sent and received without any major delays. I n cloud computing environment many algorithm are available that helps in proper load balancing between all available servers. They can be applied in the cloud system with the suitable verification. Load balancing algorithms are divided into two types in cloud environment: first type of algorithm is Batch mode heuristic scheduling algorithms (BMHA) and second type is online mode heuristic algorithms

A. Metrics For Load Balancing Algorithms In Cloud

There are many matrices are used in load balancing .These all matrices are helps in measure the performance of algorithms -

- Scalability: It means algorithm is able to perform

when numbers of tasked are increased quickly in cloud environment. Any algorithm is good when this metric should be improved as compared to other algorithms.

- Resource Utilizations: It means all the resources are

Used properly. Better resource utilization It should be better for an efficient load balancing algorithm.

- Fault Tolerance: It provide recovery from all

types of failure. Example: node failure

- Response Time: It is the time that is taken by a

particular algorithm to respond a task. A good algorithm takes minimum time to respond a task.

B. Dynamic load balancing algorithm

Dynamic load balancing algorithms have two types: distributed and non-distributed algorithms.

*a) Distributed algorithms:*In this all nodes execute algorithm together, they divide load balancing work between them properly. All the nodes are interacting with each other in two ways: one is cooperative and other is non-cooperative. In the cooperative all the nodes work together but side-by-side to achieve a common goal. .All nodes work together to improve the response time of the system. In Non-Cooperative, all nodes work independently to achieve a common goal for better response time for all tasks.

b) Non-distributed algorithms: In this load balancing is done one or many resources. Non-distributed dynamic load balancing algorithms have two types: one is centralized approach and other is semi-distributed approach. In centralized approach, the load balancing algorithm is executed only by a single node. In this only central node is responsible for all operation and controlling. In semi distributed approach, all nodes of the system are divided into clusters. Then load balancing is done in centralized form in the clusters. A central node controls all the operations of load balancing. That central node is select by choice in each cluster during load balancing operations [15]

IV. MAX MIN ALGORITHM

In these algorithms first minimum execution time and minimum completion time for all the tasks are calculated.[16][17][18]

MCT (Minimum Completion Time): MCT assigns all tasks to those resources which complete them in minimum completion time. It means this assign some tasks to those resources that do not have minimum execution time.

MET (Minimum Execution Time): MET assigns each task to that resource which performs tasks in minimum time .But it is not considered that resources are available at that time or not.

Max-Min: Max-Min start with the set of all unassigned tasks in the makespan. This algorithm also works in two phases. First, the maximum expected completion time for all the tasks is calculated. The completion time for all the tasks is calculated on all the machines. In the second phase, the task with the maximum expected completion time from makespan is selected and that tasks assigned to the corresponding resource. Then the task which is completed that is removed from the makespan and this process is repeated until all tasks are completed.

The Max-min algorithm is mostly used in distributed environment. It starts with all unexecuted tasks. Expected execution time and expected completion time of each task on the available resources is calculated. Then select the task with overall maximum expected completion time and assign that task to the resource with minimum overall execution time.

Fig.1 represents basic Max-min algorithm, r_j represents the ready time of resource R_j to execute a task, Cij and E_{ij} represent the expected completion time and Execution time of task. Then task T_k with maximum expected completion time is selected .That task assigned to resource R_j that complete task in minimum execution time.

Finally recently executed task is removed from the meta-task set and then update all calculated times, then repeat until all the tasks are completed.

1. For all submitted tasks in meta task; T_i
2. For all resources; R_j
3. Cij = Eij + rj
4. While meta-task is not empty
5. Find task T_k consumes *maximum completion time*.
6. AssignT_k to the resource R_j which gives *minimum execution time*
7. Remove T_k from meta-tasks set

Fig.1 (Basic Max Min Algorithm)

V. MODIFIED MAX-MIN ALGORITHM

Max-min algorithm assign task T_i on the resource R_j where large tasks have highest priority and smaller tasks have lower priority. It means, when we have one long task, then Max-min algorithm could execute many short tasks concurrently while executing large task. The makespan is calculated in this by the execution of long task .It would be similar to the Min-min makespan.

We try to minimize waiting time of short jobs through assigning large tasks to be executed by slower resources. On the other hand execute small tasks concurrently on fastest resource to finish large number of tasks during finalizing at least one large task on slower resource. Where meta-tasks contain tasks with different completion and execution time, we proposed a new Max-min algorithm that helps in increasing the efficiency of max min algorithm. Improved max min increases the chances of execution of tasks on resources.

Max-min algorithm is followed to implement improved Max-min. Load balancing algorithms enhances performance in distributed systems. Sometimes these algorithms not help in better makespan. There are many existing load balancing algorithms in cloud computing which used for load balancing. Some new algorithms are also implemented from existing algorithms, this will helps to researchers to carry out further work in this area. Because many new modifications are required in load balancing.[19]

Fig.2 represents the modified max min algorithm. Modified algorithm work in different way as compared to basic max min algorithm.

1. For all submitted tasks in meta-task; T_i
2. For all resources; R_j
3. Cij = Eij + rj
4. While meta-task is not empty
5. Find task T_k costs *maximum execution time*.
6. Assign T_k to the resource R_j which gives *minimum completion time.*
7. Remove T_k from meta-tasks set
8. Update r_j for selected R_j
9. update Cij for all j

In the basic max min "Select task with max execution time then assign that task to that resource which take min completion time" and in improved max min it is changed" Select task with max completion time and then assign resource which take min execution time"

Fig.2 (Modified Max Min Algorithm)

VI. STIMULATE RESULT

Cloud computing simulation platform Cloud sim is used for implementation. It was designed by Australia Melbourne University. In CloudSim we used Java as discrete event simulation engine. It using the completion time and Bandwidth task scheduling simulation to validate the algorithm model. Create a group of tasks and also create a group of virtual machines. We extend planet lab and simulate result of more than 1000 cloudlets. But in this we take the result of 5 cloudlets only. [20][21]

In this algorithm first calculates the expected completion time of the submitted tasks on each resource. Then the task with the overall maximum expected execution time is assigned to a resource that has the minimum overall completion time. After execution scheduled task is removed from meta-tasks and all calculated times are updated and the processing is repeated until all submitted tasks are executed.

The algorithm focuses on minimizing the total makespan .That is the total complete time in cloud distributed environment. In cloud computing environment tasks are executed concurrently on available resources to achieving better load.

Table.1 shows virtual machine parameters like MIPS and bandwidth.

Table.2 represents task parameters like length and file size.

Table I. The VM parameters

VM ID	MIPS	BW
0	150	300
1	300	15
2	10	4

Table II The tasks parameters

CLOUDLET ID	LENGTH	FILE SIZE
0	256000	1300
1	35000000	31
2	323242342	96
3	22343242	590
4	350	3100

Table III.Simulation results of the Modified algorithm

CLOUDLET ID	STATUS	DATACENTER ID	VM ID	TIME	START TIME	FINISH TIME	PROCESSING COST
4	SUCCESS	2	0	4.67	0.1	4.77	35
0	SUCCESS	2	2	9.87	0.1	9.97	70
3	SUCCESS	2	1	17.47	0.1	17.57	114
1	SUCCESS	2	0	23.67	0.1	23.77	150
2	SUCCESS	2	2	34.07	0.1	34.17	255

Table3. Shows the result of modified algorithm.

Improved max min algorithm implemented from basic Max-min so it has the same time complexity (mn^2) .Improved max min produces better makespan with more reliable scheduling schema. Improved Max-min helps in better load balancing of available resources and also helps concurrent execution of submitted tasks with higher probability as compared to original Max-min algorithm.

VII. CONCLUSION

We have presented a modified max min technique performed by cloud Sim .Load balancing is used to obtain better resource utilization and performance .This study is concentrated only on tasks and resources. In future we will also consider cost model for this and try to improve cost.

REFERENCES

[1] Sahu, Yatendra and Pateriya, RK, Cloud Computing Overview with Load Balancing Techniques., International Journal of Computer Applications, Year 2013,Volume 65

[2] Mell, Peter and Grance, Tim, "The NIST definition of cloud computing", National Institute of Standards and Technology, 2009,vol53, pages50, Mell2009

[3] http://www.ancoris.com/solutions/cloudcomputing.html,"Cloud Computing |Google Cloud -Ancoris."

[4] S. S. MOHARANA, R. D. RAMESH, D. POWAR "Analysis of load balancers in cloud computing".volume{2},pages={101--108}, year={2013}

[5] A. Sidhu, S. Kinger, "Analysis of load balancing techniques in cloud computing", INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY 4 (2) (2013) pages737–741.

[6] http://www.qualitytesting.info/group/cloudcomputing/forum/topics/software-as-a-s, "Software as a Service (SAAS) - Quality Testing"

[7] http://letslearncloud.wordpress.com/"Cloud Computing | Learn Cloud and its tips.

[8] Chaudhari, Anand and Kapadia, Anushka, " Load Balancing Algorithm for Azure Virtualization with Specialized VM", algorithms,vol 1,pages 2, Chaudhari

[9] Nayandeep Sran,Navdeep Kaur , "Comparative Analysis of Existing Load Balancing Techniques in Cloud Computing ",vol 2,jan 2013

[10] Bala, Anju and Chana, Inderveer, "A survey of various workflow scheduling algorithms in cloud environment", 2nd National Conference on Information and Communication Technology (NCICT), 2011

[11] Chaczko, Zenon and Mahadevan, Venkatesh and Aslanzadeh, Shahrzad and Mcdermid, Christopher, "Availability and load balancing in cloud computing", International Conference on Computer and Software Modeling, Singapore, chaczko2011availability

[12] http://www.writeaprogram.info/c/os-programs/priority/"Priority Scheduling Algorithm Example - Write a Program.''

[13] R. Alonso-Calvo, J. Crespo, M. Garcia-Remesal, A. Anguita, V. Maojo, "On distributing load in cloud computing: A real application for very-large image datasets", Procedia Computer Science 1 (1) (2010) pages 2669–2677.

[14] S.-C.Wang, K.-Q. Yan, S.-S.Wang, C.-W. Chen, "A three-phases scheduling in a hierarchical cloud computing network", in: Communications and Mobile Computing (CMC), 2011 Third International Conference on,IEEE, 2011, pp. 114–117.

[15] Y. Hu, R. Blake, D. Emerson, "An optimal migration algorithm for dynamic load balancing", Concurrency: Practice and Experience 10 (6)(1998) pages 467–483.

[16] Etminani, Kobra and Naghibzadeh, M, A min-min max-min selective algorihtm for grid task scheduling, Internet, 2007. ICI 2007. 3rd IEEE/IFIP International Conference in Central Asia on, 2007.

[17] Liu, Gang and Li, Jing and Xu, Jianchao," An Improved Min-Min Algorithm in Cloud Computing" page{47--52}, year{2013}, organization{Springer}

[18] U. Bhoi, P. N. Ramanuj, " Enhanced max-min task scheduling algorithm in cloud computing". pages={2319--4847}

[19] H. Chen, F. Wang, N. Helian, G. Akanmu, "User-priority guided min-min scheduling algorithm for load balancing in cloud computing", pages={1--8},year={2013}, organization={IEEE}

[20] Buyya, Rajkumar and Ranjan, Rajiv and Calheiros, Rodrigo N," modeling and simulation of scalable Cloud computing environments and the CloudSim toolkit: Challenges and opportunities", pages{1--11},year{2009},{IEEE}

[21] O. Elzeki, M. Reshad, M. Elsoud, "Improved max-min algorithm in cloud computing, International Journal of Computer Applications"vol 50 (12) (2012)pages 22–27

Object Oriented Metrics Followed in Different Steps of Software Development: A survey

Samreet Kaur
Dept. of CSE
CTITR, Maqsudan

Jaspreet Kaur
Dept. of CSE
CTITR, Maqsudan

Sarabjit Kaur
Dept. of CSE
CTITR, Maqsudan

***Abstract*-The software has to follow different steps during its lifetime of development. The entire process of software development requires a great effort as this is the major phase through which we can improve the quality of our final product. So the main focus for the project developers is to improve the process. Object Oriented Measures or Metrics is an approach followed in order to achieve this goal, which has shown very good results in past studies. In this paper, various object oriented metrics suites are described that are used to measure the quality of the various phases encountered in software development i.e. design, coding. Various measures are there which helps in improving the quality of project. Different types of metrics cover different aspects of software which helps to evaluate the validation and quality of software. These metrics are under a set of standards through the use of which one can easily measure the effectiveness of Object Oriented Programming.**

Keywords-WMC, MOOD, DIT.

I. INTRODUCTION

Object- Oriented Analysis and Design is a technical approach for analyzing, designing an application, business and system through the use of Object Oriented Programming. It provides a visual aspect of modeling throughout the Software Development Life Cycles for enhancing the quality of software. But Object Oriented Analysis and Design is not an easier approach, so it is necessary to provide various guidelines that helps to perform good Object Oriented Programming and also helps in enhancing the quality of Software. Object-oriented metrics is an aspect used for this purpose. There are various classes of metrics that can be analyzed to check the effectiveness of software qualities. These are: Size(Size includes various factors like number of classes, number of operations, Dynamic object count, Length of the code which mainly takes depth of inheritance and functionality of classes.), Complexity(Complexity mainly includes the interaction between the various classes.), Cohesion(Cohesion refers to the degree to which the elements of a module being together.), Coupling(Coupling refers to the interdependence between modules.),Completeness(It describes whether the class reflect all the properties of a problem domain.), Inheritance(It is a mechanism of code reuse. It describes when an object is based on another object.), Encapsulation (It refers to one of two related but distinct notions of code.), Polymorphism (It allows the programmer to define different methods for any number of derived class.)

The main metrics suits which describe various measures of object oriented programming are:

1. Chidamber & Kemerer's Metrics Suite
2. Cohesion Metrics
3. MOOD/MOOD2 Metrics
4. Miscellaneous object-oriented metrics

1. Chidamber & Kemerer's Metrics Suite

The Chidamber & Kemerer's metrics suite consists of 6 metrics calculated for each class.

Metric 1.1: Weighted Methods per Class (WMC)

WMC is simply the method count for a class.

WMC = number of methods defined per class.

WMC describes how much time and effort is required to develop and maintain the class. A large number of methods used in a class means a greater potential impact on derived classes.

High WMC leads to more faults so keep WMC down by limiting the methods of class.

One way is to limit the number of methods in a class to specify that a maximum of 10% of classes can have more than 24 methods.

Metric 1.2: Depth of Inheritance Tree (DIT)

DIT= maximum inheritance path from the class to the root class The deeper a class is in hierarchy, more methods are inherited from it which makes it more complex. A high DIT has been found to increase faults. But it is not necessarily that the classes deepest in the class hierarchy have the most faults. The most fault-prone classes are the ones in the middle of the tree. Root and deepest classes are have less faults as compared to classes in the middle. An increase in DIT increases the density of bugs and decreases quality.

Metric 1.3: Number of children (NOC)

NOC equals the number of child classes derived from a base class.

NOC measures the breadth of a class hierarchy, whereas DIT measures the depth. NOC and DIT are related to each other. Due to increase in inheritance there is increase in depth and reduction in breadth. A high NOC, a large number of child classes have:

- High reuse of base class. Inheritance is a form of reuse.
- Base class may require more testing.

Metric 1.4: Coupling between object classes (CBO)

CBO= total number of classes to which it linked.

Two classes are linked when methods declared in one class use methods or instance variables defined by the other class. High CBO is not desirable. More coupling between object classes prevents reuse. Due to increase in coupling, the classes are mostly dependent on each other.

Metric 1.5: Response for a Class (RFC and RFC')

The response for a class is a method set that can be executed in response to a message received by an object of that class.

RFC = M + R (First-step measure)

RFC' = M + R' (Full measure) M = number of methods in the class

R = number of remote methods directly called by methods of the class

R' = number of remote methods that are called.

RFC measures the potential communication between the class and other classes. A large RFC can result in more faults. Classes with a high RFC are more complex which results difficulty in understandability of that class. Testing and debugging is also complicated for that class. An increase in RFC increases the density of bugs and decreases quality. RFC' should be preferred over RFC.

Metric 1.6: Lack of Cohesion in Methods (LCOM)

To understand LCOM, Take each pair of methods in the class. If disjoint sets of instance variables are accessed by them, increase P by one. Q is increased if disjoint sets share one variable.

LCOM = P - Q, if P > Q

LCOM = 0 otherwise

LCOM = 0 indicates a cohesive class.

LCOM > 0 indicates that the class needs to be split into two or more than two classes. Classes with a high LCOM have faults.

A high LCOM value indicates disparate nature of methods in the class. The LCOM metric can be used by senior designers and project managers as a relatively simple way to track whether the cohesion principle is adhered to in the design of an application and advice changes. There are four alternatives of LCOM.

2. Cohesion Metrics

Cohesion metrics measures the relationship between the methods of classes.

2.1. TCC and LCC (Tight and Loose Class Cohesion)

For TCC and LCC only visible methods are considered. A method is visible if it is not private, also if it implements an interface or handles an event.

TCC and also considers the number of methods which are related. If two methods access the same class level variables then these two methods are related. If two methods are related in this way then they are directly connected.

TCC and LCC mainly measures NP (maximum number of possible connections), NDC (number of direct connections), and NIC (number of indirect connections).

NP=N*(N-1)/2 where n is the number of methods.

NDC= number of edges in a connection graph.

NIC= number of indirect connections.

TCC= NDC/NP

LCC= (NDC+NIC)/NP

TCC<= LCC

Cohesiveness of class directly depends on TCC.

TCC is in the range of 0.1.

LCC is also in the range of 0.1.

TCC<= LCC.

When TCC = LCC =1 then it is said that the class is highly cohesive which means that the methods of class is connected.

3. MOOD and MOOD2 Metrics (Metrics for Object Oriented Design)

MOOD metrics are designed to provide summary of overall quality of an object oriented projects. MOOD metrics describes the overall project, not an individual module or class of project. MOOD metrics have six metrics in it.

3.1. Metric: MHF (Method Hiding Factor) and AHF (Attribute hiding factor)

MHF describes how variables and methods are encapsulated in a class and AHF describes how attributes are encapsulated in a class. MHF and AHF describe the average amount of hiding among classes in a system. The private methods are the hidden one. Increased MHF decreases bug density and increase quality.

MHF= 1-Methods Visible

AHF= 1-Attributes Visible

Methods visible = sum (MV)/C-1/Number of methods

MV= the classes where the method is visible.

Attributes Visible= sum (AV)/C-1/number of attributes

AV=number of other classes where methods is visible.

C= number of classes.

3.2. Metric: MIF (Method Inheritance Factor) and AIF (Attribute Inheritance Method)

A class that inherits lots of methods/attributes from its parent classes results to a high MIF/AIF. An independent class that did not have children results to a lower MIF/AIF.

Low values of MIF/AIF indicate lack of inheritance.

MIF= Inherited methods divided by the total methods in classes.

AIF= Inherited attributes divided by total attributes in classes.

3.3. Metric: Polymorphism Factor

It is equal to the actual method overrides divides by the maximum number of possible method overrides. It mainly measures the degree of method overriding in the class.

An increase in PF leads to decrease bug density and increased quality.

PF= overrides by sum for each class (new methods *descendents)

3.4. Metric: CF (Coupling Factor)

It measures the actual coupling among classes in ratio to maximum number of possible couplings. For example class A is coupled to class B if A calls methods or accesses variables of B. Coupling increases complexity, reduce encapsulation and limit understandability and maintainability. If two classes are coupled through a standard module the coupling is not considered. Very high value of CF also avoided. It is suggested that CF should not exceed 12%.

CF= Actual coupling divided by the maximum possible couplings.

MOOD2 Metrics:

Metric: OHEF and AHEF (Operation/Attribute Hiding Effectiveness Factor)

OHEF is related to MHF. MHF measures the level of method hiding and OHEF Measures the hiding succeed.

OHEF= Classes that do access operations divided by the classes that can access operations.

AHEF= Classes that do access attributes divided by classes that can access attributes.

Metric: PPF (Parameterized polymorphic factor)

Parameterized class is a generic class. Generic classes encapsulate operations that are not specific to a particular data type.

PPF= Parameterized Classes by All Classes

4. Miscellaneous Object-Oriented Metrics

4.1. Simple Class Level Object Oriented metrics: It includes the following:

a. IMPL- It describes the total number of interfaces implemented by class.

b. WMCnp - It describes the total number of non-private methods defined by class or it ignores private methods.

c. WMCi - It describes the total number of methods defined and inherited by class.

d. VARS - It describes the number of variables defined by class.

e. VARSi- It describes the number of variables defined and inherited by class.

f. VARSnp- It describes the non-private variables defined in class.

g. EVENTS- It describes the number of variables defined by class.

h. CTORS- It describes the number of constructors defined by class.

i. CSZ - It describes the size of the class.

CSZ=WMC+VARS

j. CIS - It describes the class interface size.

CSZ=WMCnp+VARSnp

4.2. Class Hierarchy Metrics: It includes the following measures:

a. CLS- It describes the total number of classes

b. CLSa- It describes the number of abstract classes. An abstract class is a class that is not designed for intantiation at run-time.

c. CLSc- It describes the number of concrete classes. Concrete classes are opposite to abstract classes.

d. ROOTS- It describes the number of root classes. A root class is the class that inherits directly from system object.

e. LEAFS- It describes the number of leaf classes.

f. INTERFS- It describes the number of interfaces.

g. maxDIT- The maximum depth of inheritance tree.

4.3. Class Reuse and Specialization:

Reuse ratio U= Superclass / Classes

U= (CLS-Leafs) / CLS

Higher U indicates the deeper class hierarchy.

A high value of U indicates a deep class hierarchy with high reuse. Reuse ratio varies in the range 0 <= U < 1. When U=0, there is no inheritance.

Specialization ratio S= Subclasses / Superclasses

S= (CLS-ROOTS) / (CLS-LEAFS)

S is the average number of derived classes for base class. When inheritance is used, S >= 1. The deepest and narrowest inheritance tree happens when S=1 (chain form tree with exactly one root and one leaf).

A superclass is not a leaf class.

A subclass is not a root class.

4.4. MPC (Message Passing Coupling):

MPC describes the number of messages sent out from a class or the number of procedure calls.

MPC= number of defined send statements in a class.

5. Summary of Metrics

The table below summarizes the metrics discussed above. It illustrates, in general, whether a high or low value is desired from a metric for better code quality. However, one still must exercise judgment when determining the best approach for the task at hand.

Table I. Summary of Metrics

Metric	Desirable Value
Coupling Factor	Lower
Lack of Cohesion of Methods	Lower
Response for a class	Lower
Attribute Hiding Factor	Higher
Method Hiding Factor	Higher
Depth of Inheritance Tree	Low (tradeoff)
Number of Children	Low (tradeoff)
Weighted Methods Per Class	Low (tradeoff)
Number of Classes	Higher
Lines of Code	Lower

II. CONCLUSION

This paper describes the basic metrics suits for object oriented design. The use of various metrics helps the software designer to obtain good project. Projects performances are compared through these metrics. Following a particular manner earlier in designing a project can lead to a reduction in costs of the overall implementation of the project. The developers can easily conclude the performance of its project through the early use of these metrics and also improve the quality of product. The various types of metrics that can be describes in this paper may or may not suits to all languages. Different languages follow different metrics to improve their project quality. It should be also kept in mind that these metrics are not the rules, these are just guidelines which are not mandatory to use in every language.

REFERENCES

[1] F.B. Abreau and Melo, “Evaluating the impact of object oriented design on software quality”, Proc. 3rd International Software metrics Symposium (Metrics96), IEEE, Berlin, Germany, March 1996.

[2] A. Shaik, C.R.K. Reddy, and A. Damodaram, “Object Oriented Software Metrics and Quality Assessment: Current State of the Art”, International Journal of Computer Applications, vol. 37, no.11, Jan. 2012.

[3] A. Rana, “A Comprehensive Assessment of Object-Oriented Software Systems Using Metrics Approach”, International Journal on Computer Science and Engineering, vol. 02, no.

[4] Ralf Reißing, Towards a Model for Object-Oriented Design Measurement.

[5] Jaana Lindroos, Code and Design Metrics for Object-Oriented Systems, 2004

[6] R. Fichman and C. Kemerer,”Object-Oriented and conventional analysis and design methodologies: comparison and critique”, IEEE Comput.,vol.25, pp. 20-39, 1992.

[7] R. Shatnawi, “A Quantitative Investigation of the Acceptable Risk Levels of Object-Oriented Metrics in Open-Source Systems”, IEEE Transactions Software Engineering, vol. 36, no. 2, pp. 216-225, 2010.

[8] S. Kaur, “A Quantitative Investigation of Software Metrics Threshold Values At Acceptable Risk Levels”, International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 2 Issue 3, March 2013.

A Framework to Design and Update University Information Security Policy

Vikas Khullar
Department of Computer Science and Engineering
CT Institute of Engineering, Management and Technology
Shahpur, Jalandhar, India.
vikas.khullar@gmail.com

Abstract **-- Security is a complicated subject, historically only tackled by well-defined rules or policies. In this paper the main focus is on to purpose "A Framework for Analyzing and Designing University Information Security Policy" which is able to protect university resources, it is important to understand what there is to protect and what to protect it from. In the beginning of the paper presentation of the general introduction about the Information Security, CIA Trade, Information Classification, Cryptography, Information Security Attacks, Vulnerabilities and how to deal with attacks. At last a framework is proposed which helps to analyze and design the information security policy for universities. This framework is simulated and tested with an example. Finally, the paper has been concluded and future directions have been given.**

Keywords—Framework, Policy, Security, Cryptography, Authorization, Authentication and Access Control (AAA), Malicious code.

I. INTRODUCTION

In this modern technology, computer has been part of lives. Wherever we go, whatever we do, we deal with computers. Thus computer plays a vital role in our daily routines. Its make our lives faster and easier. In today's time, nothing is possible without the use of computers. They have their own importance in education, traditional, scientific, researches, mathematics, sports, entertainment, medical and many other aspects. Distributed and centralized systems are used in Universities these days.

Now major part of the data handling on database levels, stand alone system level and on transmission or network level. In the accounts department, very crucial information is stored or processed like fee submission, staff pay and their savings. Whereas examination section (it holds information about results, question papers, examination duty schedule) is also sensitive part of university. The library is another crucial section, holds information about books, journals, and research papers. The problem is arises when data about question papers are leaked out, received fee data is corrupted, or library data is misused by any unauthorized user either intentionally or by any other means.

The information security team is responsible for all these problems. They should apply constraints, generates policies and implement them to stop these vulnerabilities. Security is required because most organizations can be damaged by hostile and intruders. There may be several form of damage like:

• Damage or destruction of computer system.

• Damage or destruction of internal data.

• Loss of sensitive information by hostile parties.

•Use of sensitive information against the organization's costumers which may result in legal action by costumer against the organization or loss the costumer.

• Damage to the reputation of the organization [1].

So securing the data from unauthorized access that is confidential data remain secure from the user which can misuse that data can be done by using some security mechanisms.

The security mechanisms that can be used in common are Cryptography, Authorization, Authentication and Access Control (AAA), Implementation of rules and policies. By using the security mechanisms, data can be secured from unauthorized users. But there should be set of laws, rules, and practices that regulated how an organization manages, protects and distributes sensitive information or data which is known as security policy. Security policy should define what is to be protected and what is to be expected from the system users[2].

Establishing a security policy is the starting point in designing a secure computer network. It is essential that a set of minimum security requirements be gathered, formalized and included as the basis of your security policy [3]. This security policy must be enforceable by your organization and will create an additional cost to running and monitoring your network. This additional cost/benefit of a security policy must be understood and embraced by your organization's management in order to enhance and maintain network and system security. The lack of an accepted and well-thought-out security policy and guidelines document is one of the major security vulnerabilities in most companies today.

II. INFORMATION SECURITY POLICY

An IT Security Policy:

- Communicates clear and concise information and is realistic.
- Includes defined scope and applicability.
- Makes enforceability possible.

The information security policy is a direction document for information security within an organization. A policy is an expression of intent. A written policy is the primary means by which a board or executive team gives direction to management and staff, and informs other stakeholders. Effective security policies provide clear direction and commitment, and establish the responsibilities. A security policy defines the rules that regulate how an organization manages the information and computing resources to achieve security objectives. One of the policy's primary purposes in detecting signs of intrusion is to document important information assets and the threats to those assets that an organization chooses to address. Preparation procedures include the actions necessary to observe systems and networks for signs of unexpected behavior, including intrusion. Observation can take the form of monitoring, inspecting, and auditing. From these procedures, all concerned parties are able to determine the operational steps they need to take to comply with the policy [4], [5].

These steps will thereby uphold the security of an organization's information and networked systems. Security policies and procedures that are documented, we visibly enforced establish expected user behavior and serve to inform users of their obligations for protecting computing assets. Users include all those who access, administer, and manage the systems and have authorized accounts on the systems. They play a vital role in detecting signs of intrusion [6], [7].

An IT Security Policy is the most critical element of an IT security program. A security policy identifies the rules and procedures that all persons accessing computer resources must adhere to in order to ensure the confidentiality, integrity, and availability of data and resources.

A. Factors of Good Information Security Policy

A good information security policy is compromised of several factors:

- A good information security policy is that it usable.
- It should be concise, clearly written and detailed as possible in order to provide information necessary to implement regulations.
- It should be versioned and dated in order to easily identify the most current document.
- It should be structured internally in such a way that pertinent or required information is identified and located easily within document for references.
- A good information security policy also takes into account the existing or implicit rules into use.
- The interest of employees, third parties and the business goals of the organization should always be considered in the information security policy [8].

III. PROPOSING A FRAMEWORK TO DESIGN AND UPDATE UNIVERSITY INFORMATION SECURITY POLICY

A comprehensive and systematic framework to Design and Update Information Security Policy for University is shown in figure 1. This framework acts like a cycle of pre-attack policy

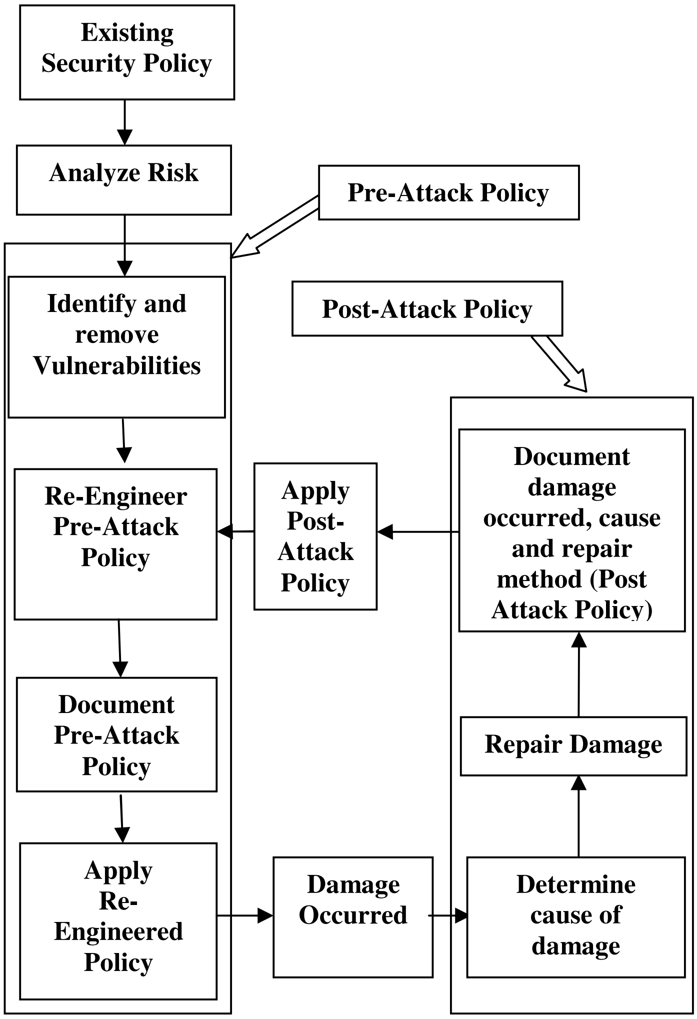


Fig.1 .A proposed framework to design and update university information security policy.

and post-attack policy. In this, pre-attack policy which is under infinite loop in which updating of existing policy in continuing that only exit when any damage occurred. When damage is come into vision, then that damage is repaired.

After repairing the damage, reason of occurred damage, its repair method and its cause is documented. Then existing pre attack policies is re-engineered by updating post-attack policy.

IV. STEP BY STEP SIMULATION OF THE FRAMEWORK TO DESIGN AND UPDATE INFORMATION SECURITY POLICY FOR UNIVERSITY

Step1. There is no any existing security pre-attack policy (Step 2 to 4 is Pre-Damage Policy).

Step 2: Analyze Risk

Threat observed, a malicious content may attack the system.

Step 3: Identify and Remove Vulnerability

After analyzing we find solution for malicious code attack that is antivirus.

Step 4: Re-Engineer Pre-Attack Policy and Document Re-Engineered Policy.

Create a malicious code policy. Now there is an information pre-attack security policy.

Malicious Code Policy

- Add antivirus

Loop back to step 2 and analyze the threat that may occurred

(Step 5 to 10 is Post-Attack Policy)

Step 5: Damage Occurred

Damage occurred (*.doc files corrupted)

Step 6: Determine cause of Damage

Find Anomaly in existing policy (Antivirus not updated)

Step 7: Repair Damage

Update Antivirus

Step 8: Document Damage occurred, Cause and Repair Method

Step 9: Apply Post-Attack Policy

Now we have post-attack policy.

Malicious Code Policy

- Update antivirus Periodically

Move to Step 4 and Update the existing Pre-Attack policy using Post-Attack policy

Step 10: Now the updated pre-attack policy is given below:

Malicious Code Policy

- Add antivirus.
- Update antivirus periodically.

V. CONCLUSION

It has been concluded from this paper that an effective information security is required so that important information should not be accessed by unauthorized users for any misuse. Security techniques and security policy should be used for information security. A security policy identifies the rules and procedures that all persons accessing computer resources must adhere to in order to ensure the confidentiality, integrity, and availability of data and resources. Most of the universities using the information security policies for handling very important data by applying basic mechanisms, but they were not using a solid information security policy for securing the data. A good security policy has many benefits. For designing a good security policy, there must be a good framework to analyze and design the policy. In the said frameworks divide the information security policy into pre-attack strategy and post-attack strategy. In this dissertation, a framework has been designed to enhance an effectiveness of existing information security policy or designing a new information security policy. After analyzing the proposed framework found that, both pre-attack and the post-attack policy depend upon each other. So, it has been concluded that the university information is not secure, it require proper security framework which analyze and design the university information security policies. This framework may also use for designing various security policies for different institutions.

REFERENCES

[1] "Network and Computer Security Tutorial", Version0.4.0, 2001,http://www.comptechdoc.org/independent/security/recommendations/

[2] "Understanding Authentication, Authorization, and Encryption",http://www.bu.edu/tech/security/resources/bestpractice/auth/,

[3] George L. Stefanek, "Information Security Best Practices, 205 basic rules", 2002.

[4] "Information Security Guidelines", Government Chief Information Office, GCIO Guidelines, Issue No 6 for NSW Government Agencies, 2007.

[5] Allegra Huxtable "Information Management Advice 35: Implementing Information Security", Department of Education, LINC Tasmania. 2013

[6] "IT Security Policy", CERT-In Indian Computer Emergency Response Team Handling Computer Security Incidents, Department of Information Technology, Ministry of Communications and Information Technology, Government of India, 2003.

[7] Julia Allen, Ed Stoner, "Detecting Signs of Intrusion", Carnegie Mellon Software Engineering Institute, 2000.

[8] SANS Institute InfoSec Reading Room, "Building and Implementing an Information Security Policy", 2002.

A Review on the Monitoring and Movement Detection System of Patient

Abhishek Bansal
Department of C.S.E.
Baba Farid College of Engg. & Tech.
Bathinda, India.
abhi1.bansal@gmail.com

Harleen Kaur
Department of I.T.
Baba Farid College of Engg. & Tech.
Bathinda, India
harleengrover@gmail.com

Abstract*–Detection of changes due to movement in a real time video is very important tool. Its importance is found in security systems, video surveillance, patient monitoring and logging etc. Patient monitoring systems mainly suffering from the false record tracking by the nurses manually in hospitals which leads to the incorrect treatment of patient and sometimes it leads to death of the patient. Due to less number of staff members in hospitals with respect to large number of patients it is impossible to take monitoring of all the patients at the same time. This paper reviews the monitoring system for capturing movement detection of the patient. By this we can capture the whole movement of patient which may not be cause any major problem. This paper provide helps to the researchers who are new in this field.

Keywords*–Patient monitoring systems, Movement detection , Motion capture, Vision based systems, Real time video.

I. INTRODUCTION

Patient movement detection system is a system that is used to detect the movement of patient. Patient movement and monitoring system still suffers from many kind of problems in local and foreign hospitals are:

(i) Less number of doctors and nurses in hospitals with respect to large number of patients which leads to the problem in keeping the record of every patient properly.
(ii) Reports written on paper which leads to the loss of information for future.
(iii) Reports generated by the 3G mobile phones and VGA cameras will give bad results which leads to the wrong treatment.
(iv) Lost of the database, when the record about he patient is saved automatically in the database of the system.

Above problems can be overcome to some extent with following steps.

(1) By providing the systematic and effective system for patient medical record, data and information about a particular patient can be saved for a long period at least 5 years for future need if that patient come again for check-up at that hospital.

(2) By integrating the system with a database management system, information about a particular patient including his present situation and progress, health information etc. can be saved properly in a database management system and can also be accessed easily by the authorized person when needed in the future. This system will also reduce the human errors and loss of information because it will provide a backup system.

(3) By reducing the time taken in searching for a particular patient medical record. It will take only few keywords in searching the medical record of a particular patient .

II. OVERVIEW OF PATIENT MONITORING SYSTEMS

In this section, we are presenting some of the research work of the prominent authors in this field and will be discuss various methods in which various patient movement monitoring systems are presented.

A. Human Motion Capture

Masashi Uchinoumi et al , "A simple structured real time motion capture system employing silhouette images", describes a novel system for human motion capture having simple architecture/structure with some accelerated computation techniques. These techniques contribute in the recovery of real time motion capturing. These techniques include use of a server-client system for computation, use of a look up table to realize quick search and use of difference images for volume recovery. The algorithm along with the simple structure works very well. One of the main advantages of a client-server system is that it is a distributed computation system.

Its significance is that the motion capture system having a simple structure architecture. The approach used is based on back projection of images with accelerated computation technique and it results to produce a better motion capture system.

B. Image Change Detection

Qiang Liu et al , "A DCT- Domain Approach to Image Change Detection and Its Application to Patient Video Monitoring ", Presents a medical application of patient monitoring via video recording, which is commonly used in hospitals for clinical diagnosis . Automatic detection of the movements of patient in the video may facilitate this application in the following aspects.

1) Detecting unusual changes and behavior in a scene so that an alarm can be whistled in case of an unusual event. 2) Controlling the camera automatically to present a neat and clean view of the object. 3) Organizing the video data according to the movement activity of the patient. 4) Compressing the video contents to reduce the size of the data.

This paper deals with the application of patient video monitoring via video recording with a novel approach in the discreet cosine transform (DCT) domain. This approach helps in testing that whether a pixel or a region between two images is significantly different or not. In this paper, a DCT domain approach is used to identify changes between two images . The application of change detection are broad including video survilliance , traffic assistance etc. Its Significance is that Noise disturbances are excluded from the meaningful changes. The Approach used is DCT Approach (Discreet Cosine Transform). And it Results is Robust.

i. Electronic medical record for patient

Shihab A. Hameed et al , "Electronic Medical Record For Effective Patient Monitoring Database" , deals with many kinds of problems that the patient monitoring system still suffer from such as 1) lost of information when data are recorded automatically from the hardware patient monitoring system into the database. 2) Blur images usually taken by the mobile phone which give incorrect or inaccurate information to the doctors in emergency cases. 3) Less number of doctors and nurses with respect to the large number of patients in both local and foreign hospitals leads to difficulties in recording all the medical information properly. This research paper is intended to overcome most of the problems faced in medical system. This research paper is focusing on developing Electronic Patients Medical Records System that will help in making patient monitoring system more effectively. This paper also deals with the remote patient monitoring system using java enabled mobile phone. Remote patient monitoring system is used to monitor information of multiple patients in Intensive care unit (ICU), Cardiology Care unit and other units in the hospital by using 3G mobile phones. By using Remote patient monitoring systems it is easy to send data to the doctor and helps the patient to monitor by the doctor.

Its significance is to overcome problems such as: lost of information and less hospital staff w.r.t large no. of patients. The approach used is EPRM (Electronic Patient Medical Record) and it results to the suitable for searching patient medical records easily with few keywords.

a. Wearable motion capture suit with full body tactile sensor

Yuki Fujimori et al, "Wearable Motion Capture Suit with Full Body Tactile sensors", Presents a system that is used for the extraction of human movement and methods for analyzing the different kinds of movements. We simply cannot extract the movement of whole body and the lack of information about the movement sometimes leads to some major problems. For better understanding of motion behavior, wearable motion capture suit with full body tactile sensors are developed. In this paper, a motion sensor is developed which can estimate its orientation with its inner CPU, and build a tactile sensor module which can fit many kinds of body shapes. In this we developed a motion capture system which can measure human motion and tactile information simultaneously, with indication of many possibilities of motion with tactile data.

Its significance is that it is used for better understanding of motion behavior. The approach used is wearable motion capture suit with full body tactile sensors and it results to the measure human motion and tactile information simultaneously.

b. A vision based System for ellery patient monitoring

Francesco Cardile et al , "A Vision Based System for Elderly Patients Monitoring", presents a remote based computer vision technology. Remote based computer vision technology improves the quality life of elder and impaired people with decreasing costs. Among the many investigated technologies, Computer vision technology has identified to be a very effective in many important scenarios where conventional the sensors fail. This paper talks about the aging population with an increasing number of elder people which is living alone or that are independent to their children and helps in improving the quality of life of elder persons. Because in old age, once a old man break his leg or foot by slipping and falling down: it would be very difficult for him to recover in old age.

Its significance is to improve the quality of life of elder people with reducing costs. The approach used is computer vision based wireless sensor system and it result is effective.

ii. Real Time Vision based respiration Monitoring Systems

K.S. Tan et al, "Real-Time Vision Based Respiration Monitoring System", presents a vision based contactless method to extract movement from the patient monitoring system. This method also involves image and signal techniques to extract movement in the chest and the information about the movements from a number of images recorded using a single video camera. This paper deals with the respiratory rate which is an important measure to examine the health of an individual used in various hospitals or clinical centers. Respiration rate can be contact or noncontact based. In contact based respiration monitoring system, the sensing device is attached to the subject's body, whereas in noncontact respiration monitoring system, the sensor is often placed much closed to the patient.

Its significance is to extract chest movement specially. The approach used: vision based contactless method and its results are Accurate, effective and easily implementable respiration monitoring system.

III. CONCLUSION

The main work of this paper is to understand the method of detecting of patients monitoring systems. This paper shows different methods by which we can develop the monitoring and detection systems by using various methods. This will help to the researchers to understand the existing systems used in monitoring and movement detection of patient.

REFRENCES

[1] Shihab A. Hameed, Shazana Mustapha et al, *"Electronic Medical Record for Effective Patient Monitoring Database"* , IEEE , 2008.

[2] Qiang Liu, Robert J. Sclabassi et al, *"A DCT- Domain Approach To Image Change detection And Its Application To Patient Video Monitoring, IEEE, 2005.*

[3] Mohamed Osman , " *Wireless ECG Monitoring System Project Report"* .

[4] K. S. Tan, R. Saatchi et al, *"Real-Time Vision Based Respiration Monitoring System"*, IEEE, 2010.

[5] Masashi Uchinoumi et al, "A simple structured real time motion capture system employing silhouette images", IEEE, 2004.

[6] Yuki Fujimori, Yoshiyuki Ohmura et al, *"Wearable Motion Capture Suit with Full-body Tactile Sensors"* , IEEE, 2009.

[7] Francesco Cardile, Giancarlo Iannizzotto et al, *"A Vision Based System for Elderly Patients Monitoring"* , IEEE, 2010.

[8] Richard J. Radke, Srinivas Andra et al, "*Image Change Detection Algorithms: A Systematic survey*" , 2005.

[9] Zhao jinmeng et al, " Key Techonologies Of Medical Monitoring System of Smart Home"IEEE,2011.

Load Balancing Techniques in Cloud Environment: A Survey

Amandeep Kaur
Computer Science & Engineering Department
Shaheed Bhagat Singh State Technical Campus
Ferozepur, Punjab,India
aman07.sbs@gmail.com

Navtej Ghumman
Computer Science & Engineering Department
Shaheed Bhagat Singh State Technical Campus
Ferozepur, Punjab,India
navtej.ghumman@yahoo.in

***Abstract*— The rapid development of Internet has given birth to a new business model: Cloud Computing This new paradigm has experienced a fantastic rise in recent years Cloud computing is the technology used in Building Internet-wide computing environment and Integrating Distributed and heterogeneous resources However, in Cloud Environment, job scheduling is confronted with a great Challenge. This paper focuses on jobs scheduling in Cloud Computing. To handle the multiple requests efficiently Scheduling of requests or jobs plays a vital role in Cloud services. This paper will show light on scheduling techniques that are being used by the cloud service provider to balance the load and its impact on the services also.**

Keywords*— *Cloud Computing, Virtualization, Scheduling, Load balancing.

I. Introduction

The concept of Cloud Computing was introduced by the arrival of Salesforce.com in 1999 that was first milepost in Cloud Computing. They instigate the Concept of delivering enterprise applications via a simple websites. The next development which spiked this field was introduction and the next development was Amazon Web Services in 2002, which provided a suite of cloud-based services including storage, computation [1]. Then in 2006, Amazon again spiked the world of cloud with the introduction of its Elastic Compute cloud (EC2) as a commercial web service that allows small companies and individuals to rent computers to run their own computer applications. Another big turning point came in 2009, as Web 2.0 hit its stride and Google and others started to offer browser-based enterprise applications, services such as Google apps [2].

Cloud computing is a type of distributed computing which is growing rapidly in the field of the computer applications. It makes use of highly scalable IT-related capabilities and providing "as a service" to the end user [3]. Its fastest growth is only due to its ease of use and facilities that is offering to the end users. It is emerging as a new software technology, using the concept of grid computing and distributed computing. We can see cloud computing as our next generation version of internet technology as it makes use of Online Storage, Online Platform and Online Infrastructure. It works on "pay as use" aim. This makes Profit to the organization as well as the user, they can access "as a service" that package as which they are not able to buy and they have to pay as per they Consumed. Cloud services are carried out on behalf of its customer on hardware that is totally managed by the cloud service provider. One way of thinking about cloud computing is to consider your experience with email. Your email client Gmail, Hotmail or any other takes care of housing all the necessary hardware and software to support your personal email account [17]. When you want to access your email you open your web browser, go to the email client, and log in. The most important part of the work is having internet access. Your email is not housed on your physical computer you access it through an internet connection, and you can access it anywhere. Cloud Computing is providing you the following main advantages and also these are the main reasons for choosing cloud as service provider [4].

- The scalability of clouds which enables clients to easily scale their environment both up and down to meet business needs.
- The end user need not to worry about the underlying hardware and supportive tool. They are all managed by cloud service provider.
- Resources are only paid for when they're used, so it's possible to reduce both capital and operating costs.
- Cloud computing offers a centralized, remote facility for computing, leading to economies of scale in both the use of hardware and software and a reduction in resources for administrative benefits.

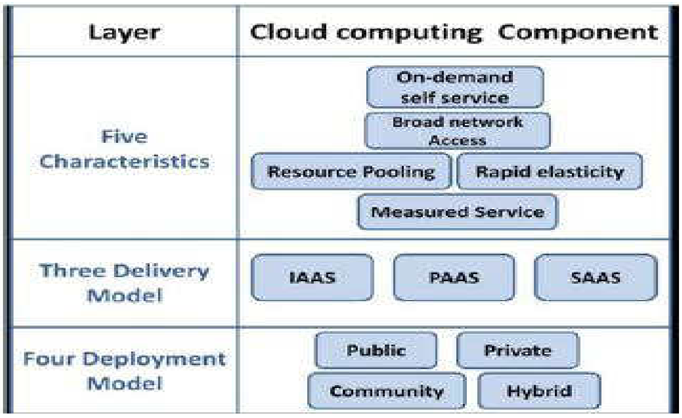


Fig .1. Model of Cloud Computing

A. Types of cloud provider

There are three types of cloud provider that one can subscribe to: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). Briefly, here is what you can expect from each type [5][16].

1. Software as a Service (Saas): A SaaS provider gives subscriber access to both resources and applications e.g. Google Docs, Dropbox. The consumer need not to manage or control the underlying cloud infrastructure including network, servers, operating systems, storage as it makes unnecessary for subscriber to have a physical copy of software installed on its device. SaaS provides the software to the subscriber as a cloud service after accessing the cloud once.
2. Cloud Platform as a Service (PaaS): A PaaS provider gives subscribers proprietary APIs to make an application that will run in a specific environment e.g. Facebook. Here the Developers can create specific applications for Facebook platform using proprietary APIs and makes that app available to any Facebook user.
3. Cloud Infrastructure as a Service (IaaS): IaaS gives subscriber the infrastructure for developing, running, and storing the application in cloud environments e.g. Amazon EC2.

B. *Deployment Model*

There are different types of clouds that you can subscribe to depending on your needs. As a home user or small business owner, you will most likely use public cloud services [5][6].

1. Public Cloud (External Cloud): Its computing resources dynamically provision to the internet by web services to the end user from an off-site third party provider. The infrastructure is shared and accessed at low cost to all customers of the organization. Example: eBay [17].
2. Private Cloud (Internal Cloud): It is mainly for the purpose of own private cloud by the specific organization whose services are just dedicated to that user of organization. So it does not utilize shared infrastructure. This type of cloud is suited where high level of security is required. Example: Amazon, Google Apps, Windows Azure [18].
3. Hybrid Cloud: It inherits both the features of Private and Public Cloud. It combines multiple private and Public Cloud. The Cloud infrastructure could be hosted by a third-party vendor or within one of the organizations in the community [8].
4. Community Cloud: When some organization constructs common Cloud having same policies, requirements, values and concerns then they may use dedicated or off site data center at their different communities [14].

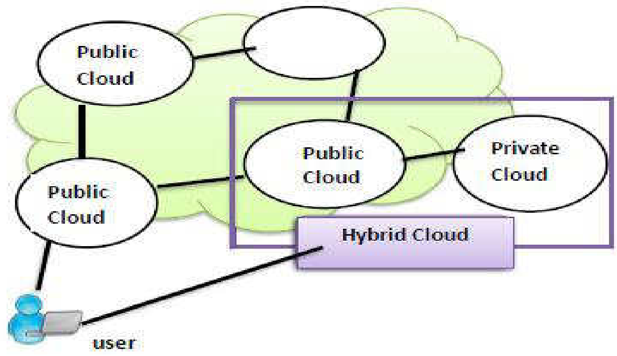


Fig .2. Different Deployment Models of Cloud

II. LOAD BALANCING TECHNIQUES

Nowadays every Organization is moving towards cloud computing, that day is not far when there will be a lot of cloud users.

As a cloud service provider who is responsible for managing and proffer Cloud services to its subscriber. Scheduling of users request is managed by cloud Scheduler [7]. Cloud Scheduler focuses on maximum utilization of resources, because without utilization there is no worth of having maximum resources [20].

To manage the load, Load balancer makes use of some load balancing techniques. Load Balancing is a technique in which the workload on the resources of a node is shifted to respective resources on the other node in a network without disturbing the running task. Task scheduling in a distributed system usually has the work of distributing the loads on processors, maximizing their utilization, while minimizing the global task execution time in which they use different scheduling techniques that help them to manage the resource demand quickly [8]. There are mainly two types of load balancing techniques used to manage load:

- Static Type: It is also called pre-scheduling technology to schedule known tasks in foregone environment [10].

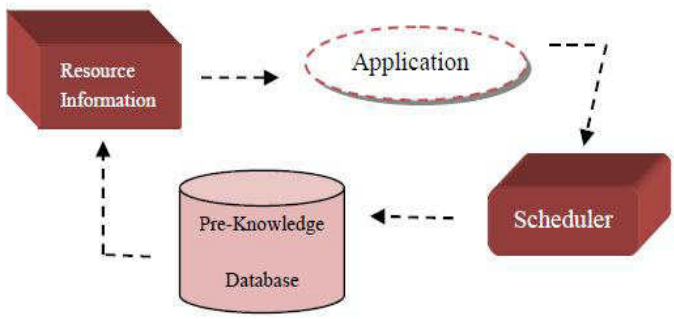


Fig 3. Static Load Balancing

- Dynamic Scheduling: It is used in real life because a cloud system is dynamically changing over time [9][8]. Task available resources with their quantity and form are changing all the way, resources capability, current load, interests and tasks requests are dynamic too. It is also two types :

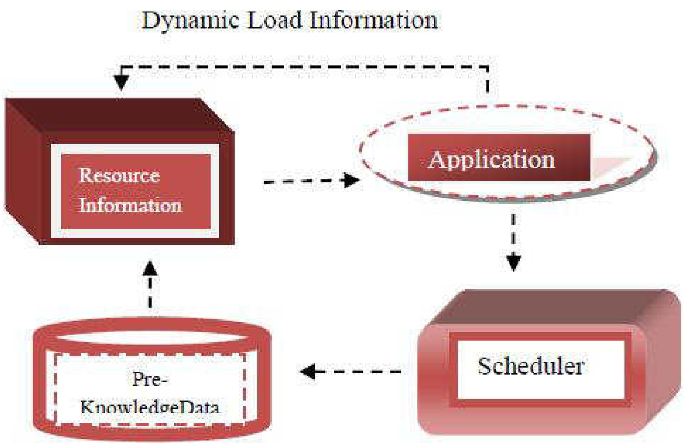


Fig 4. Dynamic Load Balancing

a) Centralized approach: - In centralized approach, only a single node is responsible for managing and distribution within the whole system. Other all nodes are not responsible for this.

b) Distributed approach: - In distributed approach, each node independently builds its own load vector. Vector collecting the load information of other nodes. All decisions are made locally using local load vectors. Distributed approach is more suitable for widely distributed systems such as cloud computing

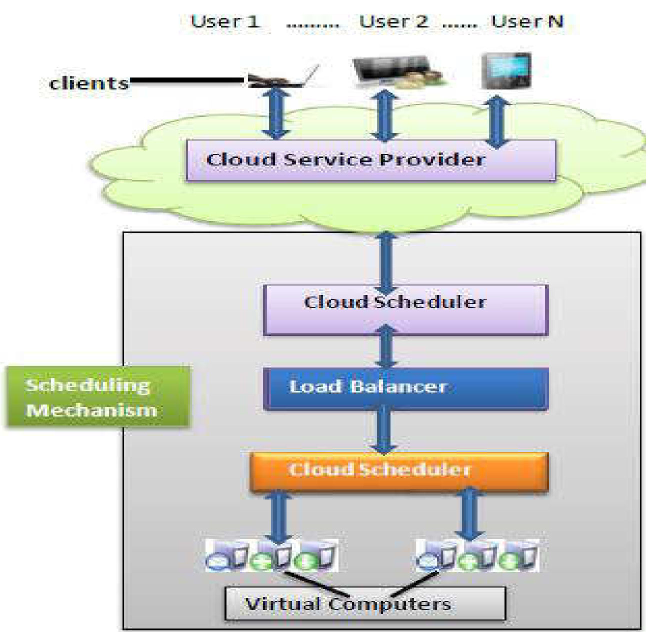


Fig 5. Client Subscriber Mechanism in Cloud

A. Round Robin Scheduling Algorithm (RR):

RR is a Static Scheduling algorithm which mainly focuses on eventually distribution of connection evenly across the virtual machines. The RR algorithm defines a circular queue and also defines a fixed time quantum [11]. During this time quantum, jobs are being allocated, if the job cannot be completed in one quantum, it will return to the queue and wait for the next round. The major advantage of this algorithm is that jobs are executed in turn and do not need to wait for the previous job completion [8]. Therefore, it does not suffer from a starvation problem. However its major drawback is, it does not consider whether the machine is heavily loaded or lightly loaded, server availability.

B. Weighted Round Robin Algorithm (WRR): It solves the problem of inconsistent server performance by adding weight, but when requesting service for a long time, it may cause load[10].

C. Greedy Algorithm: The Greedy algorithm is very simple and straight forward [11]. As a matter of fact, it was the only scheduling policy which was in use for a long time.

Only after the cloud started evolving, more complex scheduling policies came into effect. Greedy Algorithm work on the "assign jobs to a resource greedily" after the first resource is consumed it, and then it make request to the second resource.

D. RSDC (Reliable Scheduling Distributed in Cloud Computing): In this scheduling algorithm, tasks are divided into sub tasks and assigned to Vm machines. In order to balance the jobs, the request and acknowledge time is calculated and compared, and result will show which machines is utilized effectively [12]. This type of algorithm guaranteed high reliability and good QOS (quality of service) to the end user.

E .Resource-Aware-Scheduling algorithm (RASA): RASA is the combination of the both Max-min and Min-max Scheduling algorithm. This algorithm covers disadvantage of Max-min and Min-max and minimize the execution time of each task which gives minimized completion time [13]. It is both combination of resource aware scheduling algorithm and active monitoring of resources.

F. Prioritization based SJF (Shortest Job First) Scheduling:

In this scheduling algorithm the tasks are integrated as grouping, priority-aware of the resources which are used to reduce the waiting time and make span as well as to maximize resource utilization [14]. It dynamically arranges the resource according to their bandwidth and the resource capability to handle the tasks. After sorting of resources according to their priority it assigns job to the resource that will result in the proper utilization and deliver the job on time.

G. Load Balancing using Exponential Smoothing Forecast (ESF) algorithm: This algorithm is based on the prediction of algorithm on time series. This algorithm takes advantage of all historical data and distinguishes them through the smoothing factor [15]. Using the previous value of smoothing factor it predicts the future performance of server according to the load on it. This will help to choose right server by cloud server provider to balance the load.

TABLE I. COMPARISON OF EXISTING LOAD BALANCING ALGORITHMS

Parameter	**Round Robin**	**Greedy**	**ESF**	**RSBC**	**RASA**	**SJFP**
Dynamic/ Static	Static	Static	Dynamic	Static	Dynamic	Dynamic
Fault Tolerance	No	No	Yes	Yes	No	Yes
Resource Utilization	Less	Least	More	More	Medium	Medium
Network Overhead	Yes	Yes	No	Yes	No	No
Reliability	Less	Less	Medium	More	Medium	Medium

The important parameters that are being used to measure the service performance are:

1. Throughput: - It is used to calculate the all tasks whose execution has been completed. The performance of any system is improved if throughput is high.
2. Fault Tolerance: -It means recovery from failure. The load balancing should be a good fault tolerant technique.
3. Scalability: - It is the ability of an algorithm to perform Load balancing for any finite number of nodes of a system. This metric should be improved for a good system.
4. Migration time: -It is the time to migrate the jobs or resources from one node to other nodes. It should be minimized in order to enhance the performance of the system
5. Network Overhead: In real world Overhead for each job depends on the current network load and speed.

III. CONCLUSION

Cloud Computing has been widely adopted over the years by the Industries, though there are many existing issues like Load Balancing, Virtual Machine Migration, Server Consolidation, Energy Management etc. which have not been fully addressed. Load Balancing is also one of main issue in Cloud. In this paper we discussed about the important aspects of Cloud Computing. The unique services that are being provided by it and its scope in future. Then we also discussed about the importance of Load Balancing and types of Load Balancing which are being used in Cloud. Scheduling is an important issue in Cloud Computing as well. All its cloud services and efficiency would get effect by type of Load Balancing algorithm used in service providing. With proper Load Balancing, resource consumption can be kept to a minimum which will further reduce energy consumption and improve utilization which is a critical need of Cloud Computing.

REFERENCES

[1] Armbrust, Michael and Fox, Armando Armbrust, Michae and Fox, Armando Griffith, Rean and Joseph, Anthony D Katz, Ran “A view of cloud computing” in Communications journal of the ACM, vol. 53,No. 4,Year 2010.

[2] Grossman, Robert L “The case for cloud computing” In journal “ IT professional” in 2009 IEEE vol. 11, No. 2, pp-23-27.

[3] Dillon, Tharam and Wu, Chen and Chang, Elizabeth“Cloud computing: issues and challenges” IEEE year 2010.

[4] Pallis, George “Cloud Computing: The New Frontier of Internet Computing.” In journal “ IT professional” in IEEE 2010 vol. 14, No. 5.

[5] Hofmann, Paul and Woods, Dan “Cloud computing: the limits of public clouds for business applications Internet Computing.” In journal Internet Computing, IEEE 2012,vol 14,No.6.

[6] Xu, Xun “From cloud computing to cloud manufacturing”In journal “Robotics and computer-integrated Manufacturing” Elsevier 2012.

[7] Cloud Computing-Wikipedia, the free encyclopedia.htm

[8] Qi Zhang, Quanyan Zhu, Boutaba, R., “Dynamic Resource Allocation for Spot Markets in CloudComputing Environments”, IEEE, 2012.

[9] Yang, C.-T., Cheng, H.-Y., and Huang, K.“A Dynamic Resource Allocation Model for Virtual Machine Management on Cloud” in Grid an Distributed Computing, pp 581-590, Springer,2011.

[10] P. T. Endo, A. V. A. Palhares, N. N. Pereira, G. E. Gonçalves, D. Sadok, J. Kelner, B. Melander, and J. E Mangs.“Resource allocation for distributed cloud: concepts and research challenges”, IEEE Network Magazine, vol. 25, pp. 42-46, July 2011.

[11] Ma, L., Lu, Y., Zhang, F., and Sun, “Dynamic Task Scheduling in Cloud Computing Based on Greed Strategy” in Trustworthy Computing and Services pp 156-162, Springer 2013.

[12] Arash Ghorbannia Delavar , Mahdi Javanmard : Mehrdad Barzegar Shabestri and Marjan Khosravi Tal “RSDC: Reliable Scheduling Distributed in Cloud Computing” in International Journal of Computer Science Engineering and Applications(IJCESA) Vol 2, No.3, June 2012.

[13] Saeed Parsa and Reza Entezari-Maleki,” RASA: A New Task Scheduling Algorithm in Grid Environment” in World Applied Sciences Journal 7 (Special Issue of Computer & IT): 152-160, 2009.

[14] Ru, J. and KeungAn Empirical Investigation on the Simulation of Priority and Shortest-Job- First Scheduling for Cloud-Based Software Systems” in Software Engineering Conference (ASWEC), 2013 22nd Australian, pp 78-87, IEEE 2013.

[15] Ren, Xiaona and Lin, Rongheng and Zou, “A dynamic load balancing strategy for cloud computing “Platform based on exponential smoothing forecast “In IEEE 2011.

[16] Armbrust, M., Fox, A., Griffith, R., and Joseph, A. D. (2009), “Above the Clouds¬: A Berkeley View of Cloud Computing,” Science, 53 (UCB/EECS-2009 28), 07-013, Citeseer.

[17] D. Warneke and O. Kao, ”Exploiting Dynamic Resource Allocation or Efficient Parallel Data Processing in the Cloud,” IEEE Trans. on Parallel and Distributed Systems, Vol.22, No.6, pp.985-997, 2011.

[18] Chen Kang and Zheng Weimin, “Cloud computing: system instance and Research,” Journal of Software,vol. 20, 2009, pp. 1337–1348.

[19] X. Meng, V. Pappas, and L. Zhang, “Improving the scalability of data center networks with traffic-aware virtual machine placement,” in Proc. IEEE Infocom., 2010, pp. 1–9.

[20] Ardagna D, Casolari S, Panicucci B. Flexible Distributed Capacity Allocation and Load Redirect Algorithms for Cloud Systems. In: Proceedings of the 4th International Conference on Cloud Computing;2011 Jul 4-7; Washington, USA: IEEE Computer Society; 2011. p. 163- 170.

A Review on Software Maintenance Cost Evaluation

Reenu Rani
Dept. of CSE, Guru Nanak Dev University
Amritsar (Pb.) India
reenu.reenu9@gmail.com

Amandeep Kaur Cheema
Dept. of CSE, Guru Nanak Dev university
Amritsar (Pb.) India
aman.cheema2k12@gmail.com

***Abstract*-Software reuse and software maintenance are critical issue in the software reengineering. Software maintenance cost defines the time when to redevelop a software product also software define select those components among already available software for reuse. This paper evaluates the review of the software maintenance cost or software reuse. Various static and dynamic metrics are used to find the whether or not to reuse the software component and also various metrics define the software maintenance cost.**

Keywords: Software re-engineering, Software maintenance, Software reuse.

I. INTRODUCTION

Software Engineering is a coated technology. Software Engineering implements strong engineering principles to provide economical software that will reliable and provides efficient performance on real machines. It is the study of engineering to the plan, improvement, and maintenance of software.

Software Development is the improvement of software program. It includes activities involved more than one person which depends totally on the size of the software. Each phase of the development is comprised of various activities of the individual.

Software reengineering is modified or maintains the software with reverse engineering. There are different types of software's are implemented in software engineering. Software is set of lines of code that on execution provides the desired productivity. Data Structures helps the code to store the information and store them, manipulate them and finally convert them to the nearest set of output. Various programming languages are used during the software development. Due to the recent evolution of the computer's in all the fields software engineering provides a lot of features in the form of products, software's, systems and information systems. Software development have different phases in which software maintenance is last and important phase. Software maintenance is correction of errors in software and adds some functionality to make it adaptable after delivery of that software. Software maintenance is process that is used to adapt, organize, improve, prevent and maintain the performance of any software. In software development it is very important to think about software maintenance.

A. Purpose of software maintenance

In today's world, customer's requirements are changing day by day. Customers want that software which provides an accurate work. To make software adaptable with changing environment there is need of maintenance so software should be developed in this way so it can be maintain easily. Software maintenance makes a software correct, perfect, preventive and adaptable with changing environment

B. Software Maintenance

This phase focuses on corrective the bugs and errors in the software after testing the software. This is the important phase of software development as the bugs are meant to be fixed within reasonable time with less cost. This includes enhancements also that come from the changing requirements from the customer.

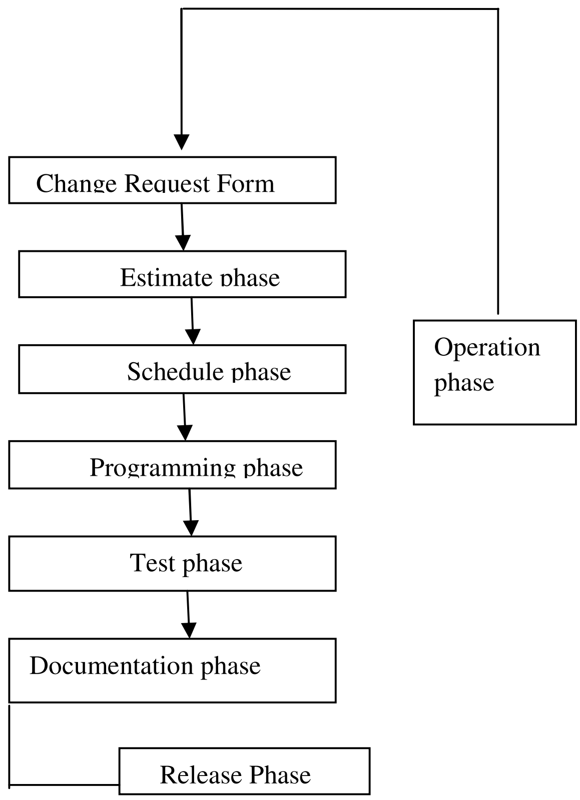


Fig.1. Software Maintenance

C. Need of maintenance

Our environment is dynamic in nature. User requirements, economical and technical changes are taking place in our environment so there is need of modification to keep it adaptable with environment. To meets today's requirement maintenance is very important part. It is last phase of software development life cycle. All the software's will need maintenance over there lifetime.

- ❖ If there is serious error in software then there is need of software maintenance to keep it compatible.
- ❖ New business has different new processes for cooperation there is need of maintenance in existing system.
- ❖ For security vulnerability of software there is need of software maintenance.
- ❖ When hardware is changed then there is need of modification of software to make it compatible with that hardware so there is also need of maintenance

D. Maintenance cost and cost factors

Maintenance cost is greater than development cost. Software maintenance damages the product formation. It makes advance maintenance more complex. The software's which are developed with old languages and compilers needs more time and cost for maintenance.

Cost factors:

1. *Team stability:* If the same staff who has developed software are involves in software maintenance then maintenance cost can be reduced
2. *Contractual responsibility*: The developers of software may have no contractual responsibility for software maintenance so there is no motivation to plan for upcoming modify.
3. *Staff skills*: Maintenance personnel arefrequentlynew and have restrictedareafamiliarity.
4. *Program age and structure*:As programs age, their formation is corrupted and theybe converted into harder to realize and modify.

E. Maintenance Cost evaluation

Changing software after post delivery is called software maintenance. Software maintenance cost is the cost required to maintain the software after initial deployment. Customer's needs are changing day by day. Software maintainability is important part for success because product does not wear out but theuse of that product will get less. Cost will be high if the software is not having reusable components as it is the important key in changing the system to new changes. Software maintenance cost include

- ❖ Corrective maintenance: Software with the best quality also encounters defects but the corrective maintenance helps to fix those defects. Cost required for bug fixing after initial deployment is known as corrective maintenance cost .It is generally 20% of all software maintenance cost.
- ❖ Adaptive maintenance: Due to the changes in the environment including CPU, Operating System, Business parameters, software meant to be changed and adopt the new changes in environment. Cost required to make software adaptable in all environments is known as adaptive maintenance cost. It is generally 25% of all software maintenance cost.
- ❖ Perfective maintenance: To improve and enhance the overall performance of the software after initial deployment is perfective maintenance .It is generally 5% of total software maintenance cost.
- ❖ Enhancements: Costs due to long-term new creations and innovations in software. It is generally 50% or more of total maintenance cost of that software.

The cost evaluation is one of the economical techniques which is used to choose a well-organized program from wide range of alternatives and to propose and execute that program. The principle of cost evaluation is do more with less budget. Cost evaluation identify most efficient approach of software. Maintaining a software with maximum profit and minimum cost is known as optimal maintenance. Cost functions are depend on the reliability and maintainability individuality of the software find out the parameters of significance to decrease. Parameters measured are – the cost of collapse, – the cost per time unit of "downtime", – the cost (per time unit) of corrective maintenance, – the cost per time unit of preventive maintenance an – the cost of repairable system replacement. Maintenance cost increases above time to time.A software which has been delivered isexclusive to change.Maintenance costs increase over time and as the system evolves Reasons: Maintenance changes, degrades the original system structure. Aging software results in high support costs.

F. Software Reuse

It is common to reuse the hardware in the engineering process, but software reuse is one of the key quality components of the software. It has been innovated to make engineer comes up with the better and efficient design. A component or modules should be developed such that the other software's or program can use these components and can save the huge development hours on these common modules.

This helps to complete the software within the dead-lines to develop software fast, less costly and reliable.

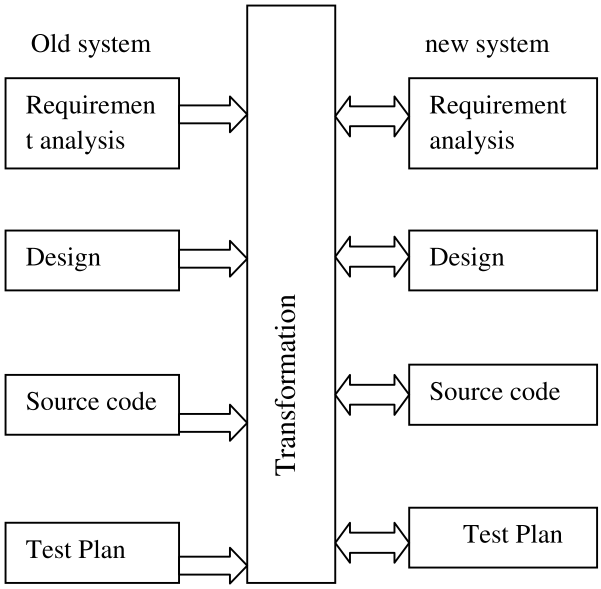


Fig.2.Old system vs New System

II. RELATED WORK

Imai et al. [1]said that, code cloning is frequently evaluated by most of the software companies. The source code which is used by more than one project can increase the maintenance cost. When software functionalities are modified and some changes are required due to changing environment then source code cloning difficulty for that software project. This paper describes the maintenance cost evaluation which evaluates the impact of reuse to maintenance. The main motive of this paper is to measure the functional redundancy. It can be measured in two steps: first to make clusters of similar functions which are used in that source code and second to make an array weighted FR-tree according to that clusters. After that, functional redundancy can be measured by weight of each node in FR-tree.

Srivastava and Biplav [2] describe that, Software Engineering is all about combining the components with each other and then assembled it into Software. These components are further comes up with combination of sub-modules. One of the main tasks in managing software project involves having a track of development and managing the individual components. Though tools are there to track the component development and maintenance but still the main and important hurdle it to evaluate trade-offs manually. Software engineering is one of the fast growing fields and new trends are there like Web Services, EJB's, so there is a need of an automated solution to help developers in creating complex applications. This paper proposed an automated decision support framework that will guide the users to make cost-effective decisions. The key approach of this paper is to develop a model of the software using the automated planning techniques to create alternative plans to develop and maintain software's taking into considerations of the developer's efforts and objectives.

Nagappan and Ball [3]have proposed that Software development is complex task in which there is need of complete knowledge about the architecture of that particular software. This paper describes, how the software dependencies and churn measures are prone to post release failures. In this paper, author analyzes the window server 2003 operating server and examines the relationship between software dependencies and churns measures and also investigates their ability to proneness post release failures. Churn measures are the measurement of code which is changing with in software unit with the analysis this paper conclude that software dependencies and churn measures arethe efficient predictors for post release failures and failure proneness. It also concludes that software quality can also be measured with this prediction and post release failure proneness is also economically useful.

Wolf et al. [4] described that Communication is the way through which ideas, thoughts and knowledge can be share in work groups. Communication and coordination plays an important role for software quality. This paper describe the relationship between successful coordination outcome and communication structure using the data from the project of IBM (Jazz TM project)**.** This paper introduces two research question and methodology in the study of integration and communication structure in Jazz project. Author conceptualizes the coordination outcome with the success or failure of the project and study communication structure of team with social network measures. The paper concludes that coordination outcome and communication structure are the efficient predictors for failure in future. Problems in communication will lead to coordination and integration failures.

Cataldo et al. [5]describe past research has shown that customer reported software faults are the result of violated dependencies that are not recognized in software developing by software developers. There are three types of dependencies which are included in this paper. The main objective of this paper is to find the performance of the dependencies and relate it to customer reported defects. Analysis of this research is based on two different projects of different companies. Research related to this topic shows that logical, syntactic and work dependencies all prone to failure. Prior research has shown that syntactic dependencies were more prone to failure on the other hand this research shows that logical dependencies and work dependencies are more prone to failure. The result of this paper suggests that architecting guided by the network structure of logical dependencies are useful for reducing defects.

Mader, Patrick, and Alexander Egyed[6],have proposed that Effects of requirements traceability regularly helps the developer's to have a track of the software changes and makes the software maintenance support better. Though it is quite popular approach, but still there is no publication about the benefits of the requirements traceability. It is quite important to learn that whether introducing requirements traceability can make the development tasks better. This paper conducted an experiment using 52 developer's performing real time maintenance on the third party software, half of the tasks with traceability and half of them without traceability. The results shows that the developers with traceability performs 21% faster on a task and provides 60% more correct solution suggesting that traceability not only saves money but provides better maintenance. Using the initial costs setup for our evaluated traceability featured systems we will implement the same to improve the maintenance support quality.

Langelier et al. [7],describe that evaluating the software quality and understanding the events leading its evolution to anomalies are two key steps in reducing the cost in software maintenance. Evaluation of the large quantity code over various versions is the most time consuming and practiced in general. To answer this, this paper described that a semi-automatic visualization framework to investigate programs having thousands of classes, over dozens of versions at much faster speed. Programs and their associated quality parameters for each version are represented graphically independently. The animations between these representations create visual quality of consistence associated with the quality belonging to the various software versions. Exploring these qualities parameters can reduces the difference between various views of software. This allows experts to use their analysis skills to investigate all quality aspects of software evolution.

Ghavami et al. [8], have proposed that a probalistic model to attain cost effective maintenance strategies. Customer's needs changes day by day so the maintainability is very important

Part for all software's. Maintenance increase the life time of any product and increase the meantime among failures, how easily software can be modified during bug fixes or other modifications is generally known as maintenance. Now days, maintenance cost effectiveness is important because of budget constraints of each software industry. Cost effective and profitable product will be more reliable. This paper describes that based on reliability indices mean duration and state probability are computed using Monte Carlo simulations and numerical examples; cost analysis is also performed with computation of all related cost like maintenance, failure cost based on reliability.

Liu et al. [9], have proposed novel maintenance strategy model, which focuses on the task of maintenance workers and the work provision method. This model can be used to estimate product reliability in different maintenance circumstances. One of the maintainability parameter is frequency of emergency state, which plays the crucial role in enforcing hierarchical planned maintenance. With reduction of hardware and software cost, total cost can be reduced. In this paper, the viability of this method can be revealed by a plain example. This assessment and model have been implemented as a software tool for dependability proposes.

Mitsuhiro et al. [10], have proposed that, product feature evaluation method in development process with the change of software metrics. It is broadly known that in software development companies, software evolution makes the product configuration difficult and it makes the product evolution more complex. This paper used huge size telecommunication software fixed in the radio network system which is described by C/C++ language. As the outcome, this Paper prove that betweenness, centralization and average distance of complex network metrics have great impacts on fault densities in software evolutions process.

III. CONCLUSION AND FUTURE WORK

This paper has evaluated various techniques for software re-engineering. Software quality metrics are used to find the software maintenance cost and also for whether the software is reusable or not. It has been found that the software maintenance plays a significant role to build the software product again called new version of that software. In near future we will use some reusable components and evaluate their reusable cost to find the best one; based upon certain quality metrics.

REFFRENCES

[1] Imai, Takeo, Yoshio Kataoka, and TetsujiFukaya. "Evaluating software maintenance cost using functional redundancy metrics." Computer Software and Applications Conference, 2002. COMPSAC 2002. Proceedings. 26th Annual International. IEEE, 2002.

[2] Srivastava, Biplav. "A decision-support framework for component reuse and maintenance in software project management." Software Maintenance and Reengineering, 2004. CSMR 2004. Proceedings. Eighth European Conference on. IEEE, 2004.

[3] NachiappanNagappan, Thomas Ball, "Using Software Dependencies and Churn Metrics to Predict Field Failures: An Empirical Case Study", First International Symposium on Empirical Software Engineering and Measurement, 2007 IEEE.

[4] Wolf, Timo, Adrian Schroter, Daniela Damian, and Thanh Nguyen. "Predicting build failures using social network analysis on developer communication." In Proceedings of the 31st International Conference on Software Engineering, pp. 1-11. IEEE Computer Society, 2009.

[5] Cataldo, Marcelo, AudrisMockus, Jeffrey A. Roberts, and James D. Herbsleb. "Software dependencies, work dependencies, and their impact on failures." Software Engineering, IEEE Transactions on 35, no. 6 (2009): 864-878.

[6] Mader, Patrick, and Alexander Egyed. "Assessing the effect of requirements traceability for software maintenance." In Software Maintenance (ICSM), 2012 28th IEEE International Conference on, pp. 171-180. IEEE, 2012.

[7] Langelier, Guillaume, HouariSahraoui, and Pierre Poulin. "Exploring the evolution of software quality with animated visualization." Visual Languages and Human-Centric Computing, 2008. VL/HCC 2008. IEEE Symposium on. IEEE, 2008..

[8] Ghavami, Mohsen, and MladenKezunovic. "Probabilistic evaluation of the effect of maintenance parameters on reliability and cost." Probabilistic Methods Applied to Power Systems (PMAPS), 2010 IEEE 11th International Conference on. IEEE, 2010.

[9] Liu, Bin, and Zhengguo Xu. "A new cost evaluation model of predictive maintenance for dynamic systems with hidden degradation." Chinese Automation Congress (CAC), 2013. IEEE, 2013.

[10] Mitsuhiro Nakamura andTomoki Hamagami," A Software quality evaluation method using thechange of source code metrics"2012 IEEE 23rd International Symposium on Software Reliability Engineering Workshops.

Cleanroom Software Engineering – A Review

Manraj
Dept. of Computer Science & Engineering
Guru Nanak Dev University RC
Gurdaspur, India
er.manrajgill@yahoo.in

***Abstract*— Clean room is process for the production of high-quality software. It is a team oriented process. It motivates the production of zero-defect software. The goal of cleanroom software engineering is to use a structured development process to produce a software that is "right the first time" instead of using a software testing process to reach a desired level of reliability. The high quality software produced with cleanroom approach improves manageability, satisfies customers, reduces risks, reduces costs and provides a competitive advantage. This paper explains cleanroom approach. Following the introduction, section two gives an overview of cleanroom approach. Technology of cleanroom approach is described in section three. Section four describes the cleanroom case study in software engineering laboratory of NASA. Some important cleanroom projects are described in section five.**

Keywords—Cleanroom approach,incremental development,box structure,correctness verification, statistical testing, certification.

I. INTRODUCTION

Cleanroom theory was established in late 1970s and early 1980s by Harlan Mills, a mathematician and IBM employee. He related the fundamental facts in statistics, mathematics and engineering to software. The foundations of cleanroom approach are drawn from the function theory of mathematics and statistical science. The function defines mapping of a domain set and a range set. Each element in domain is mapped to exactly one element of the range. The statistical science, deals with the collection and organization of data, which is followed by analysis, interpretation and presentation of data. It supports the product and process measurements and also employs the iterative process of incremental development to permit the consistency of performance to be measured and improved.

II. OVERVIEW

Cleanroom software engineering is the practical application of mathematical and statistical science to develop high-quality software for economic production. The name cleanroom was taken from the hardware manufacturing cleanrooms of the semiconductor industry, where defect prevention is preferred over fixing them later. This approach is designed to attain two goals: 1) a manageable development process and 2) zero failures in use.

A. *Manageable development:*

Managers and technical teams maintain an intellectual control of software development projects. At each step of development, the status or level of work in progress is clear. It results in management of complexity, avoiding rework, reduction of risks, and meeting business objectives.

B. Zero failure: In cleanroom, hard and careful specification, design, and verification practices are coupled with testing practices to create a zero defect software.

III. TECHNOLOGY

Cleanroom software engineering is an integration of three important principal technologies: 1) Incremental development under statistical process control; 2) Function-based specification, design, and verification; and 3) Statistical testing and software certification. These technologies can be used individually or jointly.

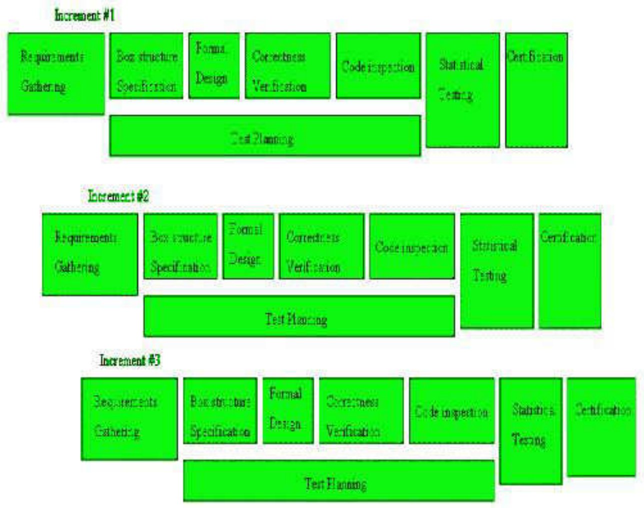


Fig.1.Cleanroom engineering process [6]

A. *Incremental development under statistical process control:*

Cleanroom approach uses a special version of the incremental process applying pipelining technique where each increment is eventually added together to create the whole system. A seven step process is followed for each increment assuring that each component is certified as reliable before moving on to the next component. It provides the benefits like intellectual control, risk management and customer feedback and control of budget. User feedback on each increment is a gauge to measure whether the right system is being built. The goal of incremental development is to grow a system with each new increment as an elaboration of the functions implemented in prior increments. An incremental strategy is applied to the project and each increment is defined and planned from start to end. Detailed description of customer's requirements is described for each functional increment. Gathering of requirements may lead to a simplified customer concept by repeatedly determining the requirements that the customer may not have initially addressed.

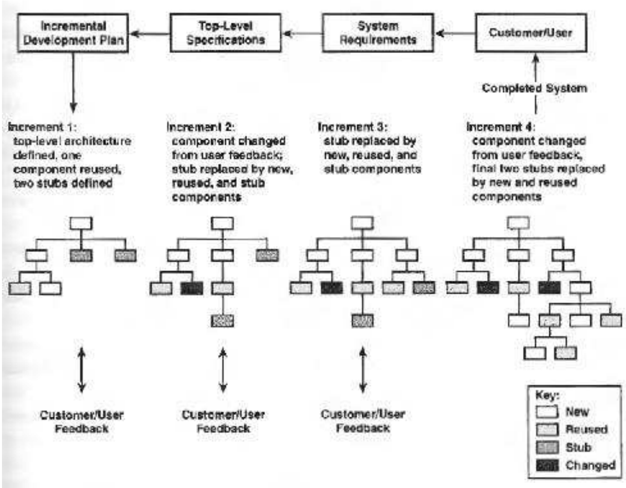


Fig.2.Incremental development and feedback [1]

B. *Function-based specification, design, and verifications:*

i. *Specification:*

Three system structures: 1) black box, 2) state box, and 3) clear box are used for the specification and design purpose. These three structures are collectively known as box structures. These structures incorporate the concepts of data encapsulation and information hiding. Box structures are developed in a step-by-step refinement and verification process that collaborate both system control and data operations. Every individual refinement is verified against the previous step. Box structures therefore separate three aspects of system development (specification of behavior, data, and procedure).

1. *Black box:*

The black box specification defines the required external behavior of the system. A system accepts a stimulus(S) as input from its environment and produces a response(R) as output to the environment.

Following function is the mathematical semantics of black box behavior:

$$SH \rightarrow R$$

(*SH* is the complete stimulus history, involving the current stimulus)

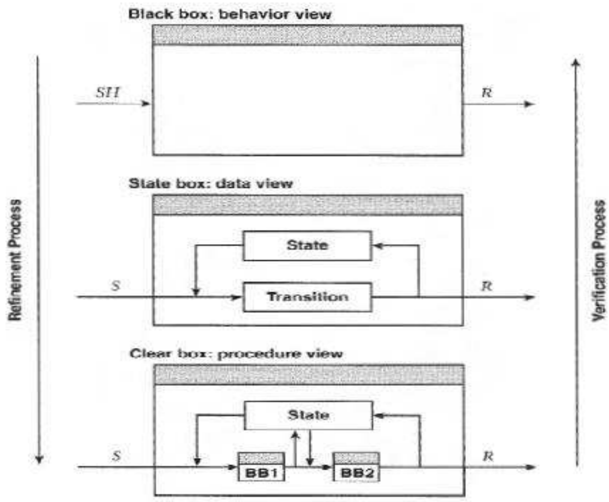


Fig.3.Box structure refinement and verification. BB =black box [1]

The response not only depends upon the current stimulus, but also on the history of stimuli. Denotation of the black box is state and procedure free. It describes externally visible behavior experienced by users only in terms of history of use. A black box concentrates on the user view in addressing questions of system behavior and does not need decisions on state and procedure design. A black box specification describes the required behavior in all possible circumstances of use. It means that the accurate responses for all possible combinations of current stimuli and stimulus histories are defined in a black box specification. In cleanroom approach, black box specifications have three major uses, all of which are important for effective system development:

i. *For system owners and users:* black boxes define required behavior for analysis and agreement before committing resources to development and testing.

ii. *For system developers:* black boxes define the required behavior to be designed and implemented.

iii. *For system testers:* black boxes define the required behavior to be validated during testing.

2. *State Box:*

The state box specification provides an intermediate view that defines the state space of a system. State boxes encapsulate stimulus history as state data. A state box definition is procedure free. A state box maps an old state (OS) and a stimulus (S) to a new state (NS) and a response (R). Hence for the next transition, the new state becomes the old state.

Following transition function is the semantics of state box behavior:

$$(OS, S) \rightarrow (NS, R)$$

A state box is refined from black box and is also verified against the corresponding black box. The state represents information from the black box stimulus history and retains this information. Therefore, a state box does not require stimulus history in its definition. In other words, the state box can be seen as a generalized state machine including state data and operations and also incorporating the black box. State boxes show the behavior of a system, its reaction and transitions due to user inputs.

3. Clear Box:

A clear box is a program that implements the state box and introduces and connects operations in a program structure for decomposition at the next level. It is derived from a state box and is also verified against the corresponding state box. A clear box is a program that accepts a stimulus(S) and, according to the program's internal state (OS) creates a new internal state (NS) and a response(R).

Following transition function is the semantics of clear box behavior:

$$(OS, S) \rightarrow (NS, R) \text{ (By procedure)}$$

The clear box breaks down the sub functions of the black box into procedural like descriptions that looks similar to structured programming. Clear box procedures may reuse existing black boxes, and can also introduce new black boxes for refinement into state and clear box forms.

1) Design:

Once the structure specification of box is complete, then cleanroom designing takes place. Cleanroom uses the structured programming approach to design its functions. Functions in clear box specification are refined from mathematical functions to logical connectives that resemble a programming language systematically. Developers use the concepts of data encapsulation, information hiding and data typing to design the product's data. At this stage much of the design is able to be implemented in the selected programming language of choice.

a) Correctness Verification:

The correctness of the software increment is verified by using the software reengineering process which use mathematically based techniques. Correctness is reviewed by the development team using function-theoretic reasoning. Specifications of black-box are verified to be correct, complete and consistent. State box specifications are then verified with respect to the corresponding black box specifications. The specification and development teams correct the faults found in verification reviews which are documented in the Increment Verification Report. Then the faulty specifications and designs are re-reviewed and re-verified. This process is the last line of defense against failures and shortcomings. The objective of Correctness Verification is to move to the testing phase with no faults in the design. Now, the increment is turned over to the certification team for the first execution of the code.

2) Code Implementation & Inspection

After verification of design, the structured programming design method can be translated into the correct programming language. Then this code is also put through correctness verification and it results in every line of code in the software to be verified as functionally correct. After code verification teams discuss any possible complications and changes, all members agree on the design and implementation, and then statistical testing is applied.

a. Statistical Testing & Certification

The Certification process is the third and last process. It is comprised of two process; 1) Usage Modeling and Test Planning, 2) Statistical Testing and Certification process.

1) Usage Modeling and Test Planning:

Usage models are created for software testing and certification. These models are expressed in terms of software usage states and probabilities of transition between them. Models are developed for the certification of expected operational use or infrequently used functions with high consequences of failure. In this phase statistical test plan is developed, test environment is prepared, and the statistical test cases are generated.

2) Statistical Testing and Certification

Statistical test cases are executed under experimental control and the results are evaluated. If failures are encountered then initiate engineering change activity. The values of certification measures obtained in statistical testing are compared with certification goals to assess the software's fitness for use. These decisions decide whether to continue testing or to stop testing for changes to the software. Evaluations and decisions are about product quality and process control is documented in the Increment Certification Report.

IV. NASA CASE STUDY

In March 1990, the Software Engineering Laboratory (SEL) at the National Aeronautics and Space Administration/Goddard Space Flight Center (NASA/GSFC) investigates the effectiveness of cleanroom software development methodology over standard SEL methodology. Following are some differences between SEL cleanroom model and the standard SEL development method:

1. Cleanroom testers and developers are on completely separate teams where as in standard SEL there is a single development and test team.
2. Clean room developers have no access to the mainframe computer for compilation and testing purposes.
3. Cleanroom developers rely on code reading instead of unit testing to verify correctness of the software prior to system testing.
4. Clean room testers use a statistical testing approach where as standard SEL use integration and system testing.

V. RESULT

There were indications of an increase in product quality and developer productivity. Cleanroom produced a lower failure rate of 3.3 errors/KSLOC while standard SEL produced 6 errors/KSLOC. There are indications of a more complete and consistent set of inline code documentation. In terms of effort, there is a distinct distribution of effort activity. More time is used in design and meetings phase, and less time is consumed in coding.

VI. CLEANROOM PROJECTS

Traditionally developed software does not experience correctness verification. It moves from the development to unit testing and then debugging. Afterwards, it passes through more debugging in function and system testing. At entrance to unit testing, traditional software generally shows 30-50 errors/KLOC. Traditional projects often show errors beginning with function testing or omitting errors found in private unit testing. There is a difference in the complexity of errors found in cleanroom and traditional code. Errors left behind by cleanroom‘s correctness verification are simple mistakes that can be easily found and fixed by statistical testing, there are no deep design or interface errors. Usually cleanroom errors are infrequent and simple as well. Following are some cleanroom projects:

A. IBM Flight Control: In year 1987, a HH-60 helicopter avionics component was developed by using cleanroom software engineering technology in three increments comprising 33 KLOC of JOVIAL. 79 corrections were required during statistical certification for an error rate of 2.3 errors per KLOC for verified software with no prior execution or debugging. [4]

B. IBM COBOL Structuring Facility (COBOL/SF): COBOL/SF was IBM’s first commercial cleanroom product. It was developed by a six-person team in 1988. The product automatically converts unstructured COBOL programs into functionally equivalent structured form. It improved understandability and maintenance. Certification testing failure rate was 3.4 errors/KLOC. Productivity was 740 LOC/PM. Deployment failures 0.2 error/KLOC. All errors were simple.

C. Martin Marietta Automated Documentation System: In 1990, a team of four persons developed a prototype of the Automated Production Control Documentation System. It was a relational database application of 1820 lines which were programmed in FOXBASE. In first compilation no errors were found. Certification testing failure rate was 0.0errors/KLOC.

D. IBM System Software: In 1991, the first increment of a system software product in C was developed by a cleanroom team of four persons. No error was found in first compilation. Certification testing failure rate was 0.0 errors/KLOC.

E. IBM System Product: In 1991, an organization of 50 people developed a complex system software product. It was developed in three increments totaling 107 KLOC. Testing failure rate was 1.6 errors/KLOC and the project reported development team productivity of 486 LOC/PM.

F. IBM Language Product: In 1991, a cleanroom team of seven persons developed an extension to a language product. The first increment of 21.9 KLOC was up and testing failure rate was 2.1 errors/KLOC.

G. IBM Image Product Component: To compress and decompress data from a Joint Photographic Expert Group (JPEG) data stream, a 3.5 KLOC image product component was developed in 1991. In first compilation there were 5 syntax errors. Certification testing failure rate was 0.9 errors/KLOC.

H. IBM Printer Application: In 1992, the first increment of a graphics layout editor in C under OS/2 Presentation Manager was developed by a cleanroom team of eleven members. The first increment of 6.7 KLOC exhibits certification testing failure rate of 5.1 errors/KLOC.

I. IBM Knowledge Based System Application: In 1992, a prototype knowledge based system for the FAA Air Traffic Control System was developed by a team of five persons. Testing failure rate was 3.5 errors/KLOC.

J. IBM Device Controller: In 1993, a team of five persons developed two increments of device controller design and microcode in 40 KLOC of C, with 30.5 KLOC of function definitions. The Certification testing failure rate was 1.8 errors/KLOC.

K. IBM Database Transaction Processor: In 1993**,** the first increment of a host based database transaction processor in 8.5 KLOC of JOVIAL was developed by a team of five persons. Testing failure rate was 1.8 errors/KLOC with no design errors and all simple fixes.

L. IBM LAN Software: In 1993, the first increment of a LAN based object server in 4.8 KLOC of C was developed by a team of four persons. Testing failure rate was 0.8 errors/KLOC.

VII. CONCLUSION

Cleanroom engineering develops a zero-defect product. Cleanroom approach provides the following benefits:

A. *Productivity improvement:* Team orientated incremental pipelining approach allows many components to work concurrently, hence increasing productivity.

B. *Quality improvement:* Continuous team meetings and reviews along with statistical testing and correctness verification develop a far better quality code than unit testing and debugging.

C. *Code size reduced:* Strict formal design and specification methods and refinement lead to reduction in code size.

D. *Return on investment:* Cost-effectiveness is increased because due to the improvement in software performance and the time, money is saved due to significantly less testing revenue.

E. Reliable certified software: Finally, Cleanroom development approach produces a reliable software and presents the user with a near zero defect product.

REFERENCES

[1] Stacy J.Prowell,Carmen J.Trammell, Richard C.Linger,Jesse H. Poore,"Cleanroom software engineering technology and process", 1999.

[2] Scott Green(NASA/GSFC), Ara Kouchakdjian, Victor Basili,David Weidow(NASA/GSFC),"The cleanroom case study in the software engineering laboratory: Project description and early analysis", march 1990, National Aeronautics and Space Administration.

[3] Linger, Richard C., Trammell, Carmen J. (November 1996). "Cleanroom Software Engineering: Reference Model", Version 1.0. Retrieved March 27, 2006.

[4] Richard C. Linger,"Cleanroom Software Engineering for Zero- Defect Software", IBM Cleanroom Software Technology Center.

[5] Richard W. Selby, Victor R. Basili,F. Terry Baker,"Cleanroom Software Engineering: An Empirical Evaluation",IEEE transactions on software engineering,VOL. SE-13, NO. 9. September, 1987.

[6] http://d.web.umkc.edu/dcw844/hw3.htm.

Progress of Data Mining in Different Application Areas

Garima Sharma
Lovely Professional University,
Phagwara, India.
garima2306@yahoo.com

Sheveta Vashisht
Lovely Professional University,
Phagwara, India
sheveta.16856@lpu.co.in

***Abstract*— With the increasing demand of IT and very fast growth in every area creates huge amount of data. Data mining plays very important role in analysis and extracting useful information from data. Data mining is also known as knowledge discovery from data. Data mining is used for forecasting future trends of market and also helps in decision making like in business, science, medical, telecommunication etc. The objective of this paper is to discuss the need of data mining in various fields and also explore the future trends.**

***Keywords*— Data mining, Data mining applications, KDD, Present and future trends, review.**

I. INTRODUCTION

Data mining is the process of discovering interesting pattern and knowledge from large amount of data and also named as knowledge discovery from data (KDD). The data sources can databases, the web, the data warehousing etc. Data converts raw or huge data into interesting and useful information [1]. KDD refers to overall process of discovering useful knowledge from data, while data mining refers to applications of algorithms for extraction patterns from data [3]. Data mining is best described as the union of historical and recent development in Statistics, AI, machine learning and database technologies [2]. The various areas of data mining financial data analysis, retail industry, telecommunication industry, bio medical, mobile computing, network intrusion detection, target marketing, currency crisis, terrorism, education etc. The objective of this paper is to reviews various application of data mining and also its development from present to future. The paper is organized as follows section 2 presents the basic and advance techniques of data mining, section 3 presents present trends in data mining, section 4 presents future trends in data mining and finally the conclusion.

II. DATA MINING ALGORITHM AND TECHNIQUES

There are many data mining techniques and algorithms have been developed and used in data mining like association, classification, clustering, prediction and sequential patterns, Regression, Neural Networks etc. Some data mining techniques are discussed below:

A. Classification

Classification is the most commonly used data mining technique, where a model is constructing to predict class. It is used to classify each item in a set of data into one of predefined set of classes or groups. Classification method uses mathematical techniques such as decision trees, linear programming, neural network and statistics. Classification classifies data items into groups. Types of classification models:

- Classification by decision tree induction
- Bayesian Classification
- Neural Networks
- Support Vector Machines (SVM)
- Classification Based on Associations

B. Association

Association is one of the best known data mining technique. In association, a pattern is making on the basis of relationship of a particular item on other items in same group. Association and correlation are used to find repeated data item from large data. The various types of association rules are:

- Multilevel association rule.
- Multidimensional association rule
- Quantitative association rule
- Direct association rule
- Indirect association rule

C. Clustering

Clustering is a data mining technique that makes useful cluster of objects that have similar characteristic. Classification approach can also be used for distinguishing groups or classes of object but it becomes costly so clustering can be used as preprocessing approach for selection and classification.

Types of clustering methods are:

- Partitioning Methods
- Hierarchical Agglomerative (divisive) methods
- Density based methods
- Grid-based methods
- Model-based methods

D. *Prediction*

It is one of a data mining technique that constructs relationship between independent variables and dependent variables. Regression technique can be used for predication. Regression analysis can be used to model the relationship between one or more independent variables and dependent variables. Types of regression methods are:

- Linear Regression
- Multivariate Linear Regression
- Nonlinear Regression
- Multivariate Nonlinear Regression

III. PRESENT TRENDS OF DATA MINING

The field of data mining is growing very fast because data mining is handling various different challenges of different domains. A number of data mining applications have been successfully implemented in various fields like customer relationship management, financial data analysis, retail industry, counter-terrorism, telecommunication industry etc.

A. *Counter-terrorism*

After 9/11 attack on America, the various countries worried about their national security, then data mining techniques and algorithms can be used for tracking the activities of terrorists. USA launched Total Information Awareness Program with the goal of creating a huge database of that consolidates all the information on population. Similar projects were launched in European countries and the rest of the world [2].

- Data mining techniques plays important role in analyzing accurate identification from a huge database of population. Accurate identification helps in finding whether the identity of person is genuine or fake. While handling the large databases some problems has arises:

a. Scalability of algorithms.

b. The target database has to be deal with different types of data like audio, video, text and multimedia data.

Data mining algorithms such as statistics, pattern recognition and machine learning helps in counter-terrorism.

B. C*ustomer relationship management*

Data mining is essential for extracting the information of customer from the database. [5]CRM defined "The process of acquiring, retaining and growing profitable customer which requires a clear focus on service attribute that represents value to customers and creates loyalty. CRM consists of four dimensions are customer identification, attraction, retention and development. Data mining techniques support CRM are:

- Classification: Classification technique is used to predict future customer behavior.
- Association: Association is used for market basket analysis and for cross selling programs.
- Clustering: Clustering technique identify grouping of records and group of similar type of customers.
- Regression: Regression is used for predict values of dependent variables.
- Forecasting: predict the future values on the basis of record.
- Visualization: The complicated data can be presented in the visual form like graphs.

C. *Banking and Financial sector*

Banking system in India had started in 17^{th} century named Bank of Hindustan but the largest and oldest existing bank is State Bank of India. Now a time banking sector is an important part of life and especially business. Development of any country depends upon the banking system of that country. Data mining techniques are used in all banks in all over the world.

- Data mining techniques are used in predicting the profit, payment default, credit scoring, fraud transaction and marketing detection etc.
- In banking sector data mining techniques extract useful information from raw data and predict the future situations.
- Classifications algorithms help in classify the data into groups.
- Prediction algorithm are used fraud detection and in reducing credit risk.
- Association algorithm is used in market basket analysis, cross marketing etc.

Before use of data mining or data warehouses Total Branch Computerization (TBC) software packages was used.

D. *Retail Industry*

In the competitive environment, data mining is used in retail industry for making better strategies and decisions. Retail industry is to make good strategy for target right customer who is profitable to them. Retail industry involves large amount of data includes transportation, sales and consumption of goods and services. The main requirement of retail industry is information of customer requirement, cost, profit and quality of product.

- Data mining techniques like clustering, classification, regression and prediction are used in identifying customer behavior, shopping pattern etc.
- Sequential pattern mining technique is used to attract customer, adjustment of price and change consumption of customer.

E. Telecommunication Industry

With the increasing demand of technology, the telecommunication industry is also growing very fast. Telecommunication industry includes many communication services like cell phones, e-mails, fax etc. The telecommunication industry is integrated with computers, internet and other communication industries.

- Data mining techniques are used to produce effective results in this field.
- Visualization tool of data mining helps in view the graphs and charts of calling resources.
- Multi dimensional data analysis techniques are used to identify the data traffic and profit etc.
- Clustering and association techniques are used in analysis of telecommunication data which includes information of call, caller, receiver, location and duration etc.

Due to fast development in this industry, it becomes difficult to analysis the complex data. Data mining techniques are used for effective result.

III. Data Mining – The Future Challenges

With the increasing demand of technology, the huge amount of data is generated. To analysis/extract data and the security of important information is handling by the use of data mining techniques and algorithms. Data mining used various fields but still there are major research challenges in some fields like currency crisis, mobile computing, biomedical and DNA analysis, security and privacy etc.

A. Currency crises

It is well known fact that currency crises exposure a critical issues for public and private industries, organization and society, it is obvious that financial and currency crises can be extremely costly to the countries, institutions and business [6].

- Adaptive neuro fuzzy inference system (ANFIS) is method used to build early warning system and also used for predict and discovering the currency crises.
- Data mining technique, quantitative method is used for identify the currency crises.
- Bayesian networks, Ridor (Ripple DOwn Rules) method, C4.5 decision tree methods are used in identify the early currency crises.

B. Mobile computing

Mobile computers can be combined with wireless communication devices that allow users to access data services from any location. In wireless communication, the server-to-client (downlink) communication bandwidth is much higher than the client-to-server (uplink) communication bandwidth [7].

- Data mining techniques are used in analysis the broadcast history.
- Rule methods are used in improving the performance of the mobile services.
- Data mining techniques like clustering, associations, and sequences are used in extracting useful information in broadcast histories.
- Sequential pattern are used in data broadcasting and prefetching.
- In mobile computing, data mining techniques are utilized for the organization and efficient querying the web data [7].

The use of association rules in scheduling broad cast requests, prefetching and cache replacement or in any other similar application has not yet been researched.

C. Biomedical and data analysis

Data mining is widely used in the field of medical sciences such as Biomedical, DNA, Genetics and Medicines etc. In the field of Genetics, the relationship between genes in human DNA sequences.

- Data mining techniques and algorithms are used in improve diagnosis, prevention and treatment of the diseases.
- Data mining cleaning and integration approaches are used in systemize the DNA data which is heterogeneous and distributed in nature.
- Association analysis method is used in detecting the diseases from combination of genes.
- Complex genes structure can be understand by visualization tools.
- Path analysis is playing important rule in linking the different genes of different stages of disease development.

D. Security and Privacy Prevention

Security and Privacy are not new concept in data mining but still there is too much can be done this area with data mining. [8] Gives a thorough analysis of impact of social networks and group dynamics. Specifying the need to understand cognitive networks, he also models knowledge network using the Enron E-mail corpus. Recording of electronic communication like email logs, and web logs have captured human process. Analysis of this can present an opportunity to understand sociological and psychological process.

- Secure Multiparty Computation (SMC) can be used where multiple parties, each having private input, want to communicate.

E. Detecting Eco-System Disturbances

Study of this area is important due to radical changes in ecosystem has led to floods, drought, ice-storms, hurricanes, tsunami and other disasters. Land Cover Change detection is also one of the areas. In a press release by NASA shows the history of natural disasters [2]. Detection algorithm applied to large data set is very expensive. For this detection the clustering framework is used.

- Data mining technique clustering is used for grouping similar events, incidents of disturbance.
- Multilevel indexing scheme based on clustering is used to explore data in hierarchical fashion.

V. CONCLUSION

This review puts focus on the present and future area of data mining. This paper provides a new perspective to a researcher about applications of data mining used in social welfare. This study shows that data mining techniques in CRM, counter-terrorism, retail industry, banking and telecommunication industry provide better prediction ability to the organizations. Data mining will have major impact on currency crises, biomedical and DNA analysis, detecting eco-system disturbances and mobile computing will present challenges for future research.

REFERENCES

[1]. J. Han and M. Kamber. Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann, San Francisco, CA, 2012.

[2]. Darminder kumar, Deepak bhardwaj, “Rise of Data Mining: Current and Future Application Areas”, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 1, September 2011.

[3]. Walter Alberto Aldana, “Data Mining Industry, Emerging Trends and New Opportunities”, Master’s Thesis Massachusetts Institute of technology, 2000.

[4]. Simmi Bagga, Dr. G.N. Singh, “Application of data mining”, International Journal for Science and emerging technologies with latest trends, 2012.

[5]. S. Janakiraman1, K. Umamaheswari, “ A Survey on Data Mining Techniques for Customer Relationship Management”, International Association of Scientific Innovation and Research.

[6]. Dilek Karahoca , Adem Karahoca ,Ozerk Yavuz, “ An early warning system approach for the identification of currency crises with data mining techniques”, Neural Comput & Applic ,2013.

[7]. Yucel Saygin, Ozgur ulusoy, “ exploting data mining techniques for broadcasting data in mobile computing environment”, IEEE transaction on knowledge and data engineering, VOL 16, NO 6, 2002.

[8]. N. Pathak, S. Mane, J. Srivastava, N. Contractor, S. Poole & D. Williams, “Analysis of Social Networks and Group Dynamic from Electronic Communication”, In Next Generation of Data Mining, Taylor and Francics, 2008.

[9]. Vikas Jayasree and Rethnamoney Vijayalakshmi Siva Balan,“A REVIEW ON DATA MINING IN BANKING SECTOR”, American Journal of Applied Sciences,2013.

[10]. Kazi Imran Moin, Dr. Qazi Baseer Ahmed, “ Use of Data Mining in Banking”, International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622, Vol. 2, Issue 2,Mar-Apr 2012, pp.738-742.

[11]. Haibin Cheng, Pang-Ning Tan, Christopher Potter, Steven Klooster, “Data Mining for Visual Exploration and Detection of Ecosystem Disturbances.

Integrating Different Tools and Process Parameters in Reuse-Based Software Engineering

Priya Suneja
Department of Computer Science
CT Institute of Technology
Jalandhar, India
priya.suneja@hotmail.com

Ankur Chauhan
Department of Computer Science
CT Institute of Technology
Jalandhar, India
ankurtheengineer@gmail.com

Rajan Sahota
Department of Computer Science
CT Institute of Technology
Jalandhar, India
rajan.sahota@gmail.com

***Abstract*— Software reuse refers to a situation in which some software is used in more than one project. By reusing high quality software components, software developers can simplify the product and make it more reliable. Reusable modules and classes reduce implementation time, increase the likelihood that prior testing and use has eliminated bugs and localizes code modifications when a change in implementation is required. Although several processes have been investigated to develop reusable software, there are not available studies that compare them. In this way, this paper presents a detailed survey on software reuse processes.**

***Keywords*— Software reuse, reuse process, reuse technologies.**

I. INTRODUCTION

Code reuse is the use of existing software to build new software. Software reuse refers uses the previously developed software resources in new applications. Because fewer total lines of code need to be written, software reuse can increase productivity, reduce development costs, and minimize schedule overruns. Since reusable software resources should be rigorously tested and verified, reuse also has the potential to improve software quality, reliability, maintainability and portability. Typically, software reuse involves the reuse of portions of code, knowledge, design, code, test cases etc. by other programmers in the same organization. There are two approaches for reuse of code: develop the reusable code from scratch or identify and extract the reusable code from already developed code [5].

II. TYPES OF SOFTWARE REUSE

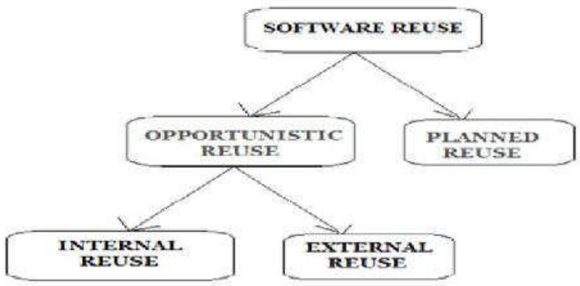


Fig.1. Types of Reuse [1]

A. *Opportunistic reuse*

This is the initial phase in which the existing components are identified and find whether they can be reused or not in future projects. This can be categorized into two parts.

B. *Internal reuse*

In this type,organisation considers its own component to reuse. So this may be aaplied to specific projects within the organizaion.

C. *External reuse*

When a team chooses thied party componen.It may consider as external reuse. But the main consider is on cost estimation because 20% of cost may be required for third party license.

D. *Planned reuse*

Planned reuse is reserved for future projects.

III. BENEFITS OF SOFTWARE REUSE

A. Increased dependability

Reused software should be more dependable than new software because reused software has been tried and tested in working systems. The primary use of the software is to find any design and implementation faults. These faults are then fixed, thus reducing the number of failures when the software is reused.

B. Reduced process risk

Software reuse reduces the margin of error in project cost estimation. If software exists, there is less uncertainty in the costs of reusing that software. This is particularly true for large software components such as sub-systems are reused.

C. Improvement in documentation and testing

Reusable components are normally accompanied by high quality documentation and by previously developed tests plans and cases. Documentation and tests will be much easier to develop when a new system is created by simple selection and altering of existing components.

D. Standards compliances

Standards compliances means if we reuse any single component of project then all the properties of that component will be inherited into whole project. Thereby creating a standard for all the users and it will improve dependability as users are less likely to make mistakes when presented with a familiar interface.

E. Accelerated development

It avoids the original development. The most important thing is to bring a system to market as early as possible. Reusing software can speed up system production because both development and validation time should be reduced.

F. Reliability and safety

The objective of software reuse is to improve the reliability. It is argued that reusable components, because of more careful design and testing and broader and more extensive usage, can be more reliable that one use equivalents. If so, then it is further argued that using these more reliable components in system architecture can increase the reliability of the system as a whole based on this idea that replacing error prone human processes in software development by automation can produce a more reliable system [3].

IV. TOOLS OF REUSABILITY

Software programming is a hard design task as complexity involved in the process. Reuse deals with the ability to combine independent software components to form a larger unit of software. To incorporate reusable components into a software system, programmers must be able to find and understand them. Thus Software reuse is software design, where the new systems are generated from previous components [2]. These are the three or four specific tools by Reusability and shown by figure 2.

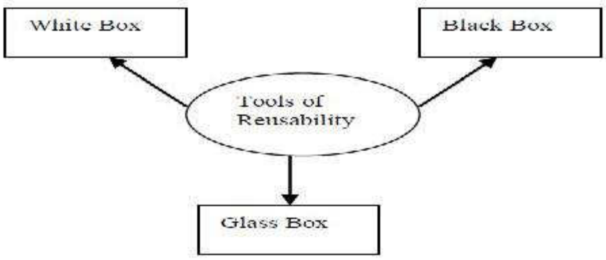


Fig.2 .Tools of reuse

A. White box reusability

White box reusability is a verification technique which is used by engineers to examine their code.

B. Black box reusability

In Black box reusability tool, programmer sees only the external features of software and all the implementation part is hidden from the programmer. So, we cannot change the logical behavior of system.

C. Glass box reusability

In Glass box reusability, implementation and interfaces of component is visible to the programmer.

V. REUSABILITY MATRICES AND MODEL

Software reuse, the use of existing software artifacts or knowledge to create new software is a key method for significantly improving software quality and productivity.

Reusability is the degree to which a thing can be reused. Software reuse reduces the amount of software that needs to be produced from scratch and thus allows a greater focus on quality. The reuse of well tested software should result in greater reliability and less testing time for new software. With reuse, software development becomes a capital investment. In this paper we survey metrics and models of software reuse and reusability. A metric is a quantitative indicator of an attribute of a thing[4]. A model specifies relationships among metrics. In reuse models and metrics are categorized into types:

- Reuse cost benefits
- Maturity assessment
- Amount of reuse
- Reusability
- Reuse library

VI. THE MAJOR STEPS IN A SOFTWARE REUSE PROCESS

The following some important steps are shown in Table 1.

Table I. Software Reuse Process

S.No.	SOFTWAR REUSE PROCESS
1	Specifying the object to be created
2	Searching the project, domain, and general databases for reuse candidates
3	Evaluating the candidates to determine which (if any) should be used
4	Modifying, if necessary, to fit specific needs
5	Integrating the reusable component(s)
6	Validating the system including the new component(s)
7	Feeding back the knowledge regarding the payoff of reuse

VII. LIMITATIONS OF SOFTWARE REUSE

Limitations for software reuse are as follows:

Table II. Limitations of Software Reuse

S.No.	LIMITATIONS
1.	Maintenance cost will be increased
2.	Needless complexity
3.	Domain irrelevance

4.	CASE toolsets do not support development with reuse

VIII. RESEARCH AREAS TO BE EXPLORED

From the standpoint of basic understanding of reuse technology, only an informal and incomplete picture of what to reuse and how to reuse it is known. Reuse experts are only beginning to characterize the total scope of information that can be utilized in software construction. Intensive research is needed to enhance the capabilities to capture, modify, and reuse a wide range of information at different stages in the software development process. This is true regardless of the software development paradigm (classic waterfall model, rapid prototyping, spiral model, etc.) that is used[6].

Research is especially needed in the following areas:

- Determination of proper search methods for reuse libraries. It is not clear if new, reuse-specific search methods are needed, or if standard library science techniques can be applied. If both approaches work, it is not clear which of the two is more efficient.
- Representation methods that are both flexible enough to incorporate existing representations and powerful enough to allow capture of appropriate knowledge.
- Accurate cost models that take into account COTS products, reuse of artifacts at multiple life cycle levels and describe both consumer and producer reuse.

IX. CONCLUSION

In this paper we surveyed different aspects of reusability for tools, metrics, models and process of software reuse. We have reviewed the current trends and existing problems, and specific difficulties, which are helpful for increasing reuse and reusability. Software reuse is not a new method; many programmers have copied and pasted snippets of code since early days of programming. Even though it might speed up the development process, this “code snippet reuse” is very limited does not work for larger projects. The full benefit of software reuse can only be achieved by systematic reuse that is conducted formally as an integral part of the software development cycle. In future the work proposed here can be used by researchers for taking this topic of software reuse to new heights where the world will be totally dependent on these types of technologies.

REFERENCES

[1] Ms. Kirandeep Kaur, Ms.Rekha Rani, Ms.Jagdeep kaur, “Code Reuse & Reusability of the Software”, International Journal of Engineering and Science (IJES), 15April, 2013.

[2] Ashwin B. Tomar, V. M. Thakare, “A Study of Software Reuse and Models”, National Conference on Innovative Paradigms in Engineering & Technology (NCIPET-2012) Proceedings published by International Journal of Computer Applications® (IJCA), PP: 14-17, 2012.

[3] Neha Budhija , Satinder Pal Ahuja,”Review of Software reusability”, International Conference on Computer Science and Information Technology,Dec,2011

[4] Sarbjeet Singh, Manjit Thapa, Sukhvinder singh and Gurpreet Singh ,“Software Engineering - Survey of Reusability Based on Software Component”, International Journal of Computer Applications (0975 – 8887), Volume 8, No.12, PP: 39-42, 2010.

[5] Sarbjeet Singh, Sukhvinder Singh, Gurpreet Singh, “Reusability of the Software”, International Journal of Computer Applications (0975 – 8887), Volume 7, No.14, PP: 38-41, 2010.

[6] Sajjan G. Shiva, Lubna Abou Shala,” Software Reuse: Research and Practice”, International Conference on Information Technology (ITNG'07), IEEE 2007.

Comparative Analysis of Different Actors in Cloud Computing Reference Model

Ankur Chauhan
Dept. of Computer Science
CT Institute of Technology
Jalandhar, India
ankurtheengineer@gmail.com

Priya Suneja
Dept. of Computer Science
CT Institute of Technology
Jalandhar, India
priya.suneja@hotmail.com

Rajan Sahota
Dept. of Computer Science
CT Institute of Technology
Jalandhar, India
rajan.sahota@gmail.com

Rachna Lekh
Dept. of Computer Science
CTIEMT, Jalandhar, India
rachna_lekh@ymail.com

Sunil Kumar
Dept. of CSE
S.V. Subharti University
Meerut
sunilcloudresearch@gmail.com

***Abstract*—Cloud Computing is the delivery of computing services over the Internet. Cloud services allow individuals and businesses to use software and hardware that are managed by third parties at remote locations. Cloud computing is a compilation of existing techniques and technologies, packaged within a new infrastructure paradigm that offers improved scalability, elasticity, business agility, faster startup time, reduced management costs, and just-in-time availability of resources. In this paper our contribution is survey on the vision of cloud computing along with cloud reference model. We also emphasize on the actors/performers in cloud computing. This paper also included the causes of problems associated with cloud computing.**

Keywords—Cloud computing, remote locations,cloud services.

I. INTRODUCTION

Cloud computing is a model for convenient, demand network access to a shared environment of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. Cloud computing is the dynamic delivery of Information technology resources and capabilities as a service over the internet. It is a style of computing in which dynamically scalable and often virtualized resources are provided as service over the Internet.

A. Characteristics Of Cloud Computing

Cloud computing have five main characteristics defined by NIST (National Institute of Standards and Technology) [1].

1) On-demand self-service

A consumer can unilaterally provision computing capabilities.

2) Broad network access

Capabilities are available over the network and accessed through specialized protocols that promote use by different thin or thick client platforms.

3) Resource pooling

The provider's computing resources are pooled to serve multiple cloud consumers, with separate physical and virtual resources assigned and reassigned dynamically according to consumer demand.

4) Rapid elasticity

Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out and rapidly released to quickly scale in.

5) Measured service

Cloud systems automatically control and optimize resource use by leveraging a metering capability

at some level of abstraction appropriate to the category of service e.g., bandwidth, processing speed, storage capability and current user accounts.

B. Benefits Of Cloud Computing

Applications and services that are based on Cloud computing have several benefits:

1) Cost Savings

The companies can reduce their capital operational expenditures.

2) Computing Capabilities

This is a lower barrier to entry and also requires fewer in-house IT resources to provide system support.

3) Scalability/Flexibility

The companies can start with a small deployment and grow to a large deployment rapidly, and then scale back according to necessity. Also, the flexibility of cloud computing allows companies to use extra resources at peak times, enabling them to satisfy consumer needs.

4) Automated backups, SLA , maintenance

The provider's cloud computing services do the system maintenance, and access is through APIs that do not require application installations onto computers, thus reduce maintenance requirements.

5) Web and mobile-access from remote locations

Mobile workers have increased productivity due to systems accessible in an infrastructure available from anywhere.

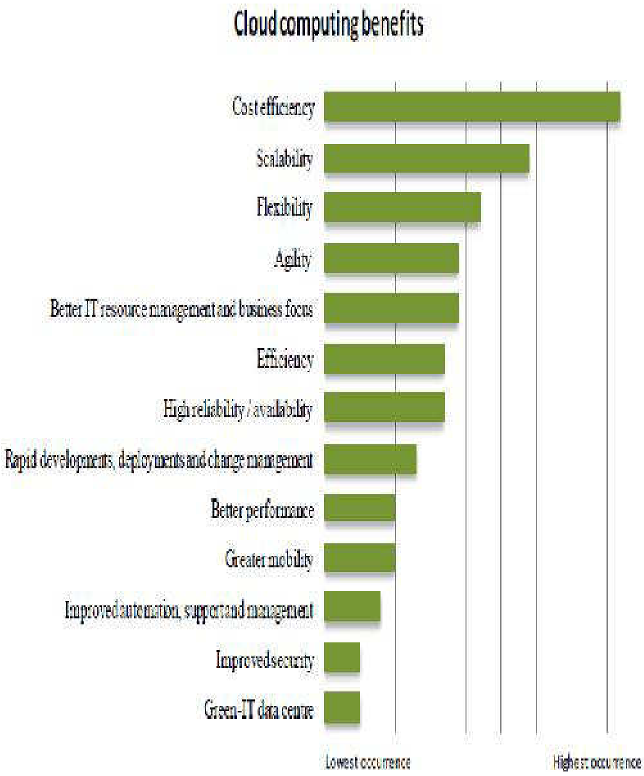


Fig.1. Cloud computing benefits [2]

II. VISION OF CLOUD COMPUTING

With the advancement of modern society basic essential services and utilities are commonly provided such that everyone can easily obtain access to them. Today utility services like electricity, gas, water and telephony are basic necessary for fulfilling daily life routines. These utilities are accessed severely that they need to be available whenever the consumer requires them at any time. Then consumers are able to pay service providers based on their usage of these utility services[7].

Consumers will be able to access applications and data from a "cloud" anywhere in the world on the basis of their demand. The consumers who avail the services are promised that the cloud infrastructure is very robust and will always be there for the required services at any time. The vision of cloud computing are:

- Service provisioning model. Computing service users or consumers need to pay providers only when they access computing services. In addition, consumers no longer need to invest heavily or encounter difficulties in building and maintaining complex IT infrastructure. Hence, software practitioners are facing numerous new challenges toward creating software for millions of consumers to use as a service, rather than to run on their individual computers.
- Computer utilities. The creation of the internet has marked the foremost milestone towards achieving this grand 21^{st} century vision of 'computer utilities' by forming a worldwide system of computer networks that enables individual computers to communicate with any other computer elsewhere in the world. This internetworking of standalone computers provide potential of utilizing endless amount of distributed computing resources owned by various owners.
- Grid computing. Grid computing enables the sharing, selection, and aggregation of wide variety of physically distributed resources including supercomputers, storage systems, data sources, and specialized devices owned by different organizations for solving large scale resource problems in science, engineering, and commerce.
- Peer-to-peer computing. The peer-to-peer (P2P) computing allows peer nodes (computers) to share content directly with one another in a decentralized manner. In pure P2P computing, there is no notion of clients and servers. The goals of P2P computing include cost sharing or reduction, resource aggregation and interoperability, improved scalability and reliability, increased autonomy, anonymity or privacy, dynamism, ad-hoc communication and collaboration.
- Service computing. A service computing focuses on the linkage between business processes and IT services so that business processes can be seamlessly automated using IT services. Examples of service computing technologies include service-oriented architecture (SOA) and web services. The SOA facilities interoperable services between distributed systems to communicate and exchange data with one another, thus providing a uniform means for service users and providers to discover and offer services. The web services provides the capability for self-contained business functions to operate over the internet.
- Market-oriented computing. It views computing resources in economic terms such that resource users will need to pay resource providers for utilizing the computing resources. Therefore, it is able to provide benefits, such as offering incentive for resource providers to contribute their resources for others to use and profit from it,

regulating the supply and demand of computing resources at market equilibrium, offering incentive for resource users to back off when necessary, removing the need for central coordinator.

- Virtualized compute and storage technologies. Today the latest paradigm to emerge is that of cloud computing which promises reliable services delivered through next-generation data centers that are built on virtualized compute and storage technologies.

III. ADVANTAGES OF CLOUD COMPUTING

Table I. Positives of cloud computing

Parameters	Cloud Computing
Data Confidentiality	Embed encryption, firewalls
Data Availability	Uses multiple cloud providers
Performance	High performance, Improved VM support
Scalability	Invent scalable storage
Data Transfer	Higher bandwidth switches
Storage capacity	Unlimited storage capacity
Data reliability	Increased data reliability as data on cloud
Cost	Reduced software cost, pay-for-use licenses

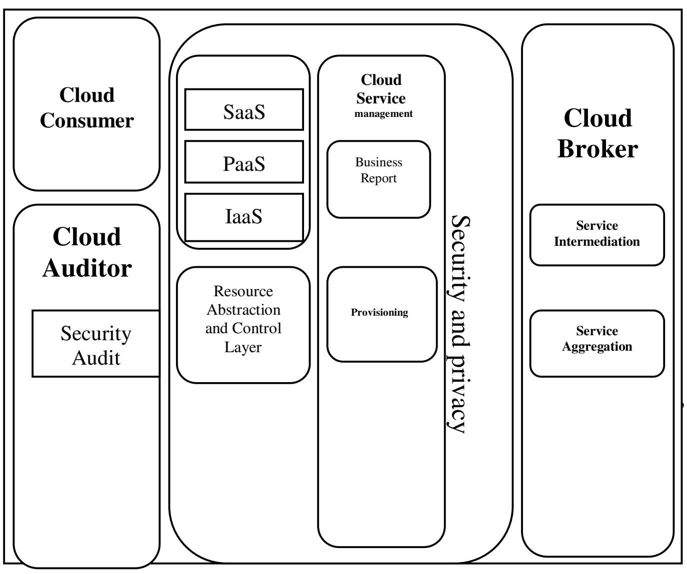
IV. DISADVANTAGES OF CLOUD COMPUTING

Table II. Limitations of Cloud Computing

Parameters	Cloud Computing
Internet Connection	Requires constant internet connection
Speed	Does not work well with low speed connection
Security	Less security as the data stores in third party server

V. CLOUD COMPUTING REFERENCE MODEL

NIST cloud computing reference architecture is which identifies the major performer/actor their activities and functions of cloud computing as shown in above figure.The Cloud computing reference model represents a generic high-level architecture and is intended to facilitate the understanding of the requirements, uses, characteristics and standards of cloud computing [4].



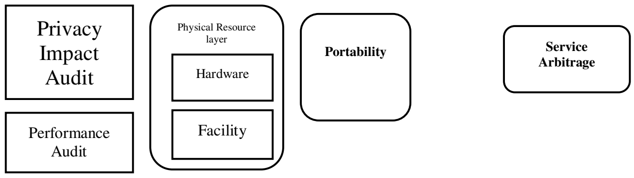


Fig 2. The Conceptual Reference Model

The actors/performers in cloud computing are:

- Cloud consumer.

A cloud consumer represents a person or organization that maintains a business relationship with and uses the service from a cloud provider. A cloud consumer browses the service catalog from a cloud provider, requests the appropriate service, sets up service contracts with the cloud provider and uses the service.

- Cloud provider.

A cloud provider is a person, an organization which is the entity responsible for making a service available to interested parties. A cloud provider acquires and manages the computing infrastructure required for providing the services, and makes arrangement to deliver the cloud services to the cloud consumers through network access.

- Cloud auditor.

A cloud auditor is a party that can perform an independent examination of cloud service controls with the intent to express an opinion thereon. Audits are performed to verify conformance to standards through review of objective evidence. A cloud auditor can evaluate the services provided by a cloud provider in terms of security controls, privacy impact, performance etc.

- Cloud broker.

As cloud computing evolves the integration of cloud services can be too complex for cloud consumers to manage. A cloud consumer may request cloud services from a cloud broker, instead of contacting a cloud provider directly. A cloud broker is an entity that manages the use, performance and delivery of cloud services and negotiates relationships between cloud providers and cloud consumers.

- Cloud carrier.

A cloud carrier acts an intermediary that provides connectivity and transport of cloud services between cloud consumers and cloud providers. A cloud carriers provide access to consumers through network, telecommunication and other access devices[8].

VI. PROBLEM STATEMENTS ASSOCIATED WITH CLOUD COMPUTING

- Loss of Control

Data, applications, resources are located with provider. Therefore, identity of user is overly managed by the cloud. All the security policies, access control rules are managed by cloud provider.

- Lack of Trust :

As consumer data, applications and resources are handled by the third party. Therefore, one has to do trust on the services that are provided by any cloud provider. And trusting a third party requires taking risks. It's hard to balance trust and risk level.

- Multi-Tenancy issues in cloud computing :

Multi-tenancy refers to a principle in software architecture where a single instance of the software runs on a server, serving multiple client-organizations (tenants). Tenants share a pool of resources and have opposite goals. Therefore sometime there occurs a conflict between tenant's opposite goals.

VII. CONCLUSION

In this paper we survey different approaches of cloud computing. Also, we discussed regarding the characteristics, advantages, disadvantages related to the arena cloud computing. Advancement of this field brought revolution in the field of Information Technology. NIST gives the cloud computing reference model which explains the actors and performers in cloud computing. We consider the causes of problems that are associated with the cloud computing. In spite of this, there are challenges that are related to the security of cloud computing. Security is the major issue in the arena called cloud computing. As consumer data is stored in the third party server and there is risk of availability assurance, maintenance and leakage of information from the provider server. In future we believe that our paper will provide better understanding of the cloud computing and researchers can work on the issues related to the security in the cloud computing.

REFERENCES

[1] The NIST Definition of Cloud Computing, version 15, by Peter Mell and Tim Grance, October 7, 2009, National Institute of Standards and Technology (NIST), Information Technology Laboratory (www.csrc.nist.gov)

[2] Wang, Lizhe; von Laszewski, Gregor; Kunze, Marcel; Tao, Jie. Cloud computing: A Perspective study, Proceedings of the Grid Computing Environments (GCE) workshop. Held at the Austin Civic Center: Austin, Texas: 16 November 2008.

[3] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, Matei Zaharia. A view of cloud computing. Communications of the ACM Volume 53 Issue 4, pages 50-58. April 2010.

[4] Open Security Architecture http://www.opensecurityarchitecture.org/

[5] Shilpashree Srinivasamurthy, David Q. Liu Indiana University – Purdue University Fort "Survey on Cloud Computing Security" http://thecloudtutorial.com/related.html

[6] Cloud Confusion Amongst IT Professionals. Version One June 6,2011

Review: Software Defined Networking

Sukhveer Kaur
Computer Science and Engineering, SBS State Technical Campus, Ferozepur, INDIA
bhullarsukh96@gmail.com

Karamjeet Kaur
Computer Science and Engineering, SBS State Technical Campus, Ferozepur, INDIA
bhullar1991@gmail.com

Japinder Singh
Computer Science and Engineering, SBS State Technical Campus, Ferozepur, INDIA
japitaneja@gmail.com

Navtej Singh Ghumman
Computer Science and Engineering, SBS State Technical Campus, Ferozepur, INDIA
navtejghumman@yahoo.com

***Abstract*—This review paper explains Software Defined Networking (SDN) where control and data plane are decoupled. Software based controller is responsible for centrally managing forwarding logic in data plane. The data Plane only handles rules specified in controller. OpenFlow is protocol used for communication between control and data plane. SDN is an alternative way to Design, Build and Operate Networks. SDN offers network programmability required for dynamically configuring and controlling networks. It replaces static, inflexible and complex networks with networks that are agile, scalable and innovative. This paper discusses various aspects of Software Defined Networking. First we give introduction about SDN, then discuss architecture of SDN, describe OpenFlow standard and list the benefits and challenges of SDN.**

Keywords- SDN, OpenFlow

I. INTRODUCTION

Internet has evolved to offer wide variety of distinct services such as IP Telephony, IP-TV, Online Commerce, Content delivery. These services require different type of parameters for the networks. This stresses our current network. It is also hard to evolve and accommodate new applications and technologies[1].

The problem with traditional networks is that software is bundled with the hardware and interfaces are vendor specific. Vendors write the code and there are long delays in introducing new features and functions. Implementing networks is expensive. In Traditional switches there was tight coupling between control and forwarding planes. In this type, due to tight coupling introduction of new applications and functionality was very difficult. Configuration of network devices was also very difficult due to lack of common control interface.

SDN separates the forwarding plane (data plane) from the control plane (Fig. 1). SDN was developed to enable innovation and simple programmatic control of forwarding plane. Separation of forwarding plane from control plane allows easier deployment of new applications and simplified protocol management[2].

Software Defined Networking (SDN) is futuristic architecture that is powerful, easy to manage, less expensive and flexible, making it suitable for today's dynamic, high bit rate applications. SDN allows control plane to be programmable. It abstracts underneath infrastructure. The OpenFlow protocol is basis for creating SDN solutions [12].

II. SOFTWARE DEFINED NETWORKING ARCHITECTURE

Data Plane, Control Plane and SDN applications are the three components of SDN architecture. The Data Plane consists of OpenFlow based physical or virtual switches. Decoupling has allowed the control plane to evolve independently resulting in short evolutionary cycle. This led to development of various control technologies. The forwarding plane evolution cycle aimed at faster packet delivery. OpenFlow switches separate the forwarding and control plane. The separated control plane is called SDN controller.

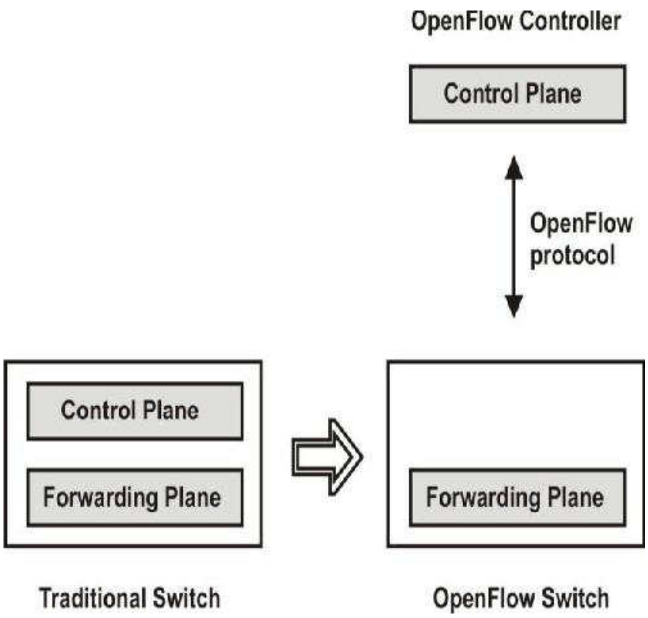


Fig. 1. Decoupling of forwarding and control plane

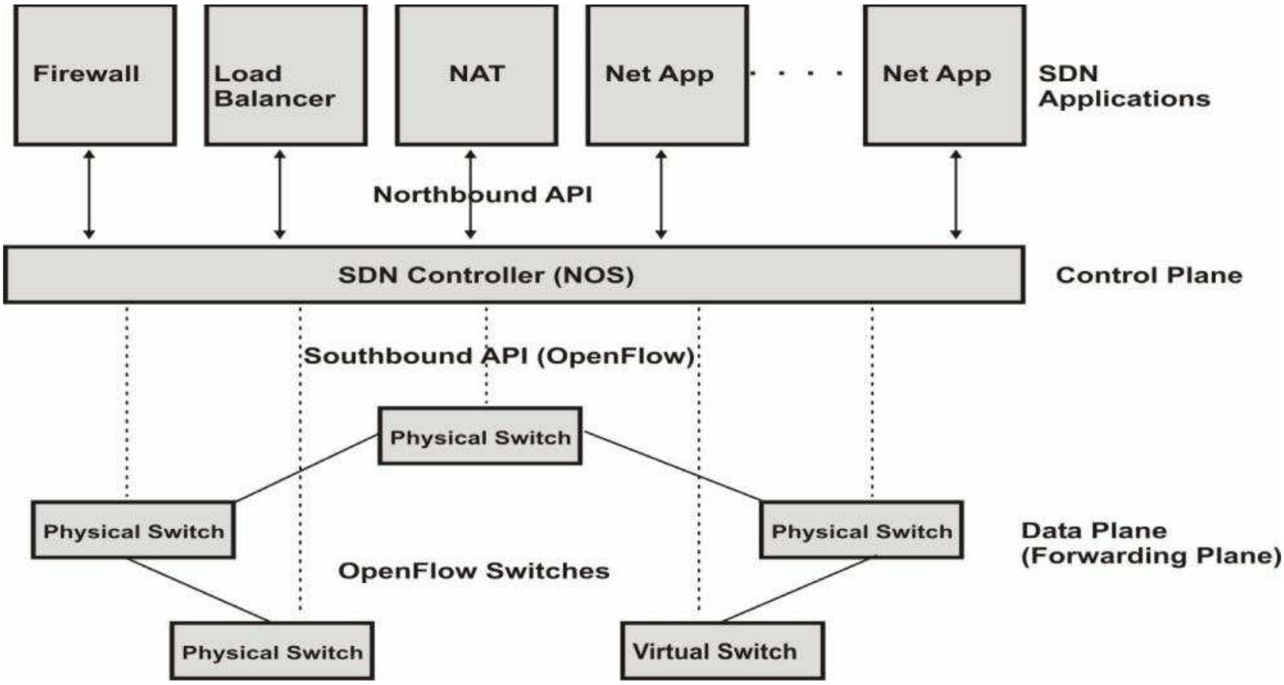


Fig. 2. SDN Architecture

Northbound interface allows applications to interact with the controller. Northbound interface is used for creating Firewall, Load Balancer, NAT applications. The SDN controller interacts with the OpenFlow switches using Southbound interface (Fig. 2). OpenFlow is the most popular Southbound interface[21], [22].

The controller is a software program responsible for manipulating the switch flow table, using the OpenFlow protocol. The secure channel is the interface that connects the controller to all switches. Through this channel, the controller manages the switches, receive packets from the switches and send packets to the switches. An OpenFlow compliant switch

must be capable of forwarding packets according to the rules

defined in the flow table[23].

SDN Controller is also defined as Network operating system (NOS)[20]. It allows abstracting the complexity of the underlying network. The idea was precisely to draw a parallel between the network operating system and the typical operating system. In an operating system, the abstraction includes the hardware components of CPU. In a network, the abstraction hides the topology and the network devices.

The information about various controllers, the languages[24] in which they are written, original developers plus brief description is listed in (Table I).

The SDN can be categorized as

i. Reactive SDN: When a new packet enters into the switch, switch performs a lookup in flow tables. If no flow match occurs switch sends the packets to the controller. OpenFlow controller creates a rule in the flow table. It makes efficient use of flow table but each flow requires additional flow setup time. The downside is that if control connection is lost, switch has limited utility.

ii. Proactive SDN: Rather than reacting to a packet, an OpenFlow controller prepopulate flow table by pre-defining all of your flows and actions ahead of time in the switches flow tables, the packet-in event never occur [3]. Advantages are no additional flow setup time plus loss of connection does not affect switches.

Centralized vs Distributed control: In the centralized control, single controller talks to distributed data planes. NOX[15], POX[16], OVS[14], Trema[17], Floodlight[18], Flowvisor[19], Beacon[4] and Maestro[5]. Centralized control plane represents a single point of failure. SDN allows the use of multiple controllers for controlling the switches. Onix[6], HyperFlow[7] and Kandoo[8] represents the idea of distributed controllers.

Packet is the basic unit of networking. It contains IP address information necessary for switch to make forwarding decisions. SDN treats similar individual packet as a flow. OpenFlow provides flow based control. Control can be further categorized as per flow and aggregated flow control [11].

TABLE I. SDN CONTROLLERS

Name	Language	Original Developers	Description
Ovs	C	Stanford/Nicira	A reference controller. Act as a learning switch
NOX	C++	Nicira	The first OpenFlow controller
POX	Python	Nicira	Open source SDN controller
Beacon	Java	Stanford	A cross platform, modular OpenFlow controller
Maestro	Java	Rice	Network operating system
Trema	Ruby, C	NEC	A framework for developing OpenFlow controller
Floodlight	Java	BigSwitch	OpenFlow controller that work with physical and virtual OpenFlow switches
Flowvisor	C	Stanford/Nicira	Special purpose controller

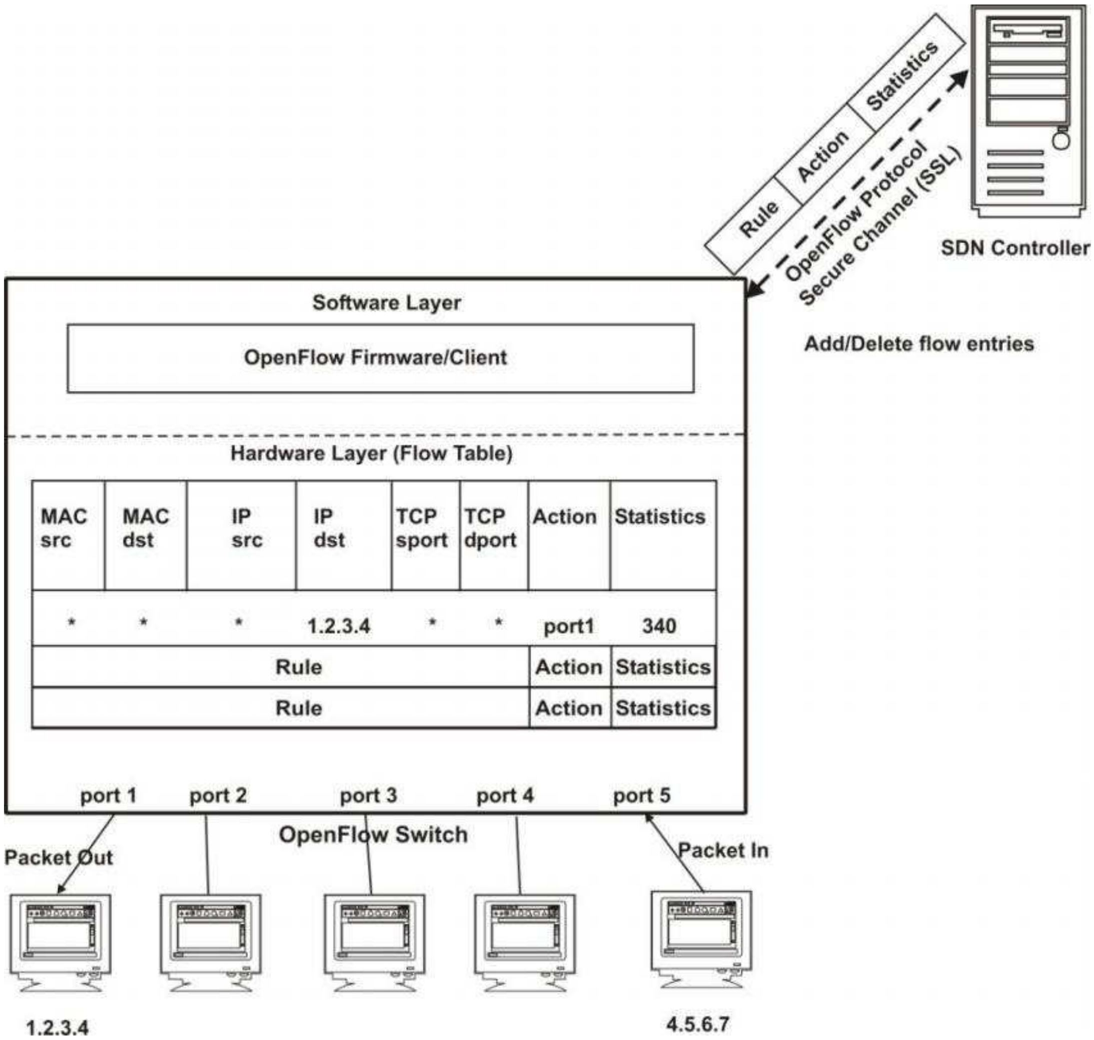


Fig. 3. OpenFlow Architecture

In per flow based routing, controller individually sets up every flow. Flow table contains one entry per flow. It performs exact match for flow entries. It is good for fine grain control. In aggregated flow routing, one flow entry covers a large groups of flows. There is one entry for each category of flows. It uses wildcard flow entries[10].

III. OPENFLOW

The OpenFlow specification allows software applications to program flow table of different switches. The OpenFlow architecture consists of three main components: an OpenFlow compliant switch, secure channel and a controller.(Fig 3) Switches use flow tables to forward packets. A flow table is a list of flow entries. Each flow entry consist of rules (match field), actions, statistics which decides how packets belonging to flow will be handled. Flow entry consists of 1) Rules used to match incoming packet. Match field(Table II) can be ingress port, dst MAC address, dst IP. 2) Actions or set of instructions determines how to handle matching packets(Table III). 3) Statistics used to collect parameters such as number of received packets, duration of flows[2], [25].

TABLE II. HEADER FIELDS

Ingress port
Ethernet source address
Ethernet destination address
Ethernet type
VLAN id
VLAN priority
IP source address
IP destination address
IP protocol
IP ToS bits
Transport source port/ICMP type

TABLE III.ACTIONS

Action	Description
Forward	Forward packets to a specified port
Enqueue	Forward packet through queue
Drop	Drop the packet
Flood	Forward packets to all ports except incoming port
Modify Field	Rewrite headers
Local	send packet to local switch networking stack
Controller	encapsulate and send the packet to controller

When packets arrive at switch it is matched against flow table entries. If match is found, appropriate action is performed that is specified in flow entry. If no match is found, action is taken as specified in flow table. The action could be dropping the packet or forwarding the packet to the controller over secure channel. Controller can then add, update or delete flow entries. This can happen reactively or proactively. By using combination of flow entries, different functionality such as firewalling, switching, routing, multicasting and flow switching can be implemented (Table IV).

OpenFlow switches can be of two types: native and OpenFlow enabled. Native openflow switches have no traditional feature set and completely depend upon controller for forwarding decisions. OpenFlow enabled switches support OpenFlow along with traditional feature set[9].

IV. BENEFITS & CHALLENGES

A. *Benefits*

1) Separating control plane and forwarding plane allows us to use commodity hardware as forwarding plane.
2) It allows you to quickly test and deploy new applications and protocols in real time.
3) Minimizes capital and operating expenses.
4) It allows centralized management of each switch.
5) By using SDN, we can prioritize different types of traffic[26].

TABLE IV.FLOW ENTRIES

	Ingress Port	MAC src	MAC dst	Eth type	VLAN ID	IP src	IP dst	TCP sport	TCP dport	Action
Flow Switching	port8	00:2e..	00:1f..	0800	vlan1	1.2.3.4	5.6.7.8	17264	80	port4
Firewall	*	*	*	*	*	*	*	*	22	drop
Switching	*	*	00:1f..	*	*	*	*	*	*	port4
Routing	*	*	*	*	*	*	5.6.7.8	*	*	port5
Multicast	*	*	00:1f	*	vlan1	*	*	*	*	port4 port5 port6
Packet In	*	*	*	*	*	*	*	*	*	controller

Flow Entries in Flow Table

A. Challenges

1) The main challenges with SDN are that it's new. Due to this infancy, many believe that SDN implementations are not ready for prime time.

2) The other challenge is of Control plane scalability and resilience.

3) By decoupling the control plane from the data plane, SDN introduce new attack surfaces such as controller APIs and protocols[13].

V. CONCLUSION

Although SDN is often termed as the solution to all networking problems but actually it is just a tool for managing network problems easily. It just places power in our hands to develop new protocols and applications to solve exiting problems. Future research areas are: Control Plane Programmability, Protocols that access, control and interact with the control plane, controller application interfaces (Northbound API), scalability and performance, controller processing delay, flow table size.

REFERENCES

[1] SUZUKI, Kazuya, Kentaro SONODA, Nobuyuki TOMIZAWA, Yutaka YAKUWA, Terutaka UCHIDA, Yuta HIGUCHI, Toshio TONOUCHI, and Hideyuki SHIMONISHI. "A Survey on OpenFlow Technologies." IEICE TRANSACTIONS on Communications 97, no. 2 (2014): 375-386.

[2] Nunes, B.; Mendonca, M.; Nguyen, X.; Obraczka, K.; Turletti, T., "A Survey of Software-Defined Networking: Past, Present, and Future of Programmable Networks," Communications Surveys & Tutorials, IEEE , vol.PP, no.99, pp.1,18.

[3] Proactive, Reactive SDN at :http://networkstatic.net/openflow-proactive-vs-reactive-flows

[4] Erickson, David. "The beacon openflow controller." In Proceedings of the second ACM SIGCOMM workshop on Hot topics in software defined networking, pp. 13-18. ACM, 2013.

[5] EugeneNg, ZhengCai AlanL Cox TS. "Maestro: Balancing Fairness, Latency and Throughput in the OpenFlow Control Plane."

[6] Koponen, Teemu, Martin Casado, Natasha Gude, Jeremy Stribling, Leon Poutievski, Min Zhu, Rajiv Ramanathan et al. "Onix: A Distributed Control Platform for Large-scale Production Networks." In OSDI, vol. 10, pp. 1-6. 2010.

[7] Tootoonchian, Amin, and Yashar Ganjali. "Hyperflow: a distributed control plane for openflow." In Proceedings of the 2010 internet network management conference on Research on enterprise networking, pp. 3-3. USENIX Association, 2010.

[8] Hassas Yeganeh, Soheil, and Yashar Ganjali. "Kandoo: a framework for efficient and scalable offloading of control applications." In Proceedings of the first workshop on Hot topics in software defined networks, pp. 19-24. ACM, 2012.

[9] Porras, Philip, Seungwon Shin, Vinod Yegneswaran, Martin Fong, Mabry Tyson, and Guofei Gu. "A security enforcement kernel for OpenFlow networks." In Proceedings of the first workshop on Hot topics in software defined networks, pp. 121-126. ACM, 2012.

[10] Northbound API at http://networkstatic.net/the-northbound-api-2/

[11] OpenFlow Specification at http://archive.openflow.org/documents/openflow-spec-v1.0.0.pdf

[12] OpenFlow definition at https://www.opennetworking.org/sdn-resources/onf-specifications

[13] SDN advantages at http://www.csoonline.com/article/2134045/mobile-security/sdn-the-security-pros-and-cons-of-using-it-in-your-organization.html

[14] Open Networking Foundation. 2013. OpenFlow Switch Specification 1.4.0. Technical Report. Open Networking Foundation.

[15] N. Gude, T. Koponen, J. Pettit, B. Pfaff, M. Casado, N. McKeown,and S. Shenker. Nox: towards an operating system for networks. ACM SIGCOMM Computer Commun. Review, 38(3):105–110, 2008.

[16] Pox. http://www.noxrepo.org/pox/about-pox/

[17] Trema openflow controller framework. https://github.com/trema/trema

[18] Floodlight, an open sdn controller. http://floodlight.openflowhub.org/ [19] R. Sherwood, M. Chan, A. Covington, G. Gibb, M. Flajslik, N.

[19] Handigol, T.Y. Huang, P. Kazemian, M. Kobayashi, J. Naous, et al. Carving research slices out of your production networks with openflow. ACM SIGCOMM Computer Commun. Review, 40(1):129–130, 2010.

[20] Shenker, Scott, M. Casado, T. Koponen, and N. McKeown. "The futureof networking, and the past of protocols." Open Networking Summit (2011).

[21] Shin, Myung-Ki, Ki-Hyuk Nam, and Hyoung-Jun Kim. "Software-defined networking (sdn): A reference architecture and open apis." In ICT Convergence (ICTC), 2012 International Conference on, pp. 360-361. IEEE,2012.

[22] Kirkpatrick, Keith. "Software-defined networking." Communications of the ACM 56, no. 9 (2013): 16-19.

[23] Vaughan-Nichols, Steven J. "OpenFlow: The next generation of the network?." Computer 44, no. 8 (2011): 13-15.

[24]] Foster, Nate, Arjun Guha, Mark Reitblatt, Alec Story, Michael J. Freedman, Naga Praveen Katta, Christopher Monsanto et al. "Languages for software-defined networks." Communications Magazine, IEEE 51, no. 2 (2013): 128-134.

[25] McKeown, Nick, Tom Anderson, Hari Balakrishnan, Guru Parulkar, Larry Peterson, Jennifer Rexford, Scott Shenker, and Jonathan Turner. "OpenFlow: enabling innovation in campus networks." ACM SIGCOMM Computer Communication Review 38, no. 2 (2008): 69-74.

[26] Casado, Martin, Teemu Koponen, Scott Shenker, and Amin Tootoonchian. "Fabric: a retrospective on evolving SDN." In Proceedings of the first workshop on Hot topics in software defined networks, pp. 85-90.ACM, 2012.

An Overview: Sky Computing

Himani
Dept. of CSE
Guru Nanak Dev University, Punjab, India
er.himani26@rediffmail.com

Harmanbir Singh Sidhu
Dept. of CSE
Guru Nanak Dev University, Punjab, India
harmansidhu27@yahoo.co.in

***Abstract*—IT infrastructure increasing without concerning about constraints like location, power, heating/cooling and other has overcome the best expectations from just a few years ago. So, the Cloud has opened a new view of development by combining multiple Clouds i.e is called Sky Computing. The management of these multiple clouds and applications are upcoming challenge.**

Keywords*—Cloud computing, Infrastructure-as-a-service, Sky computing, High-Performance Computing.

I. LITERATURE SURVEY

A. *Introduction*

The definition of Sky Computing arises as a metaphor to illustrate a layer above Cloud computing, because such dynamically provisioned distributed domains are built over several clouds [1]. It can be described as a management upper layer of an environment of Clouds, offering variable computing capacity and storage resources with dynamic support to real-time demands. Laying a virtual site over distributed resources, combining the ability to trust remote sites with a trusted networking environment [1], originates a highly elastic response to incoming requests with a seemingly infinite pool of accessible resources. The big challenge here is managing the combined resources of available Clouds, public or private and interoperating between them. The use of those resources to make intensive computing efficiently is also a requirement; the final goal is to support HPC applications.

Applications as HPC (High Performance Computing) and other scientific approaches require intensive computing, which can be compromised in the Cloud. Currently, these problems are solved with powerful data centers and grid clusters.

II. FINDINGS

A. *Using The Sky Computing*

The advantages of using intensive computing in the Cloud are the same as other applications. In addition to that, they benefit from application resilience, heterogeneity, clean computing, enhanced software development and dynamic workload migration [4]. Cloud computing is also an excellent collaboration platform that allows scientists to share information globally and replicate others' work with identical application, datasets and environmental settings [3]. However,

Cloud computing brings another relevant problem to the stake: failure either on software or hardware is critical to intensive computing and probabilities of such scenario increase with the mass addition of virtual machines; every task must be assured to complete or to resume if still unfinished. This can be guaranteed by implementing checkpoints or by predictive analysis [5], which will address the problems correctly, by resuming jobs or by allocating resources prior to other resources' fails.

The main idea in this paper is to create a turn-around model to enable intensive computing in Cloud networks. This is hoped to be achieved by enlarging the set of available resources in a way they overcome the problems referred before, like elevated latency between nodes. Also, it must be cross Cloud provider in order to combine resources. To achieve this, there must be a structure capable of receiving instructions, process and return results from all different underlying cloud systems. The proposed architecture is represented in Figure 1.

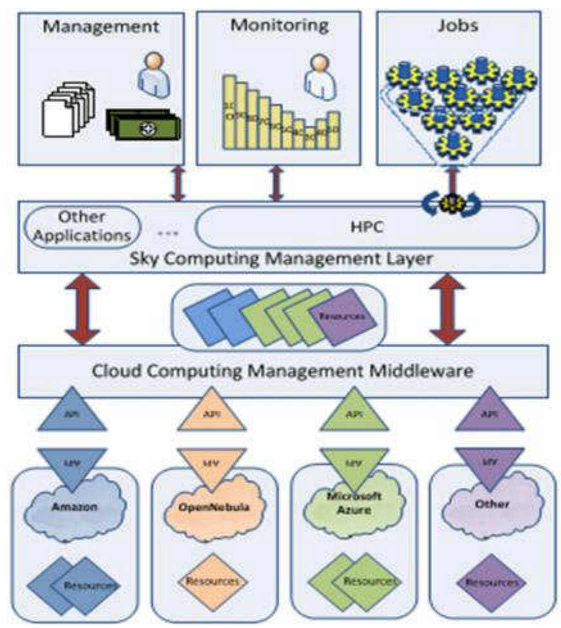


Fig. 1. Sky computing proposed architecture

As we can see in Figure 1, each Cloud provider has a specific API that makes available an interaction with their own resources. All these can be aggregated by a middleware layer, which allows controlling and managing resources by translating every command to the correspondent provider API. Abstraction, from bottom to top, is the key for building a consistent system. With the established middleware, it should be possible to use the available resources across any provider and seamlessly increase/decrease resources, adapting to demand.

The upper layer, Sky Computing, integrates the last level of Infrastructure as a Service and the next layer of Software as a Service. It allows scheduling and distributing resources to inputted tasks/requests made. This is a critical layer, as it must be as comprehensive as possible in features and capabilities. Here, our main focus is HPC, but is must be possible to deal with other applications too. Management, with scheduling, accounting and billing, should be well developed as well as Monitoring and Job submission.

B. Middleware Layer

There are some projects aiming to solve the interface problem created by the absence of standards [6]. Some are public and can be tested, like libcloud[7], Deltacloud[8], jclouds[9], or fog [10], while others, like IBM Altocumulus [11], are known to exist but with few public information.

The libcloud project is a client library for accessing clouds, with a wide range of providers, more than 20 in Python and 7 in Java. As a library, it relies on other projects to implement a UI, other than command line. Some of them are not free, as CloudKick [12].

As for jclouds, it is also a library targeting portable abstractions for cloud access. Current version is 1.0-beta-9 and present stable UI implementation – ComputeService and BlobStore, all open-source as well. Unfortunately for us, it doesn't support our private cloud provider – OpenNebula.

The analysis to libcloud also stands for project fog. It is a Ruby-based cloud computing library which allows collections to provide a simplified interface between them. Is doesn't bring a UI attached to view and manage.

Deltacloud is an open source project developing an API to deal with several Cloud service providers. It targets REST access and backward compatibility across versions, providing long-term stability for scripts, tools and applications [8]. Being a service-based REST API, it works like an intermediary layer, receiving commands and sending them through specially created Cloud drivers, which provide direct operation mapping with the provider's API. The requirements were simple: low pricing, testing status and support for existing OpenNebula private cloud. Based on available middleware, requirements and pricing suitable for the model needs, the Deltacloud project was chosen.

The install procedure was done by following the instructions of the project. As the OpenNebula server shared, the installation went on ruby packages instead of the RPM distribution, which is simplier. Deltacloud of two parts: deltacloudd (server) and deltacloudc (client); both were used in the same machine, but could be accessed differently.

1) *Amazon EC2 driver:* The EC2 driver is the default driver for Deltacloud core. By executing "deltacloud on the server, the Web UI is launched and one can access it on a browser "http://localhost:3001/api". After login with EC2 credentials, the use of the UI is very straightforward. To create an instance create a bucket first and then create an instance with the one of the available images and hardware profiles.

Table I is an example and refers to Amazon EC2 available hardware to attach to images. A useful column would be pricing, so that hardware could be directly comparable.

Table I. Amazon EC2 Hardware Profiles

Name	*Architecture*	*Memory*	*Storage*
t1.micro	i386	645.12	160
m1.small	i386	1740.8	160
m1.large	x86_64	7680.0	850
m1.xlarge	x86_64	15360	1690
c1.medium	i386	1740.8	350
c1.xlarge	x86_64	7168	1690
m2.xlarge	x86_64	17510.4	420
m2.2xlarge	x86_64	35020.8	850
m2.4xlarge	x86_64	70041.6	1690

The driver works well and fast enough. However, intensive work load leads to some wrong information data in instance listing. This is already reported and within development stage; improvements are expected in the next release.

2) *OpenNebula API driver:* The OpenNebula team has contributed a Deltacloud driver to manage OpenNebula instances through the OCCI API. Still the changes occurred with versions 2.0 and above have made the driver, developed for version 1.4, incompatible with the OpenNebula. Therefore, a prior version was needed for testing and the research did not went through. This occurred because of time contrains, since OpenNebula support was one of the requirements on middleware choice.
3) *Windows Azure driver*: Microsoft cloud provider Azure has also been tested with a driver connecting to its REST API. Here, as the tests were restricted to storage –as the driver still does not support computing – the perspective was to check how the process would flow. The results of the shortlist of operations over containers (creating/deleting) and blobs and (reading/write) were very reasonable.
4) *Multicloud manager*: Conductor: The Aelous project was the hard to get to work. Despite its late emerging, it was defined to embrace several projects, rather than just the Multicloud Web UI. It holds 5 sub-project is the User Interface that links them all.

After an enormous set of tests, and the correct OS Conductor could be deployed successfully. Being a set of WebUIs already tested with Deltacloud, only with some frontpages' addition, it lacks a set of desired features. Is has User and Provider Management, a set of the same operations of Deltacloud and it uses powerful libraries as Condor for scheduling and MongoDB as database.

C. Management Layer

The Sky computing management layer relies on the lower layer resources and interface. Once the middleware API is stabilized, running smoothly, this will be the next step to specifically explore. Cloud monitoring and management should be transparent, as well as application deployment. Looking deeper in the proposed architecture, illustrated in Figure 3 with some defined agents one can have a better glance of its structure.

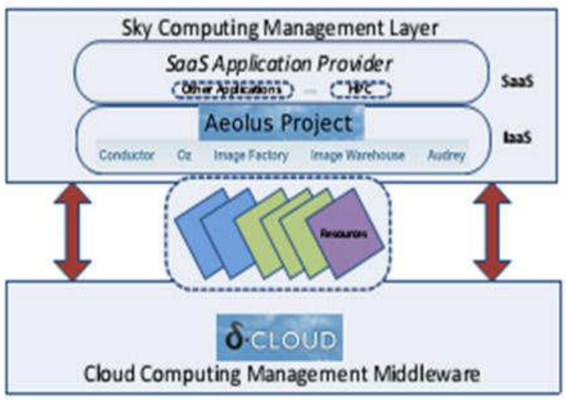


Fig.3 Proposed architecture layer details

Conductor provides a UI interface for managing the available clouds; yet a wider interface is needed in order to allow application deployment and control features like billing and accounting. Clearly this is the SaaS level that needs to be attached. However, higher abstractions than IaaS are off of the Aeolus project. At the moment, current made research has not provided a solution that fits this problem and requirements; if not available when needed it will be developed.

III. CONCLUSION

Following the proposed architecture in this paper lead to a well-structured system, with layers built from existing applications, despite its unfinished status. Still, our global idea, reusing and adapting to available system the existing, is very plain and attainable. We managed to get the first part of the architecture done, although much work is yet to be conducted.

It is possible to foresee HPC applications in the cloud more often, because of its excellent relation cost/performance. If the main problems like specialized interconnections and compatibilities are solved, it will have an assuring future.

REFERENCES

[1] K. Keahey, M. Tsugawa, A. Matsunaga and IEEE Internet Computing 13, pp. 43-51, September 2009.
[2] C. Evangelinos and C. N. Hill, "Cloud Computing for parallel Scientific HPC Applications: Feasibility of running Coupled Atmosphere Climate Models on Amazon's EC2", Proc. Cloud Computing and Its Applications, 2008.
[3] Q. He, S. Zhou, B. Kobler, D. Duffy and T. McGlynn, "Case study for running HPC applications in public clouds", in Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing (HPDC '10), ACM, New York, USA, 395.
[4] J. E. Simons and J. Buell, "Virtualizing High Performance Computing", SIGOPS Oper. Syst. Rev. 44, 4 December 2010, 136.
[5] J. Brandt, A. Gentile, J. Mayo, P. Pebay, Wong, "Resource monitoring and management with OVIS to enable HPC in cloud computing environments,"Processing, IPDPS 2009, IEEE International Symposium on.
[6] A. Monteiro, J. Sousa Pinto, C. Teixeira, T. Batista, "Cloud Interchangeability: Redefining expectations", Conference on Cloud Computing and Services Noordwijkerhout, The Netherlands, unpublished.
[7] The Apache Software Foundation, "libcloud cloud", http://incubator.apache.org/libcloud, accessed on January 2011.
[8] The Apache Software Foundation, "Deltacloud accessed on January 2011.
[9] The Apache Software Foundation, accessed on January 2011.
[10] W. Beary, "fog", github.com/geemus/fog, accessed on February 2011.
[11] IBM, Almaden Services Research Project Altocumulus, accessed on February 2011.
[12] Cloud Kick Cloud Management, www.cloudkick.com, accessed on February 2011.
[13] R. Montero, "Managing CERN: OpenNebula Team Tames the Cloud, February 2011.
[14] I. Lorente, "Deltacloud and Libcloud drivers for OpenNebula", accessed on February 2011.

A Survey Report on Role of Virtualization in Multi Cloud Service Providers

Rajesh Kumar Chakrawarti
Dept. of Computer Science Engineering
Shri Vaishnav Institute of Technology and Science
Indore, Madhya Pradesh
rajesh_kr_chakra@yahoo.com

Kajal Singhai
Dept. Computer Science Engineering Dept.
Shri Vaishnav Institute of Technology and Science
Indore, Madhya Pradesh
kajalsinghaiswt@gmail.com

Abstract—**Nowadays, not only large organization but also small and medium enterprises use computer resources for their business growth. cloud computing is a technology which provides services at the lower cost , large storage, reduce hardware and software cost valuable bandwidth, shared network, pay for the resources on consumption basis , Virtualization which is increment performance factor in business model. Virtualization is widely used in the cloud computing, it is a core technology for enabling cloud resource sharing. Virtualization is used to virtualize a single instance to the multiple instances. Because of large data store at the different locations of the cloud services it becomes so popular in between all type medical infrastructure, companies and different small and large vendors. This localization becomes possible due to virtualization concept for cloud computing.. This paper discusses the concept of virtualization in cloud environment. At every level to understand the virtualization overall concept is explained.**

Keywords—Role of Virtualization, Characteristices, Network Virtualization, Storage Virtualization, Benefits.

I. INTRODUCTION

Cloud computing is one of the most emerging and recent field of computer science engineering in which a huge amount of research is going on. Cloud computing is a flexible platform where the no. of users can get desired services as per their requirements. It is a type of distributed architecture where the user can access the services directly without any intervention to others.

According to the National Institute of Standards and Technology (NIST) Information Technology Laboratory, cloud computing is defined as follows[4]:

"Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."

Following these cloud concepts, cloud computing can be seen as a new IT infrastructure that enables users to use IT resources by renting any scale of any computing resources at anytime and anywhere from remote locations via the Internet as a service, rather than purchasing the IT resources required [3]. Cloud computing service providers build a dynamically scalable, virtualized, and integrated IT infrastructure by deployment of hardware and software resources, and then provide rapid and elastic on demand resource delivery services to customers based on their IT infrastructure over the Internet [3] in distributed architecture.

II. VIRTUALIZATION

The underpinning for the majority of high-performing clouds is a virtualized infrastructure. Virtualization has been in data centers for several years as a successful IT strategy for consolidating servers. Used more broadly to pool infrastructure resources, virtualization can also provide the basic building blocks for your cloud environment to enhance agility and flexibility. This continuing growth makes cloud computing an obvious next step for many organizations. From computing perspective, virtualization means that computing tasks or programs are run under the virtual environment rather than physical hardware, namely real environment [6]. However, in traditional approach, there are mixed hardware environment, multiple management tools, frequent application patching and updating, complex workloads and multiple software architecture. But comparatively in cloud data center far better approach like homogeneous environment, standardize management tools, minimal application patching and updating, simple workloads and single standard software architecture.[1]

Virtualization Is Not Cloud Computing Here's the difference: Virtualization abstracts compute resources typically as virtual machines (VMs) with associated storage and networking connectivity. The cloud determines how those virtualized resources are allocated, delivered, and presented. Virtualization is not necessary to create a cloud environment, but it enables rapid scaling of resources in a way that no Virtualized environments find hard to achieve [2].

The comparison table of after and before virtualization is illustrated in Table 1[5]

Table I. After and before virtualization

Characteristics	Before virtualization	After virtualization
Cost of Server Procurement	The quantity of physical server is high. High cost of hardware and electricity power consuming	Multiple servers simplify into virtual machines. Saving hardware and electricity cost.
Utilization of hardware resource	Resource utilization uneven	Improve hardware resource utilization
Deployment Time	Need to wait hardware procurement procedures.	Rapid deploying new servers in a few minute
Variability of hardware resource	Not easy	Easier
Maintenance cost	Not easy to maintenance	Not easy to maintenance Reducing legacy maintenance cost

III. WHY VIRTUALIZATION

The cloud computing virtualization environment has following characteristics are discussed below:

A. Hardware Independence

A virtual machine is configured with virtual components such as CPU, memory, network card, and SCSI controller that are completely independent of the underlying physical hardware. This gives the freedom to move a virtual machine from one physical machine to another without making any change to the device drivers, operating system, or applications.

B. Easier development

Virtualization provides easier platform for run and test application and software for heterogeneous operating environment on the same virtual machine. Isolation of different application at their respective virtual partition also helps to application developers.

C. Sharing

When a resource is too big for a single user, it is best to divide it into multiple virtual pieces, as is the case with today's Multi core processors. Each processor can run multiple virtual Machines (VMs), and each machine can be used by a different user.

D. Migration and cloning

virtual machine can be moved from one site to another to balance the workload. As the result of migration, users can access updated hardware as well as make recovery from hardware failure. Cloned virtual machines are easy to deploy in the local sites as well as remote sites.[1]

E. Consolidation

Virtualization eliminates the need of a dedicated single system to one application and multiple OS can run in the same server. Both old and advanced version of OS may capable of deploying in the same platform without purchasing additional hardware and new required applications may be run simultaneously on their on their respective OS.

F. Isolation and security

Multiple users sharing a resource may not trust each other, so it is important to provide isolation among users. Each virtual machine is isolated from each other and they are not at all interfering into the other's work which in turn helps the security aspect.

F. Aggregation

Sometimes it is possible to minimum resource and it is possible to construct a large virtual resource using virtualization behaves like large resource. In storage level, where the multiple unreliable disks can be used to make a reliable large disk.

G. Reduced Cost

Virtualization reduces the following direct costs, space for physical machines, power and cooling, Hardware including switches and Fiber Channel HBA, and annual maintenance.

IV. TYPES OF VIRTUALIZATION

After the text edit has been completed, the paper is ready for the template.

A. Compute virtualization

Compute virtualization is a technique of masking or abstracting the physical hardware from the operating system and enabling multiple operating systems to run concurrently on a single or clustered physical machine. This technique encapsulates an operating system and an application into a portable virtual machine (VM). A virtual machine is a logical entity that looks and behaves like a physical machine. Each operating system runs on its own virtual machines. In compute virtualization, a virtualization layer resides between the hardware and virtual machine. The virtualization layer is also known as hypervisor. The hypervisor provides standardized hardware resources for example: CPU, Memory, Network, etc. to all the virtual machines.

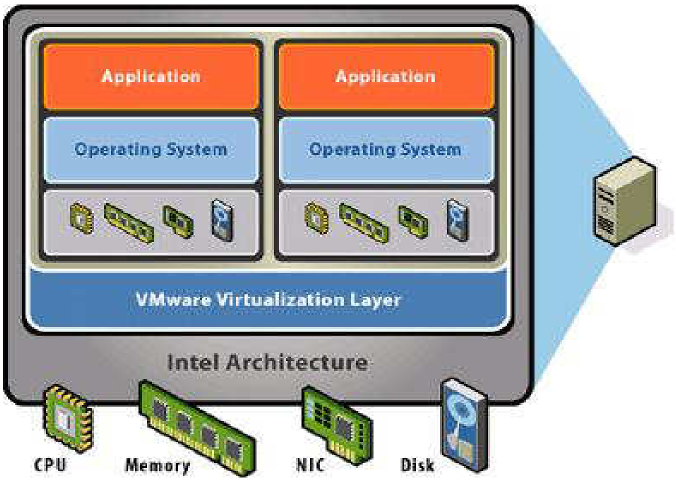


Fig. 1 Virtualization

1) Hypervisor Hypervisor is a compute virtualization software that enables multiple operating systems to run on a physical machine concurrently and to interacts directly with the physical resources. it allows multiple operating systems and applications to reside on the same physical machine[2]. And its monitoring system monitors the accesses of Virtual Machines. Hypervisor is accessible in the booting time of the system to regulate the allocation of hardware infrastructure to the multiple VMs from the resource layer.[1]

B. Storage Virtualization

Storage virtualization is the process of masking the underlying complexity of physical storage resources and presenting the logical view of these resources to compute systems in a VDC environment. Storage virtualization enables creating one or more logical storage on the physical storage resources. The logical to physical storage mapping is performed by storage virtualization layer. The virtualization layer abstracts the identity of physical storage devices and creates a storage pool by aggregating storage resources from multiple heterogeneous storage arrays.

C. Network Virtualization

Network virtualization involves logically segmenting or grouping physical network into discrete logical entities called "virtual network", and making them behave as single or multiple independent networks. Network virtualization allows multiple virtual networks to share network resources without leaking information among them. Network File System enables storing virtual machine files on remote file servers (NAS device) accessed over an IP network. The simplest example of network virtualization is virtual local area network (VLAN), which is a broadcast domain created by the switches.

1) Virtualization of Switches

A typical Ethernet switch has 32–128 ports. The number of physical machines that need to be connected on an L2 network is typically much larger than this. Therefore, several layers of switches need to be used to form an L2 network.

2) *Virtual LANs in Clouds*

The VLAN technology can divide a large LAN into smaller virtual LANs or combine separate LANs into one or more virtual LANs One additional problem in the cloud environment is that multiple VMs in a single physical machine may belong to different clients and thus need to be in different virtual LANs (VLANs). Each of these VLANs may span several data centers interconnected. Again, there are a number of competing proposals to solve this problem. VMware and several partner companies have proposed virtual extensible LANs (VXLANs).

D. Desktop Virtualization

The objective of desktop virtualization technology is to centralize the PC Operating System (OS) at the data center. Desktops hosted at the data center run as Virtual Machines (VMs) within the VDC, while end users remotely access these desktops from a variety of endpoint devices. This is to make security and management of desktops significantly easy.

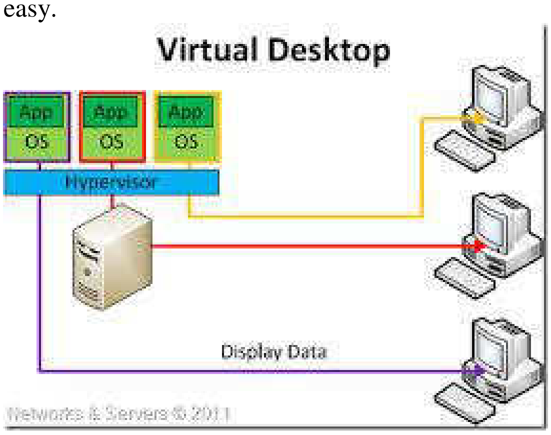


Fig 2. Virtual desktop

E. Application Virtualization

Application virtualization is technique an application to user without any installation, integration, or dependencies on the computing platform. This technology provides the ability to deploy applications without modifying or making any change to the underlying OS, file system, or registry of the computing platform in which they are deployed. Because virtualized applications run in an isolated environment, the underlying OS, the underlying OS and other applications are protected from potential corruptions which may be caused due to installation modifications.

F. *Management Virtualization*

In this technique, the physical and virtual datacenter are managed to give one single unified infrastructure for the provision of services. Resource Pools are the best example of Management virtualization.

V. Benifits of Virtualization

1) Encapsulation: A virtual machine is a package that contains a complete set of virtual hardware resources, an operating system, and applications [9]. Encapsulation makes virtual machines portable and easy to manage. For example, a virtual machine can be moved and copied from one location to another just like a file.

2) *Enhanced Security:* Network virtualization provides enhanced security by restricting access to nodes located within a virtual network from another virtual network [9]. Therefore, sensitive data of one virtual network is isolated from other virtual networks.

3) *Increase Performance:* Network broadcasts within a virtual network are not allowed to propagate to other virtual networks. Restricting broadcast preserves network bandwidth, which consequently improves virtual network performance for usual network traffic [9].

4) Simplified backup: Since centralized virtual desktops reside entirely within an organization's datacenter, it is easier to ensure full compliance with backup policies.

5) To be effective, aggregating drive space and processing power requires virtualization capabilities that present multiple components as a single logical pool [8].

6) The ability to continuously add capacity or processing power to any storage solution and then divide resources as needed.

7) In the storage market, particularly with large, virtualized storage architectures. It is not a matter of more capacity or performance, but how effectively it can be used [8].

8) Virtualization allows more users to interact with more applications that access a larger and more flexible pool of centrally managed storage.

9) Application delivery is more agile and rapid because applications are installed once on the server and can be accessed by the user without having to install the application locally.

VI. Conclusion

This paper totally discussed about the role of virtualization in cloud computing services. I also discussed the Concept of Virtualization in Cloud Computing as any discussion of cloud computing typically begins with the virtualization. Virtualization is using computer resources to imitate other computer resources or whole computers. The type of virtualization storage, networking and desktop virtualization provides greater benefits for the business and research field. Application virtualization provides the multiple app's running together due to abstraction of resources. I discussed the characteristics, benefits, and various forms of Virtualization.

References

[1] A. Souvik Pal1, B. Prasant Kumar Pattnaik "Classification of Virtualization Environment for Cloud Computing"Indian Journal of Science and Technology Vol: 6 Issue: 1 January 2013

[2] White paper on"Virtualization and Cloud Computing Intel IT center August 2013

[3] Hyeong-Ok Lee1 and Mihye Kim2 "Implementing Cloud Computing in the Current IT Environments of Korean Government Agencies" International Journal of Software Engineering and Its Applications Vol. 7, No. 1,January 2013,pp:149

[4] "NIST Cloud Computing efinition", NIST SP 800- 145.

[5] Ajit Kumar Shrivastava, Puneet Himthani "Multi Tier Authentication for Cloud Security" presented in national conference on "emerging trends in communication, control and networking, MITS Gwalior.

[6] Zongjian He , Guanqing Liang "Research and Evaluation of Network Virtualization in Cloud Computing Environment"School of Software engineering ,Tongji University,Sanghai,China

[7] Raj Jain and Subharthi Paul,Network Virtualization and Software Defined Networking for Cloud Computing: A Survey Washington University cloud networking and communications.

[8] Gary Johnson, SAN architect with Carlson Companies, from the book Storage Virtualization: Technologies for Simplifying Data Storage and Management, by Tom Clark and courtesy of Addison-Wesley.

[9] A Secure Cloud Computing Model Based on Multi Cloud Service Providers , Mooga Masthan , Dora Babu Sudarsa Volume 3, Issue 5, May 2013

[10] Vmware " www.vmware.com"

[11] Google App Engine, URL http://code.google.com/appengine.

Working of Cloud Architecture and Components

Sunil Kumar
Department of Computer Science
S.V. Subharti University
Meerut, Uttar Pardesh, INDIA
sunilcloudresearch@gmail

Varun Sharma
Department of Computer Science
CT group of Institutions
Jalandhar, Punjab, India

***Abstract*—This Cloud computing has recently emerged as a new paradigm for hosting and delivering services over the Inter- net. Cloud computing is attractive to business owners as it eliminates the requirement for users to plan ahead for provisioning, and allows enterprises to start from the small and increase resources only when there is a rise in service demand. However, despite the fact that cloud computing offers huge opportunities to the IT industry, the development of cloud computing technology is currently at its infancy, with many issues still to be addressed. In this paper, we present a survey of cloud computing, highlighting its key concepts, architectural principles, state-of-the-art implementation as well as research challenges.**

Keywords—component; formatting; style; styling; insert (key words)

I. INTRODUCTION

With the rapid development of processing and storage technologies and the success of the Internet, computing resources have become cheaper, more powerful and more ubiquitously available than ever before.[1] This technological trend has enabled the realization of a new computing mode called cloud computing, in which resources (e.g., CPU and storage) are provided as general utilities that can be leased and released by users through the Internet in an on-demand fashion. In a cloud computing environment, the traditional role of service provider is divided into two: the infrastructure providers who manage cloud platforms and lease resources according to a usage-based pricing model, and service providers , who rent resources from one or many infrastructure providers to serve the end user The emergence of cloud computing has made a tremendous impact on the Information Technology (IT) industry over the past few years, where large companies such as Google, Amazon and Microsoft strive to provide more powerful, reliable and cost-efficient cloud platforms, and business enterprises seek to reshape their business models to gain benefit from this new paradigm. Indeed, cloud computing provides several compelling features that make it attractive to business owners, as shown below-

A. No up-front investment

Cloud computing[2] uses a pay-as- you-go pricing model. A service provider does not need to invest in the infrastructure to start gaining benefit from cloud computing. It simply rents resources from the cloud according to its own needs and pay for the usage

B. Lowering operating cost

Resources in a cloud environment can be rapidly allocated and de-allocated on demand. Hence, a service provider no longer needs to provision capacities according to the peak load. This provides huge savings since resources can be released to save on operating costs when service demand is low

C. Highly scalable

Infrastructure providers pool large amount of resources from data centers and make them easily Accessible . A service provider can easily expand its service to large scales in order to handle rapid increase in service demands (e.g., flash-crowd effect). This model is sometimes called surge computing.

D. Easy access

Services[3] hosted in the cloud are generally web-based. Therefore, they are easily accessible through a variety of devices with Internet connections. These devices not only include desktop and laptop computers, but also cell phones and PDAs.

E). Reducing business risks and maintenance expenses:

[4]By outsourcing the service infrastructure to the clouds, a service provider shifts its business risks (such as hardware failures) to infrastructure providers, who often have better expertise and are better equipped for managing these risks. In addition, a service provider can cut down the hardware maintenance and the staff training costs. However, although cloud computing has shown considerable opportunities to the IT industry, it also brings many unique challenges that need to be carefully addressed. In this paper, we present a survey of cloud computing, highlighting its key concepts, architectural principles, state-of-the-art implementations as well as research challenges. Our aim is to provide a better understanding of the design challenges of cloud computing and identify important research directions in this fascinating topic

II. LITERATURE BACKGROUND

This section presents a general overview of cloud computing, including its definition and a comparison with related concepts. [5]The main idea behind cloud computing is not a new one. John McCarthy in the 1960s already envisioned that computing facilities will be provided to the general public like a utility. The term “cloud” has also been used in various contexts such as describing large ATM networks in the1990s. However, it was after Google's CEO Eric Schmidt used the word to describe the business model of providing services across the Internet in 2006, that the term re- ally started to gain popularity. Since then, the term cloud computing has been used mainly as a marketing term in a variety of contexts to represent many different ideas. Certainly, the lack of a standard definition of cloud computing has generated not only market hypes, but also a fair amount of skepticism and confusion. For this reason, recently there has been work of standardizing the definition of cloud computing. As an example, the work in compared over 20 different definitions from a variety of sources to confirm a standard definition. In this paper, we adopt the definition of cloud computing provided by The National Institute of Standards and Technology (NIST) , as it covers, in our opinion, all the essential aspects of cloud computing: NIST definition of cloud computing Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources(e.g., networks ,servers ,storage, applications ,and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

The main reason[6] for the existence of different perceptions of cloud computing is that cloud computing, unlike other technical terms, is not a new technology, but rather a new operations model that brings together a set of existing technologies to run business in a different way. Indeed, most of the technologies used by cloud computing such as virtualization and utility-based pricing , are not new. Instead, cloud computing leverages these existing technologies to meet the technological and economic requirements of today's demand for information technology.2.2 Related technologies Cloud computing is often compared to the following technologies, each of which shares certain aspects with cloud computing:

A. Grid Computing

Grid computing[7] is a distributed computing paradigm that coordinates networked resources to achieve a common computational objective. The development of Grid computing was originally driven by scientific applications which are usually computation-intensive. Cloud computing is similar to Grid computing in that it also employs distributed resources to achieve application-level objectives. However, cloud computing takes one step further by leveraging virtualization technologies at multiple levels (hardware and application platform) to realize resource sharing and dynamic resource provisioning

B. Utility Computing

Utility computing[8] represents the model of providing resources on-demand and charging customers based on usage rather than a flat rate. Cloud computing can be perceived as a realization of utility computing. It adopts a utility-based pricing scheme entirely for economic reasons. With on-demand resource provisioning and utility-based pricing, service providers can truly maximize resource utilization and minimize their operating costs.

C. Virtualization

Virtualization is a technology that abstracts away the details of physical hardware and provides virtualized resources for high-level applications.[9] A virtualized server is commonly called a virtual machine (VM). Virtualization forms the foundation of cloud computing, as it provides the capability of pooling computing resources from clusters of servers and dynamically assigning or reassigning virtual resources to applications on-demand. Autonomic

D. Computing

It is originally coined by IBM in2001[10], autonomic computing aims at building computing systems capable of self-management, i.e. reacting to internal and external observations without human intervention. The goal of autonomic computing is to overcome the management complexity of today's computer systems. Although cloud computing exhibits certain autonomic features such as automatic resource provisioning, its objective is to lower the resource cost rather than to reduce system complexity. In summary, cloud computing leverages virtualization technology to achieve the goal of providing computing re- sources as a utility. It shares certain aspects with grid com- putting and autonomic computing but differs from them in other aspects. Therefore, it offers unique benefits and imposes distinctive challenges to meet its requirements.

A layered model of cloud computing Generally speaking, the architecture of a cloud computing environment can be divided into 4 layers[11]: the hard-ware/datacenter layer, the infrastructure layer, the platform layer and the application layer, as shown in Fig.1.

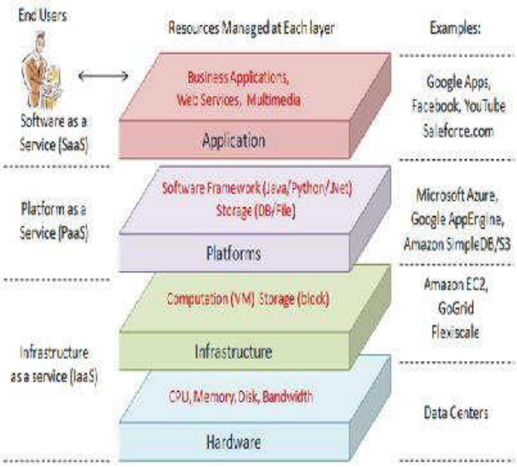


Fig. 1 OSI model

1. The hardware layer

This layer[12] is responsible for man-aging the physical resources of the cloud, including physical servers, routers, switches, power and cooling systems .In practice, the hardware layer is typically implemented in data centers. A data center usually contains thousands of servers that are organized in racks and interconnected through switches, routers or other fabrics. Typical issues at hardware layer include hardware configuration, fault-tolerance, traffic management, power and cooling resource management.

2. The infrastructure layer

Also known as the virtualization layer,[13] the infrastructure layer creates a pool of storage and computing resources by partitioning the physical re- sources using virtualization technologies such as Xen [35],KVM [30] and VMware [32]. The infrastructure layer is an essential component of cloud computing, since many key features, such as dynamic resource assignment, are only made available through virtualization technologies

3. The platform layer

Built on top of the infrastructure layer[14], the platform layer consists of operating systems and application frameworks. The purpose of the platform layer is to minimize the burden of deploying applications directly into VM containers. For example, Google App Engine operates at the platform layer to provide API support for implementing storage, database and business logic of typical web applications.

4. The application layer

At the highest level of the hierarchy[15], the application layer consists of the actual cloud applications. Different from traditional applications, cloud applications can leverage the automatic-scaling feature to achieve better performance, availability and lower operating cost. Compared to traditional service hosting environments such as dedicated server farms, the architecture of cloud computing is more modular. Each layer is loosely coupled with the layers above and below, allowing each layer to evolve separately. Cloud computing characteristics Cloud computing provides several salient features that are different from traditional service computing, which we summarize below:

a). Multi-tenancy

In a cloud environment[16], services owned by multiple providers are co-located in a single data center. The performance and management issues of these services are shared among service providers and the infrastructure provider. The layered architecture of cloud computing provides a natural division of responsibilities: the owner of each layer only needs to focus on the specific objectives associated with this layer. However, multi-tenancy also introduces difficulties in understanding and managing the interactions among various stakeholders

b). Shared resource pooling

The infrastructure provider[17] offers a pool of computing resources that can be dynamically assigned to multiple resource consumers. Such dynamic re-source assignment capability provides much flexibility to infrastructure providers for managing their own resource usage and operating costs. For instance, an IaaS provider can leverage VM migration technology to attain a high degree of server consolidation, hence maximizing resource utilization while minimizing cost such as power consumption and cooling. Geo-distribution and ubiquitous network access: Clouds are generally accessible through the Internet and use the Internet as a service delivery network. Hence any device with Internet connectivity, be it a mobile phone, a PDA or a laptop, is able to access cloud services. Additionally, to achieve high network performance and localization, many of today's clouds consist of data centers located at many locations around the globe. A service provider can easily leverage geo-diversity to achieve maximum service utility.

c). Service oriented

As mentioned previously[18], cloud com- puting adopts a service-driven operating model. Hence it places a strong emphasis on service management. In a cloud, each IaaS, PaaS and SaaS provider offers its service according to the Service Level Agreement (SLA) negotiated with its customers. SLA assurance is therefore a critical objective of every provider. Dynamic resource provisioning:One of the key features of cloud computing is that computing resources can be obtained and released on the fly.

III. RESEARCH CHALLENGES

Although cloud computing has been widely adopted by the industry, the research on cloud computing is still at an early stage. Many existing issues have not been fully addressed, while new challenges keep emerging from industry applications. In this section, we summarize some of the challenging research issues in cloud computing[19].

A). Automated service provisioning

One of the key features of cloud computing is the capability of acquiring and releasing resources on-demand. The objective of a service provider in this case is to allocate and de-allocate resources from the cloud to satisfy its service level objectives (SLOs), while minimizing its operational cost. However, it is not obvious how a service provider can achieve this objective. In particular, it is not easy to determine how to map SLOs such as QoS requirements to low-level resource requirement such as [20]CPU and memory requirements. Furthermore, to achieve high agility and respond to rapid demand fluctuations such as in flash crowd effect, the resource provisioning decisions must be made on line. Automated service provisioning is not a new problem. Dynamic resource provisioning for Internet applications has been studied extensively in the past. These approaches typically involve: (1) Constructing an application performance model that predicts the number of application instances required to handle demand at each particular level.

B). Virtual machine migration

Virtualization can provide significant benefits in cloud computing by enabling virtual machine migration to balance load across the data center. In addition, virtual machine migration enables robust and highly responsive provisioning in data centers. Virtual machine migration has evolved from process migration techniques [21]. More recently, Xen [22] and VMWare have implemented "live" migration of VMsthat involves extremely short downtimes ranging from tens of milliseconds to a second. Clark et al. pointed out that migrating an entire OS and all of its applications as one unit allows to avoid many of the difficulties faced by process-level migration approaches, and analyzed the benefits of live migration of VMs .

C). Server consolidation

Server consolidation[23] is an effective approach to maximize resource utilization while minimizing energy consumption in a cloud computing environment. Live VM migration technology is often used to consolidate VMs residing on multiple under-utilized servers onto a single server, so that the remaining servers can be set to an energy-saving state. The problem of optimally consolidating servers in a data center is often formulated as a variant of the vector bin-packing problem, which is an NP-hard optimization problem. Various heuristics have been proposed for this problem. Additionally dependencies among VMs, such as communication requirements, have also been considered recently. However, server consolidation activities should not hurt application performance. It is known that the resource usage(also known as the footprint of individual VMs may vary over time. For server resources that are shared among VMs, such as bandwidth, memory cache and disk I/O, maximally consolidating a server may result in re-source congestion when a VM changes its footprint on the server. Hence, it is sometimes important to observe the fluctuations of VM footprints and use this information for effective server consolidation. Finally, the system must quickly react to resource congestions when they occur.

IV. CONCLUSION

Cloud computing has recently emerged as a compelling paradigm for managing and delivering services over the Internet. The rise of cloud computing is rapidly changing the landscape of information technology, and ultimately turning the long-held promise of utility computing into a reality. However, despite the significant benefits offered by cloud computing, the current technologies are not matured enough to realize its full potential. Many key challenges in this domain, including automatic resource provisioning, power management and security management, are only starting to receive attention from the research community. There-fore, we believe there is still tremendous opportunity for re-searchers to make groundbreaking contributions in this field, and bring significant impact to their development in the industry. In this paper, we have surveyed the state-of-the-art of cloud computing, covering its essential concepts, architectural designs, prominent characteristics, key technologies as well as research directions. As the development of cloud computing technology is still at an early stage, we hope our work will provide a better understanding of the design challenges of cloud computing, and pave the way for further re-search in this area.

REFERENCES

[1] Al-Fares M et al (2008) A scalable, commodity data center net-work architecture. In: Proc SIGCOMM

[2] Amazon Elastic Computing Cloud,aws.amazon.com/ec2

[3] Amazon Web Services,aws.amazon.com

[4] Ananthanarayanan R, Gupta K et al (2009) Cloud analytics: do we really need to reinvent the storage stack? In: Proc of HotCloud

[5] Armbrust M et al (2009) Above the clouds: a Berkeley view of cloud computing. UC Berkeley Technical Report

[6] Berners-Lee T, Fielding R, Masinter L (2005) RFC 3986: uniform resource identifier (URI): generic syntax, January 2005

[7] Bodik P et al (2009) Statistical machine learning makes automatic control practical for Internet datacenters. In: Proc HotCloud

[8] Brooks D et al (2000) Power-aware microarchitecture: design and modeling challenges for the next-generation microprocessors,IEEE Micro.

[9] Chandra A et al (2009) Nebulas: using distributed voluntary re-sources to build clouds. In: Proc of HotCloud

[10] Chang F, Dean J et al (2006) Bigtable: a distributed storage system for structured data. In: Proc of OSDI

[11] Chekuri C, Khanna S (2004) On multi-dimensional packing prob-lems. SIAM J Comput 33(4):837–851

[12] Church K et al (2008) On delivering embarrassingly distributed cloud services. In: Proc of HotNets

[13] Clark C, Fraser K, Hand S, Hansen JG, Jul E, Limpach C, Pratt I,Warfield A (2005) Live migration of virtual machines. In: Proc ofNSDI

[14] Cloud Computing on Wikipedia,en.wikipedia.org/wiki/Cloudcomputing 20 Dec 2009

[15] Cloud Hosting, CLoud Computing and Hybrid Infrastructure from GoGrid, http://www.gogrid.com

[16] Dean J, Ghemawat S (2004) MapReduce: simplified data processing on large clusters. In: Proc of OSDI

[17] Dedicated Server, Managed Hosting, Web Hosting by Rackspace Hosting,http://www.rackspace.com

[18] FlexiScale Cloud Comp and Hosting,www.flexiscale.com

[19] Ghemawat S, Gobioff H, Leung S-T (2003) The Google file system. In: Proc of SOSP, October 2003

[20] Google App Engine, URL http://code.google.com/appengine

[21] Greenberg A, Jain N et al (2009) VL2: a scalable and flexible datacenter network. In: Proc SIGCOMM

[22] Guo C et al (2008) DCell: a scalable and fault-tolerant network structure for data centers. In: Proc SIGCOMM

[23] Guo C, Lu G, Li D et al (2009) BCube: a high performance,server-centric network architecture for modular data centers. In:Proc SIGCOMM

[24] Hadoop Distributed File System,hadoop.apache.org/hd

Current Security Problems and Solutions of Cloud Storage: An Analysis

Sunil Kumar
Dept. of CSE
Swami Vivekanand Subharti University, Meerut
sunilcloudresearch@gmail.com

Jayant Shekhar
Dept. of CSE
Subharti Institute of Technology and Engineering
jayant_shekhar@hotmail.com

Jaspreet Kaur
Dept. of CSE
CTIT, Jalandhar

Abstract-In the last few years, we have seen that cloud computing has developed as a promising business model to one of the fastest growing IT sector. Now, recession-hit companies are gradually more realizing that simply by tapping into the cloud they can gain fast access to best-of-breed business applications or drastically boost their infrastructure resources, all at negligible cost. But as more and more information on individuals and companies is placed in the cloud, concerns are beginning to grow about just how safe an environment it is. This paper discusses various data security issues and challenges that cloud service providers are facing during cloud implementation for their clients.

Keywords: Data Security issues; Cloud computing; Data security solutions etc.

I. INTRODUCTION

Recently, cloud computing has emerged as the leading technology for delivering reliable, secure, fault-tolerant, and scalable computational service. Cloud service may be offered in private data centers (private clouds), may be commercially offered for clients (public clouds), or yet it is possible that both public and private clouds are combined in hybrid clouds. There are many definitions available in various books or internet, but most accepted definition is "Cloud computing is a model for enabling convenient, on-demand network access, to a shared pool of configurable computing resources, (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [2].

The name cloud computing that is inspired by the cloud symbol that is generally used to represent the Internet in flow charts or diagrams. A distinct migration to the clouds has been taking place over recent years with end users, ''bit by bit'' maintaining a growing number of personal data, including bookmarks, photographs, music files and much more, on remote servers accessible via a network. Cloud computing is empowered by virtualization technology; a technology that actually dates back to 1967, but for decades was available only on mainframe systems. In its quintessence, a host computer runs an application known as a hypervisor; this creates one or more virtual machines, which simulate physical computers so faithfully, that the simulations can run any software, from operating systems, to end-user applications [3].

There are five essential characteristics of cloud computing which explains there relation and difference from the traditional computing.

- On-demand-self-service: Consumer can provision or un-provision the services when needed, without the human interaction with the service provider.
- Broad network Access: It has capabilities over the network and accessed though standard mechanism.
- Resource Pooling: The computing resources of the supplier are pooled to serve multiple consumers which are using a multi-tenant model, with various physical and virtual resources dynamically assigned, depending on consumer demand.
- Rapid Elasticity: Service can be rapidly and elastically provisioned.
- Measured Service: Cloud computing system automatically control and optimize resource usage by metering capability to the services (e.g.storage, processing, bandwidth, active uses accounts) [8].

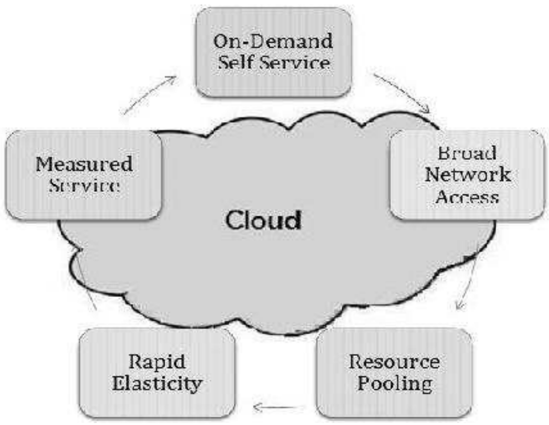


Fig.1. Cloud computing Essential Characteristics

A. *Service Model*

In addition to these five essential characteristics, the cloud community has extensively used the following three service models to categories the cloud services:

i. *Software as a Service* (SaaS). Cloud consumers release their applications on a hosting environment, which can be accessed through networks from various clients (e.g.

web browser, PDA, etc.) by application users. Cloud consumers do not have control over the Cloud infrastructure that often employs a multi-tenancy system architecture, namely, different cloud consumers' applications are organized in a single logical environment on the SaaS cloud to achieve economies of scale and optimization in terms of speed, security, availability, disaster recovery, and maintenance.

Examples of SaaS include SalesForce.com, Google Mail, Google Docs, and so forth.

ii. *Platform as a Service* (PaaS). PaaS is a development platform supporting the full "Software Lifecycle" which allows cloud consumers to develop cloud services and applications (e.g. SaaS) directly on the PaaS cloud. Hence the difference between SaaS and PaaS is that SaaS only hosts completed cloud applications whereas PaaS offers a development platform that hosts both completed and in-progress cloud applications. This requires PaaS, in addition to supporting application hosting environment, to possess development infrastructure including programming environment, tools, configuration management, and so forth. An example of PaaS is Google AppEngine.

iii. *Infrastructure as a Service* (IaaS). Cloud consumers directly use IT infrastructures (processing, storage, networks, and other fundamental computing resources) provided in the IaaS cloud. Virtualization is extensively used in IaaS cloud in order to integrate/decompose physical resources in an ad-hoc manner to meet growing or shrinking resource demand from cloud consumers. The basic strategy of virtualization is to set up independent virtual machines (VM) that are isolated from both the underlying hardware and other VMs. Notice that this strategy is different from the multi-tenancy model, which aims to transform the application software architecture so that multiple instances (from multiple cloud consumers) can run on a single application (i.e. the same logic machine). An example of IaaS is Amazon's EC2.

iv. *Data storage as a Service* (DaaS). The delivery of virtualized storage on demand becomes a separate Cloud service - data storage service. Notice that DaaS could be seen as a special type IaaS. The motivation is that on-premise enterprise database systems are often tied in a prohibitive upfront cost in dedicated server, software license, post-delivery services, and in-house IT maintenance. DaaS allows consumers to pay for what they are actually using rather than the site license for the entire database. In addition to traditional storage interfaces such as RDBMS and file systems, some DaaS offerings provide table-style abstractions that are designed to scale out to store and retrieve a huge amount of data within a very compressed timeframe, often too large, too expensive or too slow for most commercial RDBMS to cope with. Example of this kind of DaaS include Amazon S3, Google BigTable, and Apache HBase, etc.

B. Deployment Model

More recently, four cloud deployment models have been defined in the Cloud community:

i. *Private cloud*. The cloud infrastructure is operated solely within a single organization, and managed by the organization or a third party regardless whether it is located premise or off premise. The motivation to setup a private cloud within an organization has several aspects. First, to maximize and optimize the utilization of existing in-house resources. Second, security concerns including data privacy and trust also make Private Cloud an option for many firms. Third, data transfer cost from local IT infrastructure to a Public Cloud is still rather considerable. Fourth, organizations always require full control over mission-critical activities that reside behind their firewalls. Last, academics often build private cloud for research and teaching purposes.

ii. *Community cloud*. Several organizations jointly construct and share the same cloud infrastructure as well as policies, requirements, values, and concerns. The cloud community forms into a degree of economic scalability and democratic equilibrium. The cloud infrastructure could be hosted by a third-party vendor or within one of the organizations in the community.

iii. *Public cloud*. This is the dominant form of current Cloud computing deployment model. The public cloud is used by the general public cloud consumers and the cloud service provider has the full ownership of the public cloud with its own policy, value, and profit, costing, and charging model. Many popular cloud services are public clouds including Amazon EC2, S3,Google AppEngine, and Force.com.

iv. *Hybrid cloud*. The cloud infrastructure is a combination of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load-balancing between clouds). Organizations use the hybrid cloud model in order to optimize their resources to increase their core competencies by margining out peripheral business functions onto the cloud while controlling core activities on-premise through private cloud.

II. SECURITY ISSUES IN THE CLOUD

There are many security issues associated with cloud computing and they can be grouped into any number of proportions. According to Gartner [4], before making a choice of cloud vendors, users should ask the vendors for seven specific safety issues: Privileged user access, regulatory compliance, data location, data segregation, recovery, investigative support and long-term viability. In 2009, Forrester Research Inc. [5] evaluated security and privacy practices of some of the leading cloud providers (such as Salesforce.com, Amazon, Google, and Microsoft) in three major aspects: Security and privacy, compliance, and legal and contractual issues. Cloud Security Alliance (CSA) [6] is gathering solution providers, non-profits and individuals to enter into discussion about the current and future best practices for information assurance in the cloud.

The CSA has identified thirteen domains of concerns on cloud computing security [7]. S. Subashini and V. Kavitha made an investigation of cloud computing security issues from the cloud computing service delivery models (SPI model) and give a detailed analysis and assessment method description for each security issue [8]. Mohamed Al Morsy, John Grundy and Ingo Müller explored the cloud computing security issues from different perspectives, including security issues associated with cloud computing architecture, service delivery models, cloud characteristics and cloud stakeholders [9]. Yanpei Chen, Vern Paxson and Randy H. Katz believed that two aspects are to some degree new and essential to cloud: the complexities of multi-party trust considerations, and the ensuing need for mutual auditability. They also point out some new opportunities in cloud computing security [10]. According to the SPI service delivery models, deployment models and essential characteristics of cloud, there are security issues in all aspects of the infrastructure including network level, host level and application level.

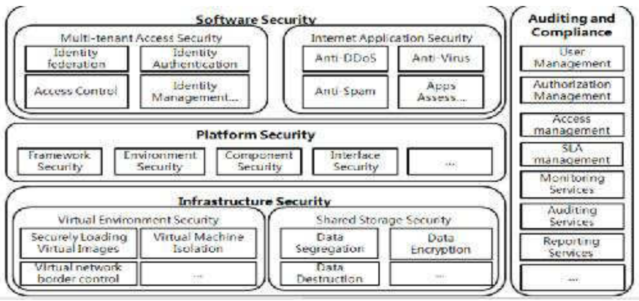


Fig. 2. Cloud computing security architecture

A. Browser Security

In a Cloud, computation is done on remote servers. The client PC is used for I/O only, and for authentication and authorization of commands to the Cloud. It thus does not make sense to develop (platform dependent) client software, but to use a universal, platform independent tool for I/O: a standard Web browser. This trend has been observed during the last years, and has been categorized under different names: Web applications, Web 2.0, or Software-as-a-Service (SaaS). Modern Web browsers with their AJAX techniques (JavaScript, XMLHttpRequest, Plugins) are ideally suited for I/O. But what about security? A partial answer is given in [8], where different browser security policies (with the notable exception of TLS) are compared for the most important browser releases. With a focus on the Same Origin Policy (SOP), this document reveals many shortcomings of browser security. If we additionally take into account TLS, which is used for host authentication and data encryption, these shortcomings become even more obvious.

Web browsers can not directly make use of XML Signature or XML Encryption: data can only be encrypted through TLS, and signatures are only used within the TLS handshake. For all other cryptographic data sets within WS-Security, the browser only serves as a passive data store.

B. Data Security

Data Security refers to the protection process for the stored data from unwanted access or modifications by unauthorised users. Confidentiality, integrity and availability (CIA) are the main security properties. Confidentiality means to make sure that only authorised clients with the appropriate rights and privileges can access the stored data. Integrity means to protect the stored data from being inappropriately modified (whether accidentally or deliberately). Availability refers to assurance that the stored data is always available to be delivered to the users.

The data in the cloud may be divided into: (1) The data in IaaS environment, such as Amazon's Simple Storage Service; (2) The data in PaaS or SaaS environment related to cloud based applications. The data stored in the cloud storages is similar with the ones stored in other places and needs to consider three aspects of information security: confidentiality, integrity and availability. The common solution for data confidentiality is data encryption. In order to ensure the effective of encryption, there needs to consider the use of both encryption algorithm and key strength. As the cloud computing environment involving large amounts of data transmission, storage and handling, there also needs to consider processing speed and computational efficiency of encrypting large amounts of data. In this case, for example, symmetric encryption algorithm is more suitable than asymmetric encryption algorithm.

Another key problem about data encryption is key management. Is who responsible for key management? Ideally, it’s the data owners. But at present, because the users have not enough expertise to manage the keys, they usually entrust the key management to the cloud providers. As the cloud providers need to maintain keys for a large number of users, key management will become more complex and difficult[22].

IBM developed a fully homomorphic encryption scheme in June 2009. This scheme allows data to be processed without being decrypted [12]. Roy I and Ramadan HE applied decentralized information flow control (DIFC) and differential privacy protection technology into data generation and calculation stages in cloud and put forth a privacy protection system called Airavat [13]. This system can prevent privacy leakage without authorization in Map-Reduce computing process. A key problem for data encryption solutions is key management. On the one hand, the users have not enough expertise to manage their keys. On the other hand, the cloud service providers need to maintain a large number of user keys. The Organization for the Advancement of Structured Information Standards (OASIS) Key Management Interoperability Protocol (KMIP) is trying to solve such issues [14]. About data integrity verification, because of data communication, transfer fees and time cost, the users cannot first download data to verify its correctness and then upload the data. And as the data is dynamic in cloud storage, traditional data integrity solutions are no longer suitable. NEC Labs's provable data integrity (PDI) solution can support public data integrity verification [15]. Cong Wang proposed a mathematical way to verify the integrity of the data dynamically stored in the cloud [16].

In the data storage and use stages, Mowbray proposed a client-based privacy management tool [17]. It provides a user centric trust model to help users to control the storage and use of their sensitive information in the cloud. Munts-Mulero discussed the problems that existing privacy protection technologies (such as K anonymous, Graph Anonymization, and data pre-processing methods) faced when applied to large data and analyzed current solutions [18]. The challenge of data privacy is sharing data while protecting personal privacy information. Randike Gajanayake proposed a privacy protection framework based on information accountability (IA) components [19]. The IA agent can identify the users who are

accessing information and the types of information they use. When inappropriate misuse is detected, the agent defines a set of methods to hold the users accountable for misuse.

About data destruction, U.S. Department of Defense (DoD) 5220.22-M (the National Industrial Security Program Operating Manual) shows two approved methods of data (destruction) security, but it does not provide any specific requirements for how these two methods are to be achieved [20]. The National Institute of Standards and Technology (NIST) Special Publication [21], 800-88, gives a"*Guidelines for Media Sanitization.*"

Kamara et al.'s work [22] can be seen as the first contribution to cryptography-based cloud storage. In the environment of public cloud infrastructure where the service provider is not completely trusted by customers, Kamara et al. designed secure cloud storage architectures for both consumer and enterprise scenarios by using non-standard cryptographic techniques, such as attributed encryption, searchable encryption, etc.

To assure the privacy of personal health information (PHI) stored in cloud storage to which the patient self-controlled access privilege, Barua et al. [23] propose access control scheme for cloud storage, that is, efficient and secure patient-centric access control (ESPAC) scheme. This scheme allows data requesters to have different access privileges based on their roles, and then assigns different attributes sets to them.

Popa et al. [24] presented a secure cloud storage system named CloudProof which can allow customers not only to detect violations of integrity, write-serializability, and freshness, and also to prove the occurrence of these violations to a third party. This cloud storage can be implemented based on conventional cloud storage services like Amazon S3 or Azure Blob Storage.

Somorovsky et al. [25] proposed a secure solution Sec2 for distributed public cloud storage by using extensive makeup language (XML) encryption. XML encryption is used in XML encryption engine which is responsible for encryption and decryption of XML payload and key data.

III. CONCLUSION

So, it is required to take security and privacy into consideration while designing and using cloud services. In this paper, various data security issues of cloud computing is covered that give challenges to the cloud service providers and try to cover various solutions of some issues.

IV. FUTURE WORK

After analysing the services of cloud computing it is clear that everyone wants to use the cloud due to cost savings and new agile business models. But, not transforming themselves fully due to existence of data security. So, to increase the faith of IT industry as well as users on cloud computing, a more secured solution required.

V. REFERENCES

[1] K. Stanoevska-Slabeva, T. Wozniak, Grid and Cloud Computing-A Business Perspective on Technology and Applications, Springer-Verlag, Berlin, Heidelberg, 2010.

[2] National Institute of Standards and Technology, The NIST Definition of Cloud Computing, Information Technology Laboratory, 2009.

[3] E. Naone, Technology overview, conjuring clouds, MIT Technology Review, July–August, 2009.

[4] Gartner: Seven cloud-computing security risks. InfoWorld. 2008-07-02. http://www.infoworld.com/d/security-central/gartner-seven-cloudcomputing-security-risks-853.

[5] Cloud Security Front and Center. Forrester Research. 2009-11-18. http://blogs.forrester.com/srm/2009/11/cloud-security-front-andcenter.html

[6] Cloud Security Alliance. http://www.cloudsecurityalliance.org.

[7] Cloud Security Alliance, Security Guidance for Critical Areas of Focus in Cloud Computing, V2.1, http://www.cloudsecurityalliance.org/ guidance/csaguide.v2.1.pdf.

[8] S. Subashini, V.Kavitha. A survey on security issues in service delivery models of cloud computing. Journal of Network and Computer Applications 34(2011)1-11.

[9] Mohamed Al Morsy, John Grundy, Ingo Müller, "An Analysis of The Cloud Computing Security Problem," in Proceedings of APSEC 2010 Cloud Workshop, Sydney, Australia, 30th Nov 2010.

[10] Yanpei Chen, Vern Paxson, Randy H. Katz, "What's New About Cloud Computing Security?" Technical Report No. UCB/EECS-2010-5. http://www.eecs.berkeley.edu/Pubs/TechRpts/2010/EECS-2010-5.html

[11] Deyan Chen, Hong Zhao, "Data Security and Privacy protection Issues in Cloud Computing", in proceedings of International Conference on

Computer Science and Electronics Engineering, 2012 IEEE DOI 10.1109/ICCSEE.2012.193.

[12] "IBM Discovers Encryption Scheme That Could Improve Cloud Security, Spam Filtering," at http://www.eweek.com/c/a/Security/IBMUncovers- Encryption-Scheme-That-Could-Improve-Cloud-Security- Spam-Filtering-135413/.

[13] Roy I, Ramadan HE, Setty STV, Kilzer A, Shmatikov V, Witchel E. "Airavat: Security and privacy for MapReduce," In: Castro M, eds. Proc. of the 7th Usenix Symp. on Networked Systems Design and Implementation. San Jose: USENIX Association, 2010. 297.312.

[14] "OASIS Key Management Interoperability Protocol (KMIP) TC", http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=kmip.

[15] Zeng K, "Publicly verifiable remote data integrity," In: Chen LQ, Ryan MD, Wang GL, eds. LNCS 5308. Birmingham: Springer-Verlag, 2008. 419.434.

[16] Cong Wang, Qian Wang, Kui Ren, and Wenjing Lou, "Ensuring Data Storage Security in Cloud Computing," in Proceedings of the 17th International Workshop on Quality of Service.2009:1-9.

[17] Bowers KD, Juels A, Oprea A. Proofs of retrievability: Theory and implementation. In: Sion R, ed. Proc. of the 2009 ACM Workshop on Cloud Computing Security, CCSW 2009, Co-Located with the 16th ACM Computer and Communications Security Conf., CCS 2009. New York: Association for Computing Machinery, 2009. 43.54. [doi: 10.1145/1655008.1655015]

[18] Muntés-Mulero V, Nin J. Privacy and anonymization for very large datasets. In: Chen P, ed. Proc of the ACM 18th Int'l Conf. On Information and Knowledge Management, CIKM 2009. New York: Association for Computing Machinery, 2009. 2117.2118. [doi:10.1145/1645953.1646333]

[19] Randike Gajanayake, Renato Iannella, and Tony Sahama, "Sharing with Care An Information Accountability Perspective," Internet Computing, IEEE, vol. 15, pp. 31-38, July-Aug. 2011.

[20] DoD, "National Industrial Security Program Operating Manual", 5220.22-M, February 28, 2006.

[21] Richard Kissel, Matthew Scholl, Steven Skolochenko, Xing Li, "Guidelines for Media Sanitization," NIST Special Publication 800-88,September 2006, http://csrc.nist.gov/publications/nistpubs/800-88/NISTSP800-88_rev1.pdf.

[22] Kamara S, Lauter K. Cryptographic cloud storage. 14th International Conference on Financial Cryptography and Data Security, LNCS, IFCA/Springer-Verlag. 2010, 6054: 136-149

[23] Barua M, Liang X, Lu X, et al. ESPAC: enabling security and patient-centric access control for eHealth in cloud computing. International Journal of Security and Networks, 2011, 6(2): 67-76

[24] Popa R A, Lorch J R, Molnar D, et al. Enabling security in cloud storage SLAs with CloudProof. Microsoft TechReport MSR-TR-2010, 2010, 46: 1-12.

[25] Somorovsky J, Meyer C, Tran T, et al. SEC2: secure moblie solution for distributed public cloud storages. 2nd International Conference on Cloud Computing and Services Science (CLOSER), 2012: 555-561.

Human Age Estimation using Computer Vision

Mukul Anand
Department of ECE,
CT Institute of Engineering
Management & Technology
Jalandhar, Punjab, India
mukul20anand@gmail.com

KomalGoel
Department of ECE,
CT Institute of Engineering
Management & Technology
Jalandhar, Punjab, India
komal_goel@rocketmail.com

Gurpreet Kaur
Department of ECE
CT Institute of Engineering
Management & Technology
Jalandhar, Punjab, India
gpreet.09@gmail.com

Abstract—**Development of computer vision system to estimate human age is of great importance in order to reduce inaccuracy of a face recognition system for not being able to recognize same person after aging. The inclusion of age estimation feature also adds higher standards of security into the system. This paper discusses various algorithms & models that are helpful in designing human age estimation based computer vision system.**

Keywords–Age Manifold, Active appearance model, Computer Vision, Craniofacial Morphology, Facial representation models, Human Age Estimation, Principle Component Analysis, Regression.

I. INTRODUCTION

Human Age Estimation (HAE) is a field of research that deals with the determination of a human's age/ age group using computer vision. Geng et al [1] discusses that the age of a subjected person can be decided by its i) Chronological age, defined by the number of years lived, ii) Person's physical Appearance iii) Perceived age, defined by other people on the basis of a person's appearance, or iv) Computer based age estimation. Researchers are focused on computer based age estimation because nowadays several face recognition systems are being employed for access control but the inclusion of age estimation in these systems guarantees additional security aspects. HAE is one of the applications of facial image classification that uses natural biometric features of human as reference for building a computational algorithm in order to develop an Age Specific Human Computer Interaction System (ASHCI) [1]. The process of age estimation uses a facial characteristic point as a standard reference point taken on human face of a 2-d facial image to estimate the age of a person. Craniofacial Morphology, [2] is a study of determining major changes in the skin and elasticity of face. Certain features of craniofacial morphology appears only in the people of certain age and changes during process [3]. The research on craniofacial morphology of individuals can be carried from different aspects. Patterson et al [4] conducted one of these studies and proposed an aging function based on the Active Appearance Model (AAM), which is based on use of

Principle Component Analysis (PCA), a technique defined by Manifold learning for Human Aging. In 1994, Kwon & Lobo [6] also proposed a theory & practical calculation method of age classification, which is based on the craniofacial morphology & wrinkle analysis. Similarly, Geng et al. [7], presented the Aging Pattern Subspace (AGES) for age estimation. This model performs aging pattern which is defined as a sequence of images sorted in time order to construct a representative subspace but there are some problems experienced in age estimation of the children.

II. PHASES OF FACIAL AGING

As discussed earlier, a facial classification system uses natural biometric features of human to perform desired analysis through sharp facial features, but the development of Human age estimation system major challenge in this regard. The natural biometric features of human change when a child evolve himself to a complete grown up & finally grow old. Geng et al [1] recognized two stages of facial aging.

A. *First phase* defines the years from birth to adulthood as shown in figure 1 [8].

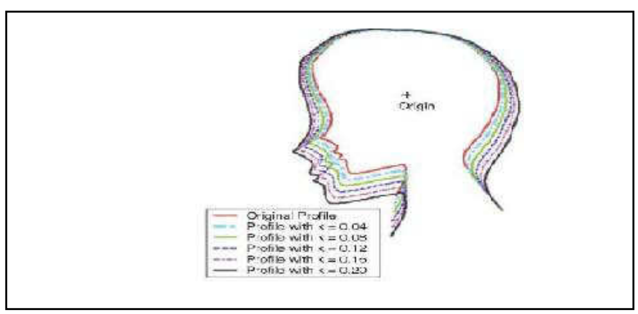


Fig.1. First Phase of facial aging

This phase includes most of the changes due to change in the craniofacial growth like Chin becomes more prominent, cheeks spread over a larger area, characteristics of the face increases and Forehead falls back to reduce the free space on skull. In addition to this some other minor changes in skin occur like Facial hair becomes denser & change color & Skin color changes.

B. *Second Phase* of the aging is duration from adulthood to old age.

The evolution of face as a result of two phases is shown in figure 2.This phase includes major changes in skin texture i.e. Skin becomes thinner, darker, less elastic and more leathery. In addition, wrinkles, blemishes, double chin, dropping cheek and eyelid bags appear.

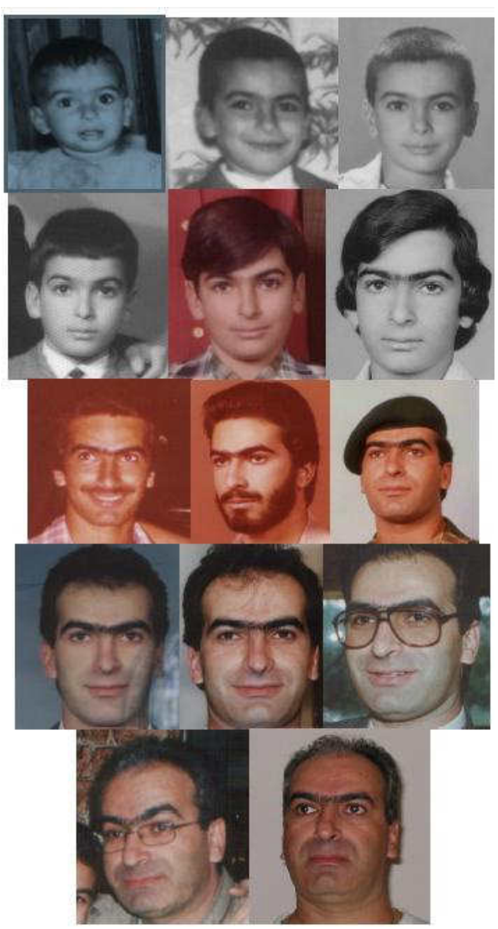


Fig.2. Evolution of human face due to aging (in FG-NET aging database)

III. REPRESENTATION MODELS

There are different models available for facial representation.

A. *Anthropometric Model*

This model which is based on the geometric ratios of faces proposed is by Kwon & Lobo [6]. The first step of this model is to define facial landmarks needed for calculation of ratios as shown on figure 3-8. After that Euclidean distances between those landmarks are calculated. Final step is presentation of research results. This model is useful for classifying minors & adults. But has a drawback that it can't distinguish between different ages of adult. This is the reason Kwon & Lobo [6] used wrinkle analysis to do the separation between young adults & seniors.

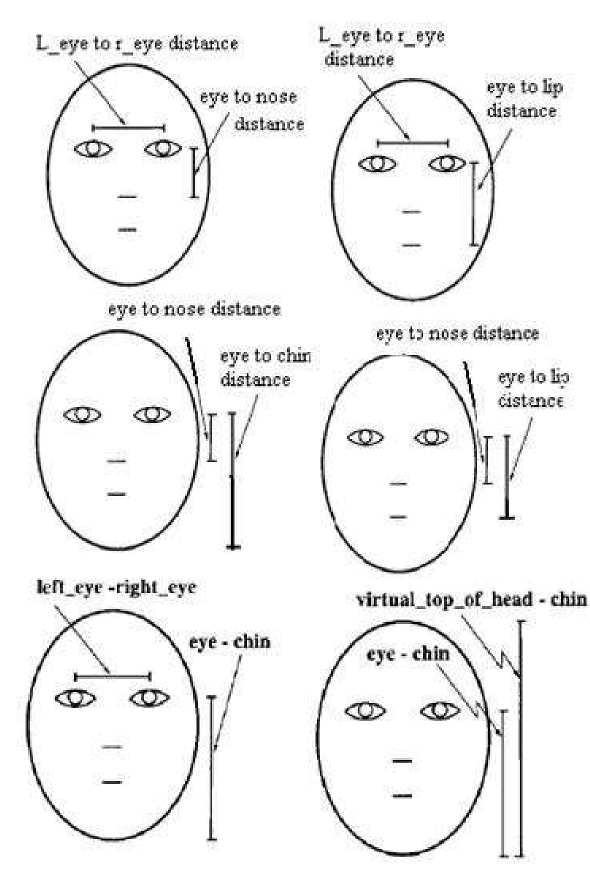


Fig.3-8.Calculation of geometric ratios of faces [3]

B. *Active Appearance Model*

In 1998, Cootes et al [9] proposed AAM model which further expanded to facial aging in 2002 by defining an age function as Age=f(b); to explain variation in years where age is the age of a person in the picture, b is a vector containing 50 parameters & f is an aging function. This function defines the relationship between person's age & facial description parameter [10]. There are different forms of an aging function, as an example,

1. quadratic age function
2. linear aging function
3. cubic aging function

The AAM is related to active shape mode. The only one disadvantage of active shape model is that it only uses shape constraints and does not take advantage of all the available information especially the texture across the target object. This can be modeled by AAM. This model required a training set of annotated images, where landmark points have been marked as shown in figure 9.

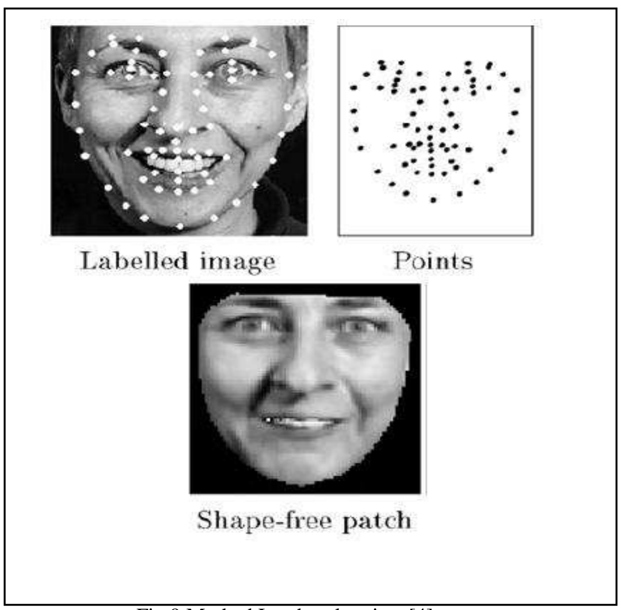


Fig.9.Marked Landmark points [4]

This should be stated that Active Appearance Model has some of the advantages over anthropometric model in terms of shape, texture, age and head poses of faces for which age is to be estimated. These advantages are summarized in Table1 below.

TABLE I. COMPARISON BETWEEN AAM AND ANTHROPMETRIC MODEL [13]

	Shape	**Texture**	**Any age**	**Head poses**
Anthropometric model	Yes	No	No	No
Active appearance model	Yes	Yes	Yes	Yes

C. *Aging Pattern Subspace*

It is possible to learn the common pattern of aging for more than one person at different ages [8], instead of learning the specific aging pattern for each person. This model was developed by Geng et al [1] named AGES (Aging pattern Subspace). Age pattern is defined as a sequence of facial image of a person, sorted by time. The AGES works in two steps:

1. Learning step, the standard PCA approach have probably images for each year for each aging pattern which don't required. So EM (Expectation Maximization) is used as differences between the available images of the face & face reconstructed images [10].
2. Age Estimation step, which is the test face image needs to find a pattern of aging that suits that images, & further extract position year in the sample. As a result position year returned is the estimated age of a person in the test image.

To deal with incomplete data effectively, due to difficulties in data collection, through the technique learning subspaces, the aging pattern subspace models the sequence of a person's aging face images. Age of tested person is determined by the projection in the subspace that can be reconstructed back.

D. *Age estimation Framework*

As shown in figure 10, age estimation framework consist of five modules, which are face detection, face normalization, manifold learning, robust regression, & local adjustment. Firstly from the image being captured a face is extracted or detected after which it goes through the normalization including geometric alignments & illumination normalization. Then the age manifold is applied to map the original face image data into a low dimensional subspace. A robust regression function is applied to fit the manifold data. Finally a local adjustment of the regression results is performed to refine the local fitting of the data.

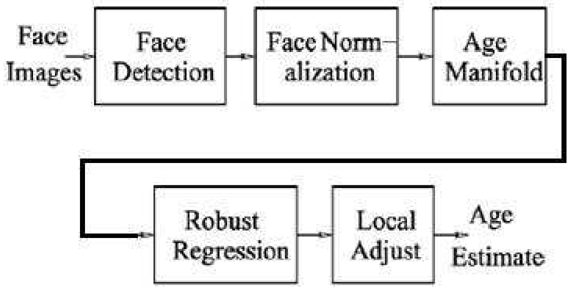


Fig.10. Modules in age estimation framework.

E. *Age manifold model*

Age manifold model defines an image space as a set of statically and dynamically aligned face images which is represented by X (eq. 1),

$$X = \{xi : xi \sum R^D\},\qquad (1)$$

Where,

i =1 to N, and

D= dimension of image in order to subject Age.

A labeling ground set true value is represented as L (eq. 2),

$$L = \{li : li \sum N\qquad (2)$$

This labeling parameter associates with the given image to provide the labeling to human age. Manifold learning provides the low learning manifold dimension in embedded subspace Y (eq. 3) which is represented as,

$$Y = \{yi : yi \sum R^d,\qquad (3)$$

Where, d<<D shows one to one mapping to X image. Now, the projection from the image space to the manifold space can be modeled as given in eq. 4,

$$Y = P(X, L)\qquad (4)$$

Here, P (.) denotes the projection function which can be linear and non-linear. The objective of manifold embedding is to find $(n\,X\,d)$ matrix P which satisfy eq. 5.

$$Y = P^T X\qquad (5)$$

Where, $Y = \{y1, y2, y3, \ldots\ldots yn\}$

$X = \{x1, x2, x3, \ldots\ldots xn\}$

$P = \{p1, p2, p3, \ldots\ldots pd\}$, and d<n;

A Linearly supervised manifold embedding technique searches the nearest neighbors in the data space. Some typical techniques used for dimension reduction and manifold embedding are as follows,

1. Principal component analysis (PCA): PCA is a technique which reduces the original variables (V) to principal components (P) [11] such that it satisfies eq. 6,

   $$P \leq V\qquad (6)$$

   This method finds the embedding that maximizes the projection variance P, as (eq. 7)

   $$P = p^T Sp;\qquad (7)$$

   Where, $S = \sum_{i=1}^{N}(xi - x')(xi - x')^T$, which is a scalar matrix,
   x' = mean vector of {xi}, and
   i=1to n.

2. Locally Linear Embedding (LLE) [9]: The LLE algorithm seeks the nonlinear embedding in neighborhood. It exploits the local symmetries of linear object class reconstructions in a preserving manner and seeks the optimal weights for local reconstruction.
3. Orthogonal Locality Preserving Projections (OLPP): The OLPP method produces orthogonal basis functions based on the LPP [12] to obtain more discriminating power for embedding. The LPP searches the embedding that preserves essential manifold structure by measuring the local neighborhood distance information. It defines the affinity weight as s_{ij} (eq. 8), when x_i and x_j are nearest neighbors, otherwise, *sij= 0,*

   $$sij = \exp(-xi - xj\,2)/t),\qquad (8)$$

   And, S(*i,j*)= *sij* is a symmetric matrix .

IV. COMPARISON OF ALGORITHMS

There is large number of algorithms available. To compare these algorithms, Mean Absolute Error (MAE) is measured [13]. The lower value of MAE signifies more accuracy for the specific algorithm. Table 2 summarizes MAE values for various algorithms used in age estimation process. This analysis is performed on FG-NET database.

TABLE II. MAE FOR VARIOUS ALGORITHMS [13]

Name of algorithm	**MAE**
K-neared neighbor	8,24
Support Vector Machines	7,25
Local feature of face image and regression	6,85
Aging Pattern Subspace	6,77
Nonlinear Aging Pattern Subspace	6,18
Ranking with uncertain labels	5,33
Metric learning and Gaussian Process Regression	5,08
Manifold Learning and locally adjusted robust regression	5,07
Regression Patch Kernel	4,95
Active Appearance Model	4,37
Enhanced bio-inspired features	3,17

CONCLUSION

Once explored & implemented properly, an automatic computer vision based Human Age Estimation system may find a wide area of useful applications. These may include:

1. *Security:* Age of an intruder who either doesn't carry any document for age verification or carry forged documents can be estimated [14].
2. *Parent control*: An improved face recognition based access control facility can be designed to provide a user control for minors (users with age less than 18 years).
3. *Investigation*: Computer Vision based database matching system can be designed while searching for a missing person after many years

This paper discussed various algorithms used for human age estimation. Although the choice of algorithm for computing Human age estimation using computer vision depends on many factors, but it may be observed that active appearance model (AAM) not only possesses many advantages (as discusses earlier in Table 1) but also have least value of MAE in comparison to most of the other available algorithms (as presented in Table 2). Apart from this algorithm, an advanced form of algorithm that uses enhanced bio-inspired features [15] is also evolved recently.

It is also suggested that Future research in human age estimation shall be carried by classifying a large sample of subjected individuals in more than one smaller samples on the basis of their facial images & then relevant classification algorithm shall be applied.

REFERENCES

[1] Xin Geng, ZhiHua Zhou, Yu Zhang, Gang Li and Honghua Dai, "Learning from Facial Aging Patterns for Automatic Age Estimation" Proceedings of 14th annual ACM international conference on multimedia, 2006.

[2] Koruga Petra, Ba a Miroslay and Schatten Markus, "Analysis of Craniofacial Morphology Changes during Aging and their Connection with Facial Age Estimation", Proceedings of the Information Technology Interfaces, 2011

[3] E. Patterson, K. Ricanek, M. Albert, E. Boone, "Automatic representation of adult aging in facial images", Proceedings of the 6th IASTED International Conference on Visualization, Imaging, and Image Processing, 2006, pp.171-176

[4] Dihong Gong , Zhifeng Li, Dahua Lin, "Hidden Factor Analysis for Age Invariant Face Recognition", International Computer Vision event (ICCV) – sydney, Australia, 2013

[5] Young H. Kwon, Niels da Vitoria Lobo, "Age Classification from Facial Images", Computer Vision and Image Understanding, Vol. 74, No. 1, April, 1999, pp. 1–21

[6] Geng Xin, Fu Yun, Smith‐Miles Kate, Automatic Facial Age Estimation, Tutorial at PRICAI 2010

[7] Todd, J. T.,Mark, L.S., Shaw, R.E., & Pittenger, J.B., "The perception of human growth", Scientific American, 1980, pp:242

[8] Cootes, T.F., Edwards G.J., Taylor C.J. "Active Appearance Models", Proceedings of European Conference on Computer Vision, Vol. 2, Springer, 1998

[9] Yun Fu, Guodong Guo, Thomas S. Huang, "Age Synthesis and Estimation via Faces: A Survey". IEEE Transactions on Pettern Analysis and Machine Intelligence, 2010, pp. 32-43

[10] Santoshi, Vikas, and Amit Kr. Gautam, "Comparative survey on age invariant face recognition and verification techniques", IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE) e-ISSN: 2278-1676, 2014, pp. 33-37

[11] Guodong Guo, Yun Fu, Charles R. Dyer and Thomas S. Huang "Image-Based Human Age Estimation by Manifold Learning and Locally Adjusted Robust Regression", IEEE transactions on image processing, vol. 17, 2008, pp. 1178- 1188.

[12] Petra GRD, "Introduction to human age estimation using face images", Research papers, Faculty of material science and technology in TRNAVA, De Gruyter online, 2013

[13] Petra Koruga, Miroslav Ba a and Jurica Ševa " Application of Modified Anthropometric model in Facial Age Estimation", ELMAR proceedings, 2011

[14] Guodong Guo, Guowang Mu, Yun Fu, Thomas S. Huang, "Human Age Estimation Using Bio-inspired Features", IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp.112-119.

Improved Encryption Algorithm for Enhanced Security Based on Position Index

Shipra Jain
Ambedkar Institute of Advanced Communication Technologies and Research
Delhi
sjshiprajain@gmil.com

Vishal Bhatnagar
Ambedkar Institute of Advanced Communication Technologies and Research
Delhi
vishalbhatnagar@yahoo.com

***Abstract*—For providing security to the data, DNA cryptography is a new emerging field. It uses the DNA as an information carrier to transmit information between the sender and the receiver through the open environment. DNA cryptography encryption algorithms make use of various traditional methods and biological process like translation, splicing etc. to provide security to the data. In this paper, the authors present an encryption decryption algorithm by using the reference DNA sequence as a key. The authors also used a traditional base4 conversion method and a position change method to provide the encryption to the data.**

Keywords—component, formatting, style, styling, insert (key words)

I. INTRODUCTION

Sensitive information such as financial transaction, medical and personal records, passwords and other crucial data are transmitted through the public communication networks. The security of the sensitive information from the unintended receiver is very important. It poses a great threat if sensitive information is delivered to an unintended receiver. The storage devices used for storing the information such as optical drives, hard drives, pen drives, DNA also require security. Various Steganographic and cryptographic algorithms are used to provide security of the stored and transmitted data. In this paper authors present a new way for providing security to the data store in the DNA. DNA is a small component of the human body. It can store data of 10^6TB in 1gm of DNA. As every storage device requires security, the DNA also requires security for its data. Security in DNA is provided by both cryptography and steganography. DNA cryptography is a new cryptographic field based on the concept of DNA computing. In this paper authors propose a new method of providing security to the data stored in DNA. The authors used traditional methods for providing security to the binary data and indexing method for providing security of the equivalent DNA code.

II. DNA: AN INTRODUCTION

DNA is an abbreviation for Deoxyribonucleic Acid. It is a long molecule contained in the nuclei of cells. It is the building block of every living organism. DNA was identified by Swiss physician Friedrich Miescherin 1869. In 1878 Albrecht Kossel isolated the non-protein components of nucleic acid and later isolated its five primary nucleobases [1].DNA is a double helix, complementary structure consists of four bases: adenine (A), thymine (T), cytosine (C) and guanine (G). Watson Crick proposed the complementary rule that adenine will always make a pair with thymine (A-T) and cytosine always make a pair with guanine (G-C) [2]. Every cell in the living organism has a complete set of DNA. DNA is unique for each individual. Different biological operations can be performed on DNA like ligation, polymerase chain reaction (PCR), gel electrophoresis.

III. DNA CRYPTOGRAPHY

DNA cryptography is the new branch of cryptography. In DNA cryptography, DNA is used to store and transmit information. DNA cryptography uses the properties and storage power of DNA to hide the information from the intruder. DNA cryptography was introduced in 1994 by Dr. Leonard M. Adelman of the University of South California to solve the complex mathematical problem [3]. Here, data is converted in the form of nucleotides to store in DNA. For storing data in to data is converted into binary sequence or in any other number system like base4 or base5. The conversion method is different for different researchers or it may be same. There is no standard method for this [4]. Modern technology allows expressing of data in the form of DNA [5]. DNA cryptography can be classified into three subfields:

1. Symmetric Key DNA Cryptography
2. Asymmetric Key DNA Cryptography
3. Pseudo DNA Cryptography

Symmetric key DNA cryptography uses the same key for encryption and decryption. A key is shared between the sender and the receiver. Asymmetric key cryptography uses two keys, a public key and a private key. The Public key is known to everyone and is used for decrypting the data. The Private Key is used for encrypting the data and is secret to the person. It is not shared with anyone. Pseudo DNA cryptography depends on the function of DNA not on the DNA strands. In this method encryption and decryption is based on the flow of genetic information within a biological system. It adds the artificial features to the cipher text. The artificial features help in protecting data from an attacker [6].

IV. PROPOSED ALGORITHM

Here, an encryption algorithm is proposed using the base4 conversion and position index of DNA sequence. In this method only the reference sequence is shared between the receiver and sender of the data. Hence, there are very less chances of the security breach. A large database of the chromosome is freely available on the internet. It is very difficult for the attacker to try the different chromosome DNA sequence and get the useful data. The framework of a proposed method is shown in the figure1 below. By this method, the sender can send any type of data like audio, video, image and text including the special characters.

Step 1: Input any audio, video or text file.

Step2: Find its ASCII value.

Step3: Convert the ASCII value into base4 number system.

Step4: Scan the first digit and the last digit from the base4 digits calculated above.

Step5: Step 4 is repeated until every digit is scanned.

Step 6: Convert the base4 digits into DNA sequence according to the coding rule.

Step7: Choose a reference DNA sequence from the ncbi database as a key.

Step8: Share the reference sequence between the sender and the receiver.

Step9: Group two elements from the DNA code sequentially.

Step10: Find the position of a DNA pair in the reference DNA sequence.

Step11: Pass this position sequence to the receiver of the message.

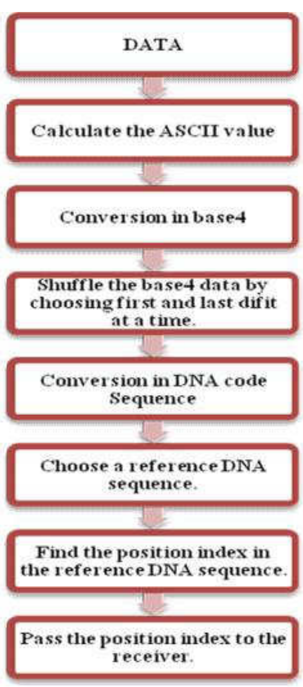


Fig. 1. Encryption Method

A. Coding Method

In order to convert the plaintext into the DNA sequence, authors first calculate the ASCII value of the data. ASCII value is calculated because every data in the digital world have some specific and unique ASCII value. By this ASCII value, the data is converted into DNA four nucleotide sequences {A, T, G and C}. After calculating the ASCII value, it is converted into a base4 sequence. The base 4 sequence consists of 4 elements {0, 1, 2 and 3}. The base4 sequence is chosen because DNA also has the 4 elements.

B. First Level Encryption

In this method the authors provide the encryption of data by two levels. First level encryption at the base4 data and second level encryption at the DNA sequence element. The encryption or information hiding at two levels increases the security of the data. Encryption at the first level is provided by shuffling digits after the base4 conversion. The digits are shuffled by choosing first digit from the starting and second digit from the last. This shuffling changes the sequence of digits of plain text. This shuffling does not require any key. Hence reduces the risk of information breach by loosing key. The shuffling is shown below by example:

0	1	2	3
0	3	1	2

C. DNA Sequence

After encrypting the base4 sequence, this sequence is converted into the DNA sequence. The authors make the following coding rules to map the base4 sequence into the DNA sequence.

Table I. DNA Code

DNA Element	Base4 Digit
A	0
T	1
G	2
C	3

These rules may vary from author to author. There is no standard rule for choosing the mapping relation between the base4 digits and DNA elements.

D. DNA Sequence Encryption

This is an encryption at the second level. Here the authors encrypt the DNA sequence. The authors use a reference DNA sequence which is share between the sender and the receiver to encrypt the DNA data. This reference DNA sequence will act as a key for data encryption. The reference DNA sequence is chosen from the large database available at the internet. After choosing and sharing the reference DNA sequence, the sender makes a group of two DNA elements and found their position in the reference DNA sequence. This position index is send to the receiver through the open or unsecured environment. This is an encrypted DNA sequence delivered to the receiver. This technique is used by researchers in Index Based Symmetric DNA encryption algorithm [7].

E. Decryption

Decryption is the reverse of complete encryption process. For the decryption of coded message, sender and receiver have shared the reference DNA sequence. This reference DNA sequence will act as a key between the sender and receiver. This key should be keep confidential and secret to provide security of the data. The Receiver receives the position index of DNA sequence as the encrypted data. This position index gives the DNA sequence from the reference DNA sequence. After getting the DNA sequence, receiver will convert it into the base4 digit by reversing the coding table and gets the encrypted base4 data. The decryption of base4 data is done by putting first digit in the starting and second digit in the extreme end. Then put third digit at the second position and fourth digit at the left of end digit. Complete decryption is done by repeating the above step. This will give the plaintext base4 data. Convert it into the decimal value to get the ASCII value of the actual plaintext data. After getting the ASCII value, the receiver will get the original message. This step completes the decryption process. The steps for decryption are shown in the figure 2.

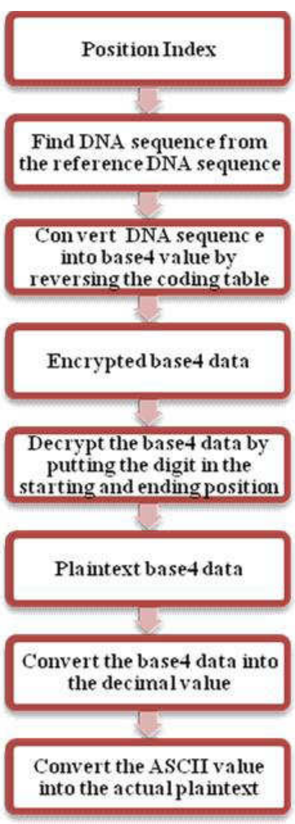


Fig. 2. Decryption Method

V. ILLUSTRATION

In this example a plaintext has been taken by the authors. For example, the plain text is “HELLO”. The authors then find its ASCII value and convert it into base4 number system.

Plaintext: HELLO

ASCII value: 73 70 77 77 79.

Base4 Value: 1021 1012 1031 1031 1033

After calculating the base4 values, the authors applied the technique of encryption at the first level.

Plaintext: 10211012103110311033

Cipher Text: 13032011110310211103

Now convert this binary cipher text into DNA sequence. This DNA sequence will act as a plaintext for the encryption at the second level.

Plaintext: 13032011110310211103

DNA Sequence: TCACGATTTTACTAGTTTAC.

Now this sequence is a plaintext and its position is found in the reference DNA sequence.

In this example the reference DNA sequence is CANIS FAMILIARIS GENOME. This chromosome name is shared between the sender and the receiver. Its DNA sequence is taken from the ncbi database [8] shown below:

AGGTCTGGGACCCACCTCCCCTGAGCTGACACTGAG
CTTGAATCTGAAAGAGCAGGCTGCCAAGGTCTAAAG
CAGCAGAAGCTGGTATGGGTGGTGGATCCTGAGGGG
GGAACGGTTGCAGTGTCTCTGAGTGACAAGGACAAA
GTCTGCTGCATCCAGCCTCAATGTTTCATCCCCAGTG
GTCATCAAGTCCTGCCAAAGTTCTTGCATTCACGCTG
GGCGGGATCTTAGGCCTTCTGCTTCTAATTGGCTTTG
CGTCTTCTGCTGTGTTAAGACTCTGCCCAAGGCCCAG
CAGCAGCGGATCT.

Now the position index of a plaintext DNA sequence is found in the reference DNA sequence and send to the receiver. The position sequence is found by making a group of two nucleotide elements sequentially.

The position sequence of the DNA sequence of HELLO plaintext is:

Position Index: {4,10,23,28,116,134,223,252,255,271}.

VI. ADVANTAGES

- Key Requirement for Binary Encryption: No need of key for encrypting the base4 data. Hence, the security of data is not depends upon the key.
- Padding Requirement: There is no requirement of extra digit padding. Every digit encryption and decryption is independent.

- Large Key Space: The proposed method uses a reference DNA sequence. The database of a chromosome is so vast and easily available. Hence brute force attack requires more time and it is very complex to occur.
- No Mathematical Calculation: the proposed method has any complex mathematical calculation. It is a quite simple method requires only conversion technique.
- This method can use any type of data whether it is audio/video file or image or text or any other document.

VII. DISADVANTAGES

- As there is no complex mathematical calculation is required in the proposed system. The complete security of the system is depending upon the key. If the intruder is able to detect the reference DNA sequence. Then, he will easily decrypt the secret message.
- The sender and receiver must have the knowledge of biological domain to understand the DNA sequence of reference chromosome

VIII. CONCLUSION

In this paper, the author present a simple and secure way of providing security to the data stored in the form of DNA sequence. The authors proposed a symmetric key cryptosystem by using the reference DNA sequence and applying index. This method provides the two level securities, first at the base4digits and second at the DNA sequence. The authors have used a transposition cipher for providing security at the first level. The transposition cipher does not alter the data it only changed its place. At the second level, the authors use a reference DNA sequence to find the position of the plaintext DNA sequence. By providing the example, the authors can conclude that the proposed method is simple and secure.

REFERENCES

[1] M.E. Jones, "Albrecht Kossel, A Biographical Sketch", Yale Journal of Biology and Medicine, Vol. 26, Issue No. 1,pp. 80-97, 1953.

[2] J.D.Watson, F.H.C, Crick, "A structure for deoxyribose nucleic acid", Nature, Vol. 25, pp. 737-738, 1953.

[3] L. Adelman, "Molecular computation of solutions to combinatorial problems," Science in JSTOR, vol. 266, pp. 1021–1025, 1994.

[4] Monica Borda, "DNA secret writing Techniques" 8th International Conference on Communication, 2010

[5] G. Z. Xiao, "New field of cryptography: DNA cryptography," Chinese Science Bulletin, vol. 51, pp. 1139–1144, 2006.

[6] Ning Kang, A pseudo DNA cryptography Method, http://arxiv.org/abs/0903.2693 ,2009

[7] Z. Yunpeng, Z.Yu, W.Zhong and R.O.Sinnott,"Index-Based Symmetric DNA Encryption Algorithm", 14th International Congress on Image and Signal Processing (CISP) , Shanghai, Vol. 5,pp. 22990-2294, 2011.

[8] DNA sequence access , "http://www.ncbi.nlm.nih.gov/nuccore/L34212.1, on 21 Jan, 2014.

Comparative Analysis of Software Process Models

Sehaj Rai
Dept. of Computer Science & Engg.
GNDU, Amritsar, India
sehajrai86@gmail.com

Navjot Kaur Jassi
Dept. of Computer Science & Engg.
GNDU, Amritsar, India
navjotkjassi@gmail.com

***Abstract* - The field of software engineering is related to the development of software. Large software needs systematic development unlike simple programs which can be developed in isolation and there may not be any systematic approach being followed. In the last few decades, the computer industry has undergone revolutionary changes in hardware. This research deals with a critical and important issue in computer world. It is concerned with the software management processes that examine the area of software development through the development models, which are known as software development life cycle. It represents three of the development models namely, waterfall, V-shaped and spiral. These models have advantages and disadvantages as well. Therefore, the main objective of this research is to represent different models of software development and make a comparison between them and focuses on security activities involve in developing secure software's. Identifying security risks and managing those risks based on spiral model. A new spiral model with Multilevel Security Spiral (MSS) has been proposed with security activities.**

Keywords - Software process, Waterfall Model, V-Shaped Process Model, Spiral Model, MSS.

I. INTRODUCTION

No one can deny the importance of computer in our life, especially during the present time. In fact, computer has become indispensible in today's life as it is used in many fields of life such as industry, medicine, commerce, education and even agriculture. It has become an important part in the industry and technology of advanced as well as for developing countries.[5] Computer is considered a time- saving device and its progress helps in executing complex, long, repeated processes in a very short time with a high speed performance. During the previous four decades, software has been developed from a tool used for analyzing information or solving a problem to produce the final product. However, the early programming stages have created a number of problems turning software an obstacle to software development particularly those relying on computers. Software consists of documents and programs that contain a collection that has been established to be a part of software engineering procedures. Moreover, the aim of software engineering is to create a suitable work that construct high quality programs.[6]

In the early years, small size of the software was being used and those were developed either by a single programmer or by a small programming team. Security for a software system has always inverted and address solely within the production environment through perimeter security like firewall, proxy, antivirus, platform security, and intrusion prevention system[4]. This is the reason for considering security as a non-functional requirement. The non functional requirements are separate from procedural requirements that includes; maintainability, reusability, reliability portability, and security [8]. The security criteria which include authentication, privacy, authorization, and integrity of some applications exhibit to less variation.

The three main aspects of security are Confidentiality, Integrity, and Availability. The given aspect of security is also referred as CIA [8]. Confidentiality is all about maintaining privacy from unauthorized users. Integrity is about ensuring the accuracy and completeness of information. Availability is ensuring about the information available to authorized users [4]. The main focus of security activity is to identify security risks and managing those risks from the early stage of each phase. Based on spiral model and its risk perception, a new model has been proposed and named as Multilevel Security Spiral (MSS).

II. SOFTWARE PROCESS

The software process is a critical factor for delivering quality software systems, as it aims to manage and transform the user need into a software product that meets this need. In this context, software process means the set of tools, methods, and activities required to produce a software system, executed by a group of people organized according to a given organizational structure and counting on the support of techno-conceptual tools. Historically, software developments have largely been product centered. Recently, many researchers and practitioners have refocused their efforts on the process dimension of software engineering. Software processes performed during software Development and evolution are becoming rather complex and resource intensive. They involve people who execute actions with the primary goal to create quality software in accordance with the previously set user requirements. Constant monitoring and improvement of software processes is therefore of a significant interest for organizational performing software development and maintenance. in order to improve the process an objective description and evolution of the existing process is needed.[2]

A. Software Process Models

Software process models represent a networked sequence of activities, objects, transformations, and events that embody strategies for accomplishing software evolution. Such models can be used to develop more precise and formalized descriptions of software life cycle activities [1].Their power emerges from their utilization of a sufficiently rich notation, syntax, or semantics, often suitable for computational processing. Software process networks can be viewed as representing multiple interconnected task chains. Task chains represent a non-linear sequence of actions that structure and transform available computational objects (resources) into intermediate or finished products. Task actions in turn can be viewed a non-linear sequence of primitive actions which denote atomic units of computing work. Task chains are also known to be as a "workflow" .Task chains can be employed to characterize either prescriptive or descriptive action sequences. Prescriptive task chains are idealized plans of what actions should be accomplished, and in what order. The primary function of software development process models is to determine the order of the stages involved in software development and evolution and to establish the transition criteria for progressing from one stage to the next. In history various models were proposed. [5]

Software Process Models are:

1. Waterfall model: Separate and distinct phases of specification and development.
2. Prototype model.
3. Rapid application development model (RAD).
4. Evolutionary development
5. Incremental model.
6. Iterative model.
7. V-shaped model
8. Spiral model.
9. Component-based software engineering

This paper presents the three process models: i) waterfall process model, ii) V-shaped model, and iii) spiral model and then MSS.

1. Waterfall Process Model

The Classical Life Cycle or the Waterfall Process Model was the first process model to present a sequential framework, describing basic stages that are mandatory for a successful software development model. It formed the basis for most software development standards and consists of the following phases:

i. Requirement Gathering and analysis: All possible requirements of the system to be developed are captured in this phase and documented in a requirement specification doc.[2]

ii. System Design: The requirement specifications from first phase are studied in this phase and system design is prepared. System Design helps in specifying hardware and system requirements and also helps in defining overall system architecture.

iii. Implementation: With inputs from system design, the system is first developed in small programs called units, which are integrated in the next phase. Each unit is developed and tested for its functionality which is referred to as Unit Testing.

iv. Integration and Testing: All the units developed in the implementation phase are integrated into a system after testing of each unit. Post integration the entire system is tested for any faults and failures.

v. Deployment of system: Once the functional and non functional testing is done, the product is deployed in the customer environment or released into the market.

vi. Maintenance: There are some issues which come up in the client environment and also to enhance the product some better versions are released. Maintenance is done to deliver these changes in the customer environment.

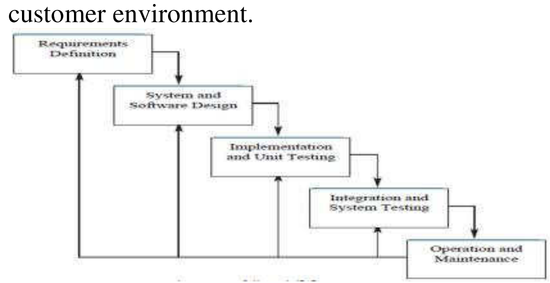


Fig. 1. Waterfall model

In each stage, documents that explain the objectives and describe the requirements for that phase are created. At the end of each stage, a review to determine whether the project can proceed to the next stage is held. Your prototyping can also be incorporated into any stage from the architectural design and after.

a) Waterfall Model Application

Every software developed is different and requires a suitable SDLC approach to be followed based on the internal and external factors. Some situations where the use of Waterfall model is most appropriate are:

i. Requirements are very well documented, clear and fixed.

ii. Product definition is stable.

iii.Technology is understood and is not dynamic. [3]

iv.There are no ambiguous requirements.

v.The project is short.

Advantages and Disadvantages of waterfall model are shown below in Table1.

Table I. Advantages and Disadvantages of waterfall model

Sr.no	Advantages	Disadvantages
1.	Simple and easy to understand and use.	No working software is produced until late during the life cycle.
2.	Easy to manage due to the rigidity of the model – each phase has specific deliverables and a review process.	High amounts of risk and uncertainty.
3.	Phases are processed and completed one at a time.	Not a good model for complex and object-oriented projects.
4.	Works well for smaller projects where requirements are very well understood.	Poor model for long and ongoing projects.[5]
5.	Clearly defined stages.	Not suitable for the projects where requirements are at a moderate to high risk of changing. So risk and uncertainty is high with this process model.
6.	Easy to arrange tasks.	It is difficult to measure progress within stages and Integration is done as a "big-bang" at the very end, which doesn't allow identifying any technological or business bottleneck or challenges early.
7.	Process and results are well documented.	Cannot accommodate changing requirements.

2. V- Shaped Process Model

V- Shaped Model is an extension of the waterfall model and is based on association of a testing phase for each corresponding development stage. This means that for every single phase in the development cycle there is a directly associated testing phase. This is a highly disciplined model and next phase starts only after completion of the previous phase. Just like the waterfall model, the V-Shaped life cycle is a sequential path of execution of processes. Each phase must be completed before the next phase begins. Testing is emphasized in this model more than the waterfall model. The testing procedures are developed early in the life cycle before any coding is done, during each of the phases preceding implementation. Requirements begin the life cycle model just like the waterfall model. Before development is started, a system test plan is created. [7] The test plan focuses on meeting the functionality specified in requirements gathering. The high-level design phase focuses on system architecture and design. An integration test plan is created in this phase in order to test the pieces of the software systems ability to work together. However, the low-level design phase lies where the actual software components are designed, and unit tests are created in this phase as well. The implementation phase is, again, where all coding takes place. Once coding is complete, the path of execution continues up the right side of the V where the test plans developed earlier are now put to use. [5]

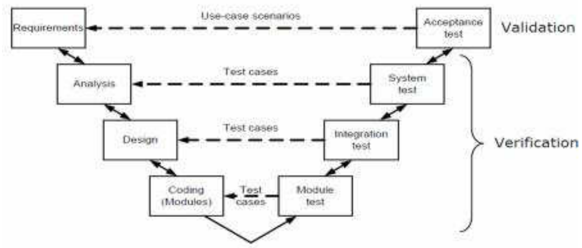


Fig. 2. V- Shaped process Model

a) V-Model Application

This model is used in the medical development field, as it is strictly disciplined domain. Following are the suitable scenarios to use V-Model:

i. Requirements are well defined, clearly documented and fixed.

ii.Product definition is stable.[6]

iii.Technology is not dynamic and is well understood by the project team.

iv.There are no ambiguous or undefined requirements

v. The project is short.

Advantages and Disadvantages of V-model are shown below in Table2.

Table II. Advantages and Disadvantages of V-Model model

Sr.no	Advantages	Disadvantages
1.	Simple and easy to use.	Very rigid like the waterfall model.
2.	Each phase has specific deliverables.	Little flexibility and adjusting scope is difficult and expensive.
3.	Higher chance of success over the waterfall model due to the early development of test plans during the life cycle.	Software is developed during the implementation phase, so no early prototypes of the software are produced.
4.	Works well for small projects where requirements are easily understood.	This Model does not provide a clear path for problems found during testing phases.

3. Spiral Model

In Spiral model, instead of presenting a sequence of activities with some backtracking from one activity to the other, the process model followed a spiral organization of activities. It combines characteristics of both prototype and waterfall process model. The model is divided into some task regions, which are as follows: Customer Communication, Planning, Risk Analysis, and Engineering, Construction and release and Customer evaluation. The distinctive feature of this model is that each stage is controlled by a specific risk management criteria ensuring decision making using critical factors. [2]

a) Spiral Model Application

Spiral Model is very widely used in the software industry with the natural development process of any product i.e. learning with maturity and also involves minimum risk for the customer as well as the development firms.

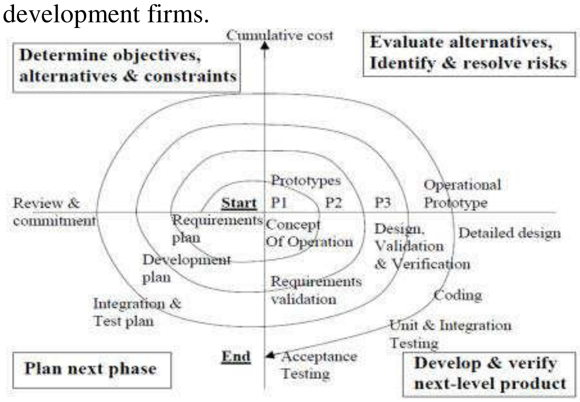


Fig. 3. Spiral Model

Following are the typical uses of Spiral model:

i. When cost there is a budget constraint and risk evaluation is important.

ii. For medium to high-risk projects.

iii. Long-term project commitment because of potential changes to economic priorities as the requirements change with time.

iv. Customer is not sure of their requirements which are usually the case.

v. Requirements are complex and need evaluation to get clarity.

vi. New product line which should be released in phases to get enough customer feedback.

vii. Significant changes are expected in the product during the development cycle. [1]

Advantages and Disadvantages of Spiral model are shown below in Table3.

Table III. Advantages and Disadvantages of spiral model

Sr.no	Advantages	Disadvantages
1.	Changing requirements can be accommodated.	A number of risks, constraints, alternatives, models etc. need to be analyzed but never are these risks or objectives listed and no specific risk analysis technique is mentioned. If risk analysis is poor the end product will surely suffer.
2.	Allows for extensive use of prototypes	Another difficulty of the spiral model is adjustment of contract deadlines using the spiral model.
3.	Requirements can be captured more accurately	For large projects expert software developers can produce efficient software products but in case of a complex large project absence of specific risk analysis techniques and presence of varying expertise can create a chaos.
4.	Users see the system early.	Risk analysis expertise is vital
5.	Development can be divided in to smaller parts and more risky parts can be developed earlier which helps better risk management.	Doesn't work well for smaller projects

III. MULTILEVEL SECURITY SPIRAL (MSS) MODEL

Security activities for development of secure software models should be followed right from the beginning. A careful consideration for security is required in pre-requirements, requirements, design, implementation, testing, deployment and maintenance phase. The main focus of security spiral is to identify security risks and managing those risks [4]. Multilevel Security Spiral (MSS) considers security from the very beginning of the software development life cycle [2]. Spiral model is as organized approach for developing software. Spiral model is based on risk perception for project.

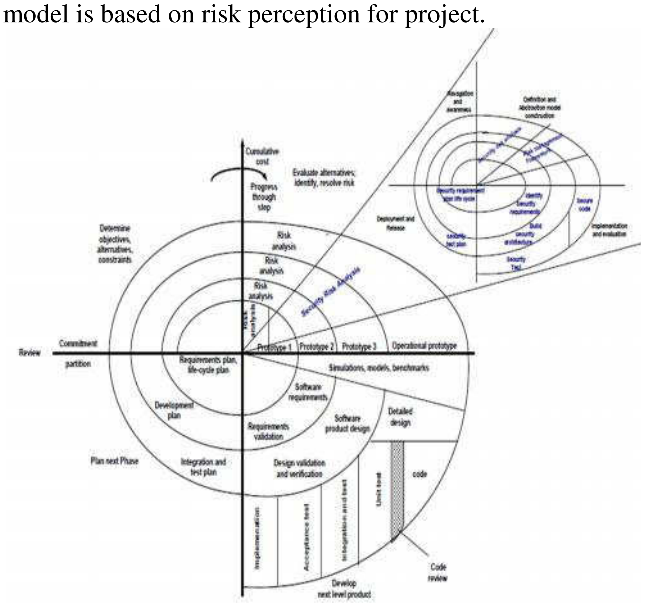


Fig. 4. Multilevel Security Spiral (MSS) Model

Each cycle of the multilevel spiral model begin with identification of objectives, constraints, alternatives doable for achieving software. Fig 5 shows the security risk analysis in beginning of each cycle in spiral. After identifying the risk, next step is to develop strategy to resolve the uncertainty & risks [4]. This can be done by security spiral model which is included in the security risk analysis area. The activities in secure spiral model are organized very much like a traditional spiral that has many cycles. In the beginning of cycle, develop framework for risk management. After resolve risk, next step is to develop next product in which we create the software. Including security in requirement phase and in design phase, after every step it must be revised by the programmer considering the risks. If any possible risk is encountered then it should be resolve in that step. At each step, extensions and design modification should be made. Secure code should written code must be reviewed by the programmer himself considering the risks. Security test should be there after integration test before implementation. So that the Multilevel Security Spiral (MSS) model should be risk free and highly secure. Multilevel Security Spiral (MSS) model is better in all prospective from the other models.

Comparison of Waterfall, V-Shaped, Spiral and MSS Process Models is shown below in Table4.

Table IV. Comparison of Waterfall, V-Shaped, Spiral and MSS Process Models

Features	Waterfall Model	V-shaped Model	Spiral Model	MSS
Requirement Specification	Beginning	Beginning	Beginning	Beginning
Cost	Low	Expensive	Expensive	Expensive
Overlapping phases	No	Yes	Yes	Yes
Understanding requirements	Not well understood	Well understood	Well understood	Well understood
Complexity of system	Simple	Complex	Complex	Complex
Risk analysis	Only at the beginning	No	Yes	Yes
Flexibility	Rigid	Rigid	Flexible	Flexible
Cost control	Yes	Yes	Yes	Yes
Changes Incorporated	Difficult	Easy	Easy	Easy
Resource Control	Yes	Yes	Yes	Yes
Provide Work for Security	No	Yes	Yes	Yes
Operating in organized environment	Yes	Yes	Yes	Yes
Early Functionality	No	Yes	Yes	Yes
Provides Security Awareness	No	No	No	Yes
Complete set of requirements is not required at the onset	No	Yes	Yes	Yes
Dynamic reporting for risks and risk assessment	No	No	Yes	Yes
Allocation of resources and facilities	Yes	Yes	Yes	Yes

CONCLUSION

It has been found that the original water fall model is used by various big companies for their internal projects .Since the development team is familiar to the environment and it is feasible to specify all requirements of working environment. Spiral model is used for development of large, complicated and expensive projects like scientific Projects .Since spiral model approach enables the project term to address the highest risk at the lowest total cost. The main focus of security spiral is to identify security risks and managing those risks. A new model based on spiral model and its risk perception, has been proposed and named as Multilevel Security Spiral (MSS). In MSS, security spiral model is included within the spiral model right from the beginning of the software development life cycle. The software should be developed keeping in mind the secure software development activities and risks. After every step it must be revised by the programmer considering the risks. Secure code should written, code must be reviewed by the programmer himself considering the risks before implementation. So that the Multilevel Security Spiral (MSS) model should be risk free and highly secure.

REFERENCES

[1] Pankaj Vohra, Ashima Singh, "A Contrast and Comparison of Modern Software Process Models", International Conference on Advances in Management and Technology, 2013.

[2] Shams Tabrez Siddiqui, Hatem S. A. Hamatta, "Multilevel Security Spiral (MSS) Model: NOVEL Approach", International Journal of Computer Applications (0975– 8887) Volume 65– No.20, March 2013.

[3] Akhilesh," Assessment of software process models", International Journal of Engineering and Management Sciences", Vol.3 (4) 2012: 500-506.

[4] Daljit kaur, Parminder kaur and Hardeep Singh, "Secure Spiral: A Secure Software development Model", Journal of Software Engineering 6(1): 10-15, 2012.

[5] S.Thulasee Krishna,Dr. S.Sreekanth ,K.Perumal1, K.Rajesh Kumar Reddy," Explore 10 Different Types of Software Development Process Models" , S. Thulasee Krishna et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (4) , 2012.

[6] Nabil Mohammed Ali Munassar1 and A. Govardhan, "A Comparison between Five Models Of Software Engineering", IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 5, September 2010.

[7] Dahiya, D., and S. Dahiya. "Role of software process models in enterprise system development." Roedunet International Conference (RoEduNet), 2010 9th. IEEE, 2010.

[8] G.McGraw, "Managing Software Security Risks",IEEE Security & Privacy, Volume 2, Issue 2, Mar- Apr 2004, Page(s): 80-83.

PhysioBank, PhysioToolkit, and PhysioNet: A Research Resource for Intricate Physiological Signals

Manjit Singh
Dept. of ECE
Guru Nanak Dev University, Regional Campus, Jalandhar, India

Butta Singh
Dept. of ECE
Guru Nanak Dev University, Regional Campus, Jalandhar, India
bsl.khanna@gmail.com

Vijay Kumar Banga
Amrirtsar College of Engineering and Technology, Amritsar, India

Abstract- **PhysioNet is a web-based resource for Complex Physiologic Signals created under the support of the National Center for Research Resources of the National Institutes for the biomedical research community. Resources have 3 interdependent components, PhysioBank, PhysioToolkit, and PhysioNet PhysioBank is a large and growing collection of well-characterized digital recordings of physiological signals and related data for use by the biomedical researchers. PhysioNet is a public service of the Research Resource for multifaceted Physiologic Signals, a cooperative project initiated by researchers at Boston's Beth Israel Deaconess Medical Center/Harvard Medical School, Boston University, McGill University, and Massachusetts Institute of Technology (MIT). In addition to providing free electronic access to PhysioBank data and PhysioToolkit software via the world wide PhysioNet offer services and training via on-line tutorials to assist users with varying levels of expertise.**

Keywords- PhysioNet, PhysioBank, Electrocardiogram (ECG), Data base.

I. INTRODUCTION

In biomedical research the development of public databases and open source programs has been prolonged as public databases allow open access of data for developing and testing new computer models and algorithms. Open source programs share the source code of the computer models and algorithms, and permit users to copy, modify, and redistribute the source code [1-3].

One popular website is the PhysioNet which offers free access via the web to large collections of recorded physiologic signals and related open source software. The purpose of the PhysioNet is to provide research data for complex physiologic signals and to invite participation by the bio- medical community in a cooperative research activity [4].

II. PHYSIOTOOLKIT

PhysioToolkit is the open source software package, which includes the tools for a wide variety of classic signal-processing functions such as filtering, sampling frequency conversion, signal averaging, QRS detection, and power spectral density estimation. Other major component of PhysioToolkit includes the Implementations of novel algorithms for physiologic signal processing, simulation and modeling, and time-series analysis, such as applications for deriving a respiration signal from one or more ECG signals [5-8], and for detrended fluctuation analysis [9-12].

III. PHYSIOBANK

The PhysioBank is the collection of physiologic signals in digitised form which can be used by the research community in biomedical field.

A. *Massachusetts Institute of Technology/Beth Israel Hospital (MIT-BIH) Supraventricular Arrhythmia Database*

This contains 78 excerpts of two-lead ECG recordings, each 30-min long, selected to supplement the examples of supra- ventricular arrhythmias in the MIT-BIH Arrhythmia Database [13].

B. *Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) Database*

This database contains continuous recordings of multiple physiologic signals gathered from intensive care unit (ICU) monitors; it contains 72 complete records sampled at 500Hz. It also includes one or more blood pressure signals (e.g. arterial blood pressure, pulmonary arterial pressure, and central venous pres- sure) and respiration sampled at 125 Hz. The records also include measurements, computed by the ICU monitors at intervals of 1.024s of various parameters like heart rate, systolic and diastolic blood pressure and respiration rate. These measurements are now available for all 121 records of the database, including multiple recordings of some of the 90 subjects. The lengths of these records vary but average nearly 40 hours each, the total length is approximately 4658 hours [14].

C. *Apnea- ECG Database*

This database is consists of 70 ECG records, each is having eight hours length with accompanying sleep apnea annotations obtained from study of simultaneously recorded respiration signals which are included for eight of the recordings.

D. *Beth Israel Deaconess Medical Center (BIDMC) Congestive Heart Failure Database*

This contains 15 recordings, each about 20 hours long, from subjects with severe congestive heart failure [15].

E. MIT-BIH Arrhythmia Data- base

This database consists of 48 half-hour excerpt soft channel ambulatory ECG recordings, obtained from 47 subjects, Twenty-three recordings were chosen at random from a set of 4000 for 24-hours collected from a mixed population of inpatients (about 60%) and outpatients (about 40%), 25 recordings were selected from the same set to include less common but clinically significant arrhythmias that would not be well represented in a small random sample. In this category available records are 100-109, 111-119, 121-123, 124, 201-205, 207-210, 212-215, 217,219-223, 228, and 230-231 available in their entirety. The complete reference annotation files and the first 10min of the signal files for each of the remaining records are also available [15-17].

F. MIT-BIH Noise Stress Test Database

This consists of 12 half-hour records, consisting of two records from the MIT-BIH Arrhythmia Database to which calibrated amounts of noise have been added, for the purpose of assessing the noise tolerance of auto- mated analyzers. [18].

G. European ST-T Database

This standard database for evaluation of algorithms for detection of ST and T-wave changes consists of 90 two-hour recordings, with beat, rhythm, signal quality, and ST and T-wave change annotations. [19-20].

H. Post-Ictal Heart Rate Oscillations in Partial Epilepsy:

This consists of seven annotated single lead ECG recordings, with times of seizures indicated [21].

I. Fantasia Database

This consists of heart-beat time series from five young (two hours each) and five elderly subjects all healthy in sinus rhythm during a resting state [22].

J. Exaggerated Heart Rate

Oscillations during Two Meditation Techniques with additional data from spontaneously and metronomically breathing controls, and from highly trained athletes. These data include times of beat occurrence only [23].

K. Creighton University Ventricular Tachyarrhythmia Database

This database is having 35 records, each of which shows the onset of ventricular fibrillation. Most of these records are original digital recordings from patient monitors, each 8.5min long and contains one ECG signal and is collected by Floyd Nolle at the Creighton University Cardiac Center.

L. MIT-BIH Normal Sinus Rhythm Database

This database is 18 two-lead ECG recordings, each between 20 and 24 hours in length, from subjects without diagnosed cardiac abnormalities, with beat annotations. Also available are two recordings that were excluded from this database because of the presence of occasional ectopic beats [24-26].

M. MIT-BIH Atrial Fibrillation Database

This database consists of 25 records, each ten-hour long, containing about 300 episodes of atrial fibrillation and 40 episodes of atrial flutter. The ECG signals are currently available for only one of the records, the others are represented by annotation files only, from which RR-interval time series can be derived [27].

N. MIT-BIH ST Change Data-base

This contains 28 one and two-lead ECG recordings, mostly from exercise stress tests. [28].

O. MIT-BIH Polysomnographic Database

This is a collection over 80 hours of recordings of multiple physiologic signals during sleep [29].

P. MIT-BIH Malignant Ventricular Arrhythmia Database

This consists of 22 excerpts of two-channel ECG recordings, each 35 min in length, containing 89 episodes of ventricular tachycardia, 60 episodes of ventricular flutter, and 42 episodes of ventricular fibrillation. This database contains rhythm and signal quality annotations only [25-26].

Q. QT Database

This data- base has over 100 15-minute two-lead ECG recordings , with onset, peak, and end markers for P, QRS, T, and waves of 30-50 selected beats in each recording[30].

IV. PHYSIONET WEB SITE

All of the software used to support PhysioNet is freely available and open-source, just as are the PhysioToolkit components .The PhysioNet web site is a public service of the PhysioNet Resource, established as its mechanism for free and open dissemination and exchange of recorded biomedical signals and open-source software for analyzing them. The PhysioNet web site provides free electronic access to Physio Bank data and PhysioTool kit software, and facilities for cooperative analysis of data and evaluation of proposed new algorithms.

In addition, to this it offers service and training via on-line tutorials to assist users at entry as well as at advanced levels. PhysioNet tutorials provide introductions to the data and software available from Physio Bank and PhysioToolkit. There are many tutorials are available on various topics such as nonlinearity, heart rate variability, multifractal time series, multiscale entropy, physiologic signal processing. Also the tutorials are also available for understanding that how-to use and customizing Physio Tool kit software.

PhysioNet servers run the Apache web server under Linux, on 200 MHz or faster x86-compatible PCs, typically connected to the Internet with T1 (10 Mbps) or faster 3connections. During PhysioNet's first year, the PhysioNet servers at MIT providedover160 Gb of data to visitors from over 15,000 locations in 90 countries. PhysioNet's master web servers are located at MIT in Cambridge, Massachusetts. Mirrors are located in Italy, Israel, Spain, Slovenia, Taiwan, China, and elsewhere in the US [31]. To avoid overloading the master server, mirroring is performed using rsync [32].

V. CONCLUSION

PhysioNet invites the research community to contribute to PhysioNet, by participating in the ongoing review process, and by sharing common data sets and software in source form for peer review and possible inclusion in PhysioBank and PhysioToolkit. PhysioNet has received a number of contributions of physiologic signals, time series data, and soft- ware from published studies. PhysioNet alert the users to about the problems occurring in development of algorithms discovered and used in solving in nonlinear domain. Also permits to correct them when they are discovered, PhysioNet in particular offer an effective remedy for such problems and will also allow the developers and users of new or established software to let the community know promptly about updates or refinements.

REFERENCES

[1] http://www.physionet.org.

[2] W.-K. Chen, "Linear Networks and Systems", Brooks/code engineering division, pp. 123-135, 1993.

[3] H. Poor, An Introduction to Signal Detection and Estimation. New York: Springer-Verlag, 1985.

[4] A.L .Goldberger, L.A Amaral, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals", journal of American heart association, vol. 101, no. 23, pp. 215-20, 2000.

[5] G. O. Young, "Synthetic structure of industrial plastics", vol. 3, J. Peters, 2nd Ed. New York: McGraw-Hill, pp. 15-64. 1964.

[6] G.B. Moody, R.G. Mark, A. Zoccola, and S. Mantero,"Derivation of respiratory signals from multi-lead ECGs," Computers in Cardiology, vol.12, pp. 113-116, 1986.

[7] G.B. Moody, R.G. Mark, M.A. Bump et al., "Clinical validation of the ECG-derived respiration (EDR) technique," Computers in Cardiology, vol 13, pp. 507-510, 1986.

[8] 8. C-K Peng, S.V. Buldyrev, S. Havlin, "Onthemosaic organization of DNA sequences," Phys. Rev. E, vol. 49, pp.1685-1689, 1994

[9] C-K. Peng, S. Havlin, H.E. Stanley, and A.L. Goldberger, "Quantification of scaling exponents and crossover phenomena in nonstationary heart- beat time series," Chaos, vol.5, pp.82-87, 1995.

[10] M. Young, The Techincal Writers Handbook. Mill Valley, CA: University Science, 1989.

[11] J. U. Duncombe, "Infrared navigation-Part I: An assessment of feasibility (Periodical style),"IEEE Trans. Electron Devices, vol. 11, pp. 34-39, Jan. 1959.

[12] S.D. Greenwald, "Improved detection and classification of arrhythmias in noise-corrupted electrocardiograms using contextual information," Ph.D. thesis, Harvard-MIT Div. of Health Sciences and Technology, 1990.

[13] G.B. Moody and R.G. Mark, "A data base to support development and evaluation of intelligent intensive care monitoring", Computers in Cardiology, vol. 23, pp. 657-660, 1996.

[14] D.S. Baim, W.S. Colucci, E.S. Monrad et al., "Survival of patients with severe congestive heart failure treated with oral milrinone," Journal of American College of Cardiology, vol. 7, no. 3, pp. 661-670, Mar. 1986.

[15] R.G. Mark, P.S. Schluter, G.B. Moody et al., "An annotated ECG database for evaluating arrhythmia detectors", Frontiers of Engineering in Health Care, vol.29, no.8, pp. 205-210, 1982.

[16] G.B. Moody and R.G. Mark, "The MIT-BIH Arrhythmia Database on CD-ROM and software for use with it," Computers in Cardiology, vol.17, pp. 185-188,1990.

[17] G.B. Moody, R.G. Mark, "The Impact of the MIT-BIH Arrhythmia Database," IEEE Eng. Med. Biol. Mag., vol 20, no. 3, pp. 45-50, 2001.

[18] G.B. Moody, W.K. Muldrow, and R.G. Mark "A noise stress test for arrhythmia detectors, Computers in Cardiology, vol.11, pp.381-384, 1984.

[19] A. Taddei, A. Benassi, A. Biaginietal,"ST-T change analysis in ECG ambulatory monitoring: a European standard for performance evaluation," Computers in Cardiology vol. 14, pp. 63-68. 1987.

[20] A.Taddei, G.Distante, M.Emdinetal ," The European ST-T Database: standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography," European Heart journal, vol. 13, pp. 1164-1172, 1992.

[21] I.C. Al-Aweel, K.B. Krishnamurthy, J.M. Hausdorff, "Post-ictal heart rate oscillations in partial epilepsy" Neurology, vol. 53, no. 7, pp. 1590-1592, 1999.

[22] N. Iyengar,C-K. Peng, R. Morin, "Age-related alterations in the fractal scaling of cardiac inter- beat interval dynamics" American Journal of Physiology, vol. 271, pp.1078-1084, 1996.

[23] C.K. Peng, J.E. Mietus, Y. Liu , "Exag- gerated heart rate oscillations during two medita- tion techniques," International Journal of Cardiology, vol. 70, pp. 101-107, 1999.

[24] N. Iyengar,C-K. Peng, R.Morin , "Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics" American Journal of Physiology, vol. 271, pp. 1078-1084, 1996.

[25] S.D. Greenwald, P. Albrecht, G.B. Moody, and R.G. Mark, "Estimating confidence limits for arrhythmia detector

performance," Computers in Cardiology, IEEE Computer Society, vol.12, pp. 383-386, 1986.

[26] 26. S.D. Greenwald, "Development and analysis of a ventricular fibrillation detector," M.S. thesis, MIT Department of Electrical Engineering and Computer Science.1986.

[27] G.B.MoodyandR.G.Mark,"A new method for detecting atrial fibrillation using RR intervals," Computers in Cardiology, vol.10, pp. 227-230, 1984.

[28] P. Albrecht, "S-T segment characterization for long-term automated ECG analysis" M.S. thesis, MIT Department of Electrical Eng. and Computer Science, 1983.

[29] Y. Ichimaru, G.B. Moody, "Development of the polysomnographic database on CD-ROM," Psychiatry and Clinical Neuroscience, vol. 53, no. 2, pp. 175-177, 1999.

[30] P. Laguna, R.G. Mark, A.L. Goldberger, and G.B. Moody, "A database for evaluation of algorithms for measurement of QT and other wave- form intervals in the ECG", Computers in Cardiology, pp.673-676,1997.

[31] http://www.physionet.org/mirrors

[32] http://rsync.samba.org.

Track-1

Technical Session 4-(A & B)

IMAGE/SPEECH PROCESSING

A Systematic Way of Image Segmentation based on the Bacteria Foraging Optimization Algorithm Implementation and Analysis

Amrinder Singh
Dept. of Computer Science
Akal Degree College, Mastuana, Sangrur (Punjab, India)

Sonika Jindal
Dept. of CSE
Shaheed Bhagat Singh State Technical Campus, Ferozepur (Punjab, India)

***Abstract*--The goal of image segmentation is to cluster pixels into salient image regions, i.e., regions corresponding to individual surfaces, objects, or natural parts of objects. Segmentation could be used for object recognition, occlusion boundary estimation within motion or stereo systems, image compression, image editing, or image database look-up. Segmentations of simple gray-level images can provide useful information about the surfaces in the scene. A great variety of segmentation methods has been proposed in the past decades, and a disjunction categorization does not seem to be possible though, because even two very different segmentation approaches may share properties that defy singular categorization. In this paper we used BFO technique on image to find the optimum edges where the significant intensity change takes place. And for analyzing the optimum behavior of proposed work we are using the parameters like PSNR, MSE and BER.**

Keyword-Segmentation, Edge Detection, Clustering, Histogram, Thresholding.

I. INTRODUCTION

Image Segmentation is the process of partitioning a digital image into multiple regions or sets of pixels [1]. Actually, partitions are different objects in image which have the same texture or color. The result of image segmentation is a set of regions that collectively cover the entire image, or a set of contours extracted from the image. All of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristics. Edge detection is one of the most frequently used techniques in digital image processing [3]. The boundaries of object surfaces in a scene often lead to oriented localized changes in intensity of an image, called edges.

A. *Different categories for image segmentation*

The categorization presented in this section is therefore rather a categorization regarding the *emphasis* of an approach than a strict division.

The following categories are used:

i. Threshold based segmentation. Histogram thresholding and slicing techniques are used to segment the image. They may be applied directly to an image, but can also be combined with pre- and post-processing techniques.

ii. Edge based segmentation with this technique, detected edges in an image are assumed to represent object boundaries, and used to identify these objects.

iii. Region based segmentation. Where an edge based technique may attempt to find the object boundaries and then locate the object itself by filling them in, a region based technique takes the opposite approach, by (*e.g.*) starting in the middle of an object and then "growing" outward until it meets the object boundaries.

iv. Clustering techniques. Although clustering is sometimes used as a synonym for (agglomerative) segmentation techniques, we use it here to denote techniques that are primarily used in exploratory data analysis of high-dimensional measurement patterns. In this context, clustering methods attempt to group together patterns that are similar in some sense. This goal is very similar to what we are attempting to do when we segment an image, and indeed some clustering techniques can readily be applied for image segmentation.

v. Matching. When we know what an object we wish to identify in an image (approximately) looks like, we can use this knowledge to locate the object in an image. This approach to segmentation is called matching.

II. EDGE DETECTION FOR IMAGE SEGMENTATION

Edge detection techniques transform images to edge images benefiting from the changes of grey tones in the images. Edges are the sign of lack of continuity, and ending. Edges define the boundaries between regions in an image, which helps with segmentation and object recognition. It is a local property of an individual pixel and is calculated from the image function in a neighborhood of the pixel [Rakesh, 2004].

A. Edge properties

The edges extracted from a 2D image of a 3D scene can be classified as either viewpoint dependent or viewpoint independent. Basic properties of the three- dimensional objects, such as surface markings and surface shape are typically showed by viewpoint independent edge. A viewpoint dependent edge may change as the viewpoint changes, and typically reflects the geometry of the scene, such as objects occluding one another.

B. Purpose of Edge detection

Edge detection refers to process of identifying and locating sharp discontinuities in an image so Edge detection is an important stage in pattern recognition, vision-based inspection and image processing [Ratnam,2000], particularly used in the areas of feature detection and feature extraction. The purpose of edge detection is not only to extract the edges of the interested objects from an image, but also to lay the foundation for image fusion, shape extraction, image segmentation, image matching, and image tracking [Hildreth,1980]. The aim of edge detection is to provide a meaningful description of object boundaries in a scene from intensity surface i.e. to identify points of a digital image at which the intensity changes sharply. The reason behind these intensity changes may be either discontinuities in depth, surface orientation, lighting condition changes and many other factors. There are different sources for sharp changes in images which are created by structure (e.g. texture, occlusion) or illumination (e.g. shadows, highlights). Extracting edges from a still image is certainly the most significant stage of any computer vision algorithm requiring high accuracy of location in the presence of noise [Kurt,2008].

C. General steps in Edge detection

Mainly, three steps are involved in Edge detection

named as:

- Filtering
- Enhancement
- Detection

D. Edge Detection Methods

Three most frequently used edge detection methods are used for comparison. These are (1) Roberts Edge Detection, (2) Sobel Edge Detection and (3) Prewitt edge detection. The details of methods as follows,

i. The Roberts Detection

The Roberts Cross operator performs a simple, quick

to compute, 2-D spatial gradient measurement on an image. It thus highlights regions of high spatial frequency which often correspond to edges. In its most common usage, the input to the operator is a grayscale image, as is the output. Pixel values at each point in the output represent the estimated absolute magnitude of the spatial gradient of the input image at that point.

+1	0
0	-1

Gx

0	+1
-1	0

Gy

Fig. 1. Roberts Mask

ii. The Prewitt Detection

The Prewitt edge detector is an appropriate way to estimate the magnitude and orientation of an edge. Although differential gradient edge detection needs a rather time consuming calculation to estimate the orientation from the magnitudes in the x and y- directions, the compass edge detection obtains the orientation directly from the kernel with the maximum response. The Prewitt operator is limited to 8 possible orientations, however experience shows that most direct orientation estimates are not much more accurate. This gradient based edge detector is estimated in the 3x3 neighborhood for eight directions. All the eight convolution masks are calculated. One convolution mask is then selected, namely that with the largest module.

-1	+1	+1
-1	-2	+1
-1	+1	+1

0°

+1	+1	+1
-1	-2	+1
-1	-1	+1

45°

Fig. 2. Prewitt Mask

iii. The Sobel Detection

The Sobel operator performs a 2-D spatial gradient measurement on an image and so emphasizes regions of high spatial frequency that correspond to edges. Typically it is used to find the approximate absolute gradient magnitude at each point in an input grayscale image. In theory at least, the operator consists of a pair of 3x3 convolution kernels as shown in Figure 2.3. One kernel is simply the other rotated by 90^{o}. This is very similar to the Roberts Cross Operator. The convolution masks of the Sobel detector are given below:

-1	0	+1
-2	0	+2
-1	0	+1

Gx

+1	+2	+1
0	0	0
-1	-2	-1

Gy

Fig. 3. Sobel Mask

III. BACTERIA FORAGING OPTIMIZATION

Bacterial Foraging Optimization (BFO) is a population-based numerical optimization algorithm. Until date, BFO has been applied successfully to some engineering problems, such as optimal control [3], harmonic estimation transmission loss reduction [5] and machine learning [6]. The original Bacterial Foraging Optimization system consists of three principal mechanisms, namely, chemo taxis, reproduction, and elimination & dispersal. These are described as follows [2]

A. Chemotaxis

In the original BFO, a unit walk of the bacteria with random direction represents a "tumble" and a unit walk with the same direction in the last step indicates a "run."

B. Reproduction

The fitness value of each bacterium is calculated as the sum of the step fitness during its life and all bacteria are sorted in descending order according to health status. In the reproduction step, only the first half of population survives. The surviving population is divided into two identical ones, which are then placed in the same locations at which their parents were. Thus, the total population of bacteria remains constant [2].

C. Elimination and Dispersal

The chemotaxis provides a basis for searching the local best solution, and the reproduction process speeds up the convergence which has been simulated by the classical BFO. The bacteria with the best positions are kept and the remaining bacteria population is killed. The bacteria with best positions are then moved to another position within the environment [2].

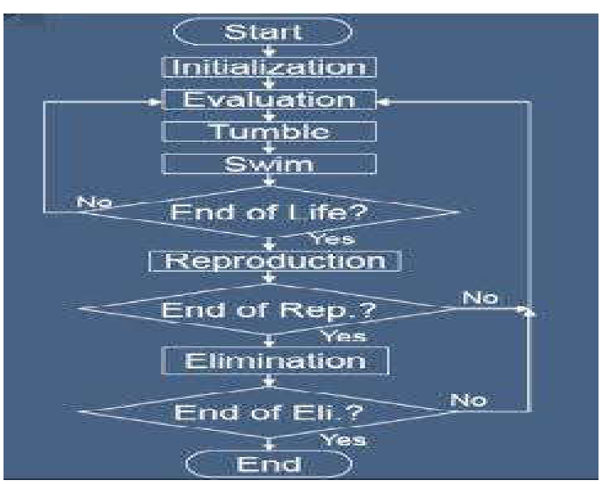


Fig.4. Flow chart for Bacterial Optimization Algorithm

IV. LITERATURE REVIEW

Nileshsingh V. Thakur., Pushpajit A. Khaire et al. [1] Edge detection is one of the most important techniques used for image segmentation. Image segmentation remains a puzzled problem even after four decades of research. In this paper, a soft computing approach based on fuzzy logic is applied on histogram of an image to enhance edge detection technique. We used BSD images for experimentation and their respective ground truths for qualitative evaluation of proposed approach.

Sunil Kumar, Vijay Kumar Sharma, Somya Saxena et al.[2] Edge detection is a primary operation of most of the image processing applications such as image detection, boundary detection, image classification, image registration. Edge detection filters out less important information and preserves the structural properties of image. The Proposed technique uses ANFIS edge detector for edge detection on digital images. It involves a neuro fuzzy system with the learning capability of neural network and the advantages of rule-based fuzzy system. This work follows hybrid algorithm to resolve the edge detection issues with the help of least square method and gradient descent method. Bhakwad K.M. et al. [3] in the paper entitled " Bacterial Foraging Optimization Technique Cascaded with Adaptive Filter to enhance Peak signal to Noise Ratio fron Single Image" proposed a new approach to enhance PSNR of highly corrupted image affected by impulse noise. The adaptive median filter is used to identify pixels affected by noise and replace them with median value to keep the information uncorrupted. The BFO technique minimizes error between adaptive median filter output image and noisy image to maintain an error percentage of

0.0001.

Zhang et al. [4] in the paper entitled "Ant Colony Optimization and Statistical Estimation Approach to Image Edge Detection" proved that the combination of ACO and statistical estimation in edge detection has been successfully. Clearer detection results are obtained and, especially, the proposed approach is robust to noise. In future research, ACO and more statistical features can be expected to have more excellent performances in edge detection and are hoped to be applied in a wider field of image processing.

Lakshmi et al. [5] in the paper entitled "A study of Edge Detection Techniques for Segmentation Computing Approaches" proposed that the purpose of this paper is to present a survey of various approaches for image segmentation based on edge detection techniques. In future, they planned to design a novel approach for edge detection and object recognition.

V. PROBLEM FORMULATION

Edge detection is a very important process in image processing and image analysis. Its importance arises from the fact that the edges contain most important information in image and reduce the dimensionality of data. Edge is a boundary of the target and the background. In images, edges are marked with discontinuities or significant variation in intensity or grey levels. Useful in applications such as: object detection, structure from motion, tracking and so on. There are various methods of edge detection like Sobel, Prewitt, Robert etc. In this paper, we have applied the BFO technique on image to find the optimum edges where the significant intensity change takes place.

VI. OBJECTIVE OF THE WORK

The Objective of the work include following points.

A. To study the various image segmentation Techniques.

B. To study the various edge detection methods for image segmentation.

C. To apply the BFO optimization algorithm for edge detection.

D. To implement the BFO algorithm for edge detection technique.

E. To apply the various parameters for analyze the behavior of the proposed approach.

VII. RESULT AND DISCUSSION

To validate the proposed BFO for Edge detection of an image, the image is firstly read and then converted to gray scale image as shown in fig 8.1.Apply BFO on the resulted image.

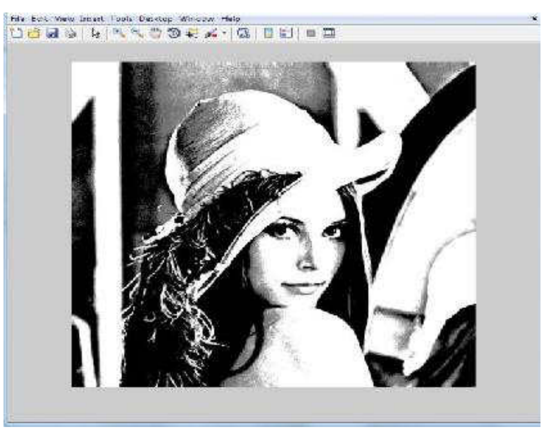


Fig. 5 Shows Image to be worked on

After working on all parameters and going through basic steps : chemotactic step , reproduction step and elimination& dispersal step ,BFO show the edges in image.

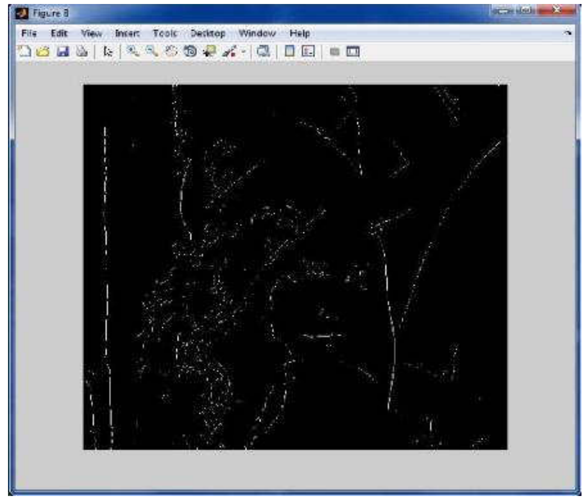


Fig 7.2: Edge detection using BFO

Edge detection using BFO do not contain the noise. I = 44 means out of 50 bacteria's, 44 number bacteria find the optimum result.

In this paper, we have compared the results of our work with Sobel, Prewitt and Roberts, in terms of bit error rate, mean square error and peak signal to noise ratio.

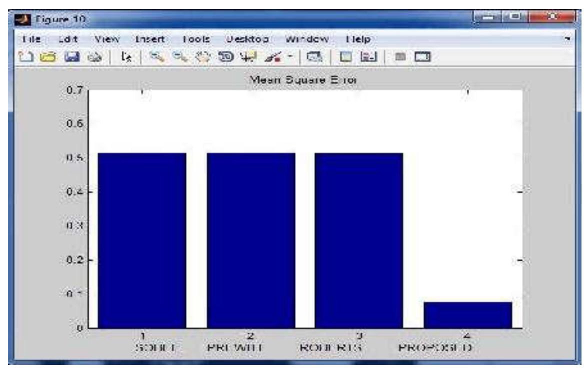


Fig. 6. Shows the result analysis of all techniques in terms of Mean Square Error

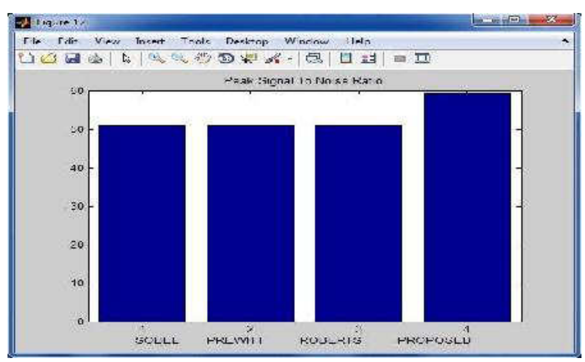


Fig. 7.Shows the result analysis of all techniques in terms of PSNR

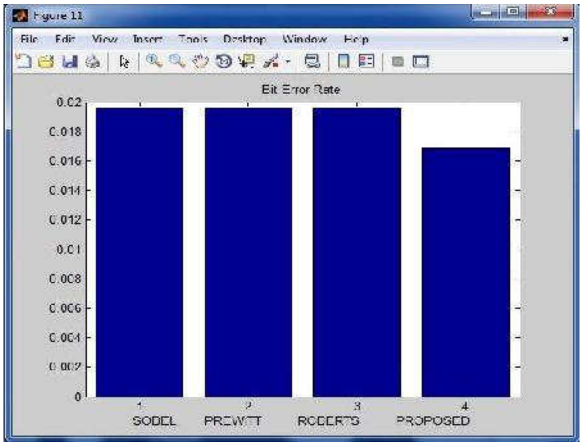


Fig.8 . Shows the result analysis of all techniques in terms of Bit Error Rate

The graph analysis shows the results of our technique in comparison to those obtained by applying Roberts, Sobel and Prewitt Edge detectors on the same image showing the higher value for Peak signal noise ratio (PSNR- measure the quality of reconstructed image, Higher the value of PSNR more optimized are the edges detected), least value for Mean square error (MSE- to quantify the difference between values implied by an estimator and the true values of the quantity being estimated) and least value for Bit Error rate (BER- to represent the ratio of the pixels affected by noise with all the pixels of the image) with BFO in comparison to Roberts, Sobel and Prewitt Edge detectors.

VIII. CONCLUSION

Edge detection is one of the key stages in image processing & object recognition. Sobel, Prewitt and Roberts edge detectors are also the most widely used detection algorithms due to their good performance. They produce thinned edges of the image .But for high resolution images, they give spurious branches corresponding to noise and texture of the image, unable to detect weak edges and distinguish the gray scale with little change, the detailed edge is uncontinuos. In this paper, we present a new optimization technique named as BFO to detect edges that results in significantly reduced memory, decreased latency &increased throughput with no loss in edge detection performance as compared to the original Sobel, Prewitt and Roberts edge detectors. BFO also results in high PSNR i.e. it is less sensitive to noise, yields high entropy value & low values for MSE & BER.

REFERENCES

[1] Nileshsingh V. Thakur., Pushpajit A. Khaire et al. (2012) , "Image Edge detection based on Soft Computing Approach" International Journal of Computer applications(0975-8887) Vol. 51 – No.8.

[2] Sunil Kumar, Vijay Kumar Sharma, Somya Saxena et al.(2013) ,"Edge detection using Soft Computing in MATLAB" International Journal of Advanced Research in Computer Science and Software Engineering(ISSN: 2277 128X) Vol. 3- Issue- 7.

[3] Xu, Q., Chakrabarti, C., and Karam, L.J., (2011), "A Distributed Canny Edge Detector And Its Implementation On FPGA" Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE), Sedona, pp. 500-505.

[4] Zhao, H., Qin, G., and Wang, X., (2010), "Improvement of Canny Algorithm Based on Pavement Edge Detection" 3rd International Congress on Image and Signal Processing. Yantai, pp. 964-967.

[5] Lakshmi, S., and Sankaranarayanan, S.V., (2010), "A study of Edge Detection Techniques for Segmentation Computing Approaches" IJCA Special Issue on "Computer Aided Soft Computing Techniques for Imaging and Biomedical Applications" CASCT.

[6] Kurt, B., and Gokmen, M., (2008), "Goal Oriented Edge Detection", Signal Processing, Communication and Applications Conference, pp. 1-4.

[7] Qiang, S., and Liu, L., (2009), "Compare between several Linear Image Edge Detection Algorithm", Second International Conference on Machine Vision.

[8] Shi, J., Zhu, Y., Chu, X., and Chen, X., (2010), "A Novel Approach of Edge Detection Based on Gray Weighted Absolute Correlation Degree and Prewitt Operator", Intelligent Computing and Integrated Systems (ICISS), Guilin, pp. 232-234.

[9] Passino, K. M., (2002), "Biomimicry of bacterial foraging for distributed optimization and control," Control Systems Magazine, vol. 22, no. 3, pp. 52-67.

Text Embedding Based on Nonsubsampled Contour LET Transform Using Steganographic Method

Rathod Rangu Vithal
Dept. of Digital Communication
Siddaganga Institute of Technology
Tumkur, India
rangu.v.rathod@gmail.com

K Viswanath
Dept. of Telecommunication Engineering
Siddaganga Institute of Technology
Tumkur, India
viswasit@gmail.com

Abstract—**Steganography, the science of invisible communication is one of the branches of information hiding. steganography, technique embeds secret data without making explicit modifications to the image. This paper presents a new data hiding technique for embedding text data in image using (NSCT). The proposed method provides less detectability and invisibility of information. We decompose image by using non sub sampled contour let transform and determine non-smooth regions. Embedding data in these regions cause less degradation in image quality. Contour let sub bands are divided into 3*3 blocks. Central coefficient of each block is considered for embedding of text data if they belong to edgy regions. The accuracy of the stego_image with respect to the original one was checked by evaluating PSNR, MSE and the correlation factor. Experiments show that this method achieves high embedding capacity and extract data without any loss of information while remains undetectable by using blind steganalysis technique.**

Keywords—Steganography, NSCT, Stego_Image, Cover Image.

I. INTRODUCTION

Steganography is an art of secret communication. Its main aim is to hide the data over an unsecured channel. Cryptography is “secret writting”whereas steganography is “covered writting”that hides the presence of message itself. Watermarking is the copyright protection. Data hiding can be achieved using various file formats such as image, text, audio, video etc. Image used for data embedding is known as cover image and the image with embedded data is stego_image. The amount of data that can hide without affecting the image quality is referred as embedding capacity. Embedding efficiency is defined in terms of expected number of embedded message bits per modified carrier bit. The importance of steganography lies in the fact that it hides the very existence of the secret it is protecting. High embedding capacity is the most important requirement in secure communication. An ideal image steganography system should satisfy following requirements: a) Maximum embedding rate b) Withstand in the presence of stegansiyzer c) Secure enough against various statistical attacks d) No severe visual artifacts. Steganalysis is art of detecting covert communication. Various forms of steganalysis include the detecting the presence of secret information and finding its approximate length or Even to extract the hidden Information from the stego_image. Image steganography can be classified into spatial domain, transform domain. In spatial domain[3]-[5] steganography, secrete data is embedded by altering the pixel values of cover image directly Inserting bit of information into pixel value of image using LSB technique we can reduces noise of the image. Drawback of LSB scheme is that it applicable for small amount of data and breaking this kind of embedding is easy. Drawback of spatial domain is that the size of secret data is less than cover data; the quality of stego_image is low. Whereas; transform domain technique hides data in transform coefficients. Hence transform domain approach makes the system more robust against various images processing operation such as rotation, compression, enhancement; this also improves security against stego_attacks. Here we have considered transform domain steganography. High robustness is achieved in transform domain. Many works in data hiding have been done in DCT and DWT. DCT [1]-[8] used in JPEG takes correlated input data and concentrates its energy in first few transform coefficients, providing better energy compaction. The drawback of DCT is that fewer coefficients are not sufficient to hide large data, robustness is low against compression. DWT and IWT used in JPEG2000[8]-[9] has many advantages over DCT such as better energy compaction without any blocking artifacts, DWT decomposes the image as L level dyadic wavelet pyramid and resultant wavelet coefficients can be easily scaled in resolution and perfect reconstruction of image can be achieved. But, for the images rich in directional details, DWT results in poor directional selectivity and lack of shift invariance capability. Contour let transform is the key to overcome these drawbacks. Advantages of contour let and NSCT transform over other transforms (wavelet & DCT) is that NSCT is the existence of linearly independent sub-bands. This issue decreases the possibility of detection of the stego_image by steganalysis algorithms [8]. The remainder of the paper is organized as follows. Section 2 provides review of contour let transform and NSCT. Aim of paper is defined in section 3. In section 4, data embedding and extraction algorithm is presented. Results and Discussions are provided in section 5&6 respectively. Conclusions are drawn in section 7.

II. CONTOUR LET TRANSFORM

The contourlet transform provides a multi scale and multidirectional representation of an image. It captures the smooth contour image. It consists of a DFB structure for obtaining sparse expansions for typical image having smooth contours. In this double directional filter, the laplacian pyramid (LP) is first used to capture the point discontinuities, and then followed by a directional filter bank to link point discontinuities into linear structures shown in Fig. 1. The required number of directions can be specified by the user. Since contour lets gives more edges, it is more suitable for data hiding applications as more data can be hidden in the high frequency regions without perceptually distorting the original image. The overall results in an image expansion using basic elements like contour segments, and, thus, are named Contour let [4], [6] and [13]. In this method, first cover image can be decomposed by contour let transform then we will get the contour let coefficients with larger magnitude corresponds to the edges are selected for embedding the secret data of the bits. It is flexible multi resolution, local and directional image expansion using contour segment. Some additional features like contour let possess high degree of directionality and anisotropy. Contour let based steganography hide the more information and provide higher protection against state of the art steganalyzer compare to previous steganography. Fig. 2 shows that decomposition of image in first level with four directions.

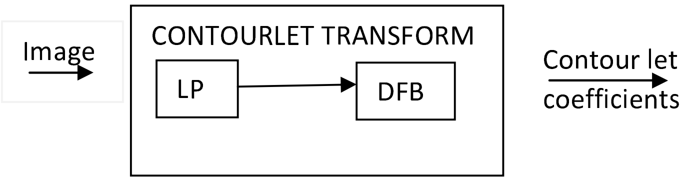
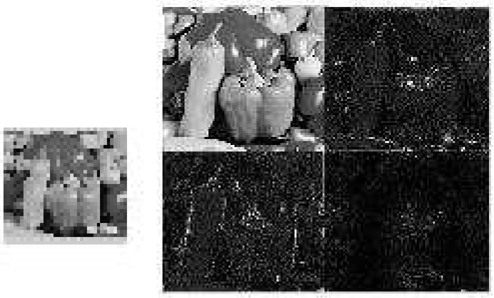


Fig.1. Contour let Transform



(a) (b)

Fig.2. The contour let transform in level one and four directions:
a) original image b) decomposed image

III. NON SUB SAMPLED CONTOURLET TRANSFORM

The non-subsample contourlet transform is a new image decomposition scheme introduced by Arthur L.Cunha, Jianping Zhou and Minh N.Do [10]-[11] NSCT is more effective in representing smooth contours in different directions of in an image than contourlet transform and discrete wavelet transform. The NSCT is fully shift invariant, Multi scale and multi direction expansion that has a fast implementation. The NSCT exhibits similar sub band decomposition as that of Contourlets, but without down samplers and up samplers in it. The NSCT is constructed by combining nonsubsampled pyramids and nonsubsampled directional filter bank as shown in Fig. 3.The non-subsamples pyramid structure results the multi scale property and nonsubsampled directional filter bank results the directional property.

IV. NONSUBSAMPLED PYRAMIDS

The nonsubsampled pyramid is completely different from the contourlet transform, the building block of the nonsubsampled filter bank as shown in fig. 4(a). A nonsubsampled filter bank has no down sampling or up sampling, and hence shifts invariant. The perfect reconstruction condition is given as

$$H_0(z).G_0(z) + H_1(z).G_1(z) = 1. \qquad (1)$$

Equation (1) shows that perfect reconstruction condition for critically sampled filter banks, and thus allows better filters to be designed.

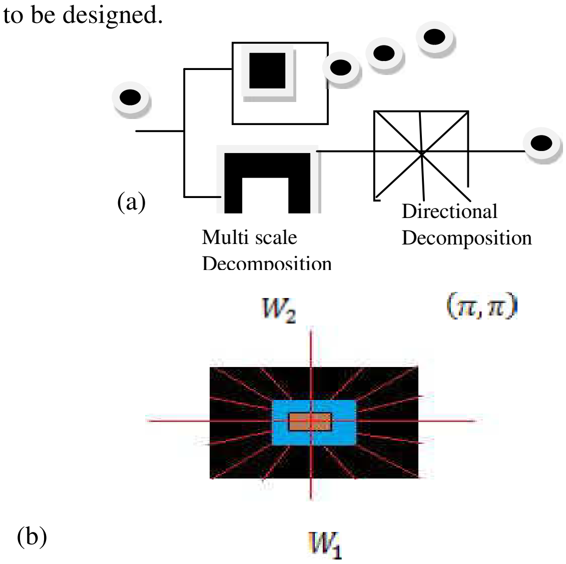


Fig. 3.The NSCT: (a) Nonsampled pyramid (b) frequency division of DFB.

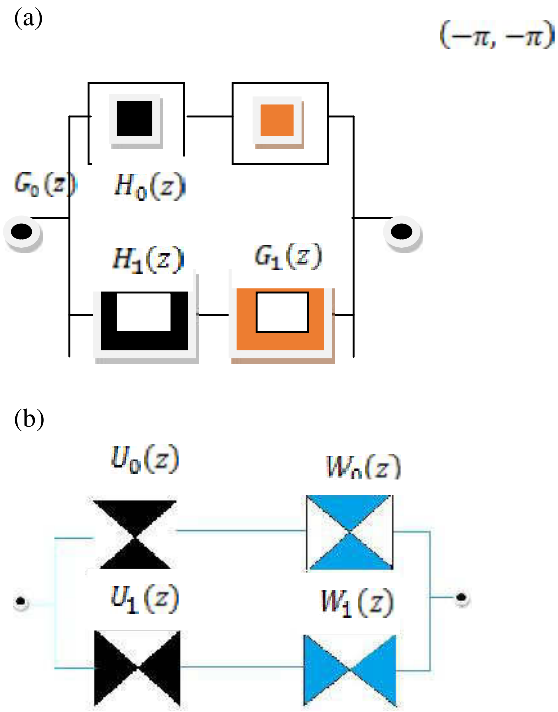


Fig. 4. Ideal frequency response of the building block of:
(a) nonsubsampled pyramid ;(b) nonsubsampled DFB.

V. NONSUBSAMPLED DIRECTIONAL FILTER BANKS

The nonsubsampled DFB is a shift invariant version of the critically sampled DFB in the contourlet transform. The building block of nonsubsampled DFB is also a two-channel non sampled filter bank. However, the ideal frequency response for a nonsubsampled DFB is different, as shown in Fig. 4(b). To obtain finer directional decomposition, we iterate nonsubsampled DFB's. For next level, we upsample all filters by a quincunx matrix given by

$$Q = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \qquad (2)$$

The frequency responses of two upsampled filters are given in Fig. 5.The analysis part of an iterated nonsubsampled di rectional filter bank is shown in Fig. 5(b).The frequency division of the directional filter bank is shown below Fig. 6(b).

A. AIM OF PAPER

In this paper a new technique for data hiding depend on dividing one of the contourlet coefficients to blocks and embed message bits in its center.

B. PROPOSED ALGORITHM

Fig (7), show the block diagram of proposed algorithm. The algorithm contain two phase the first one handle the embedding data inside image while the second phase cover the extract the embedded data as follows:

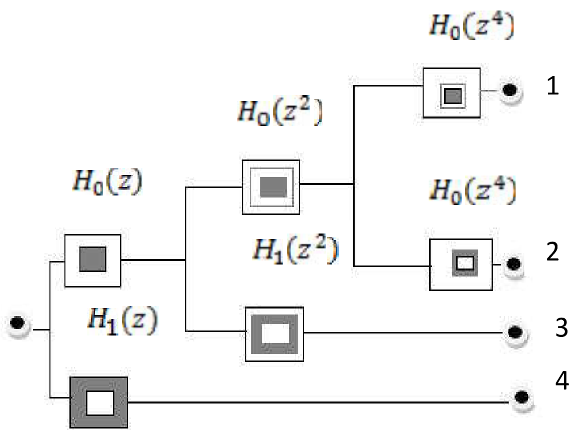


Fig.5.(a)Iteration of two-channel nonsampled filter banks in the analysis part of a nonsampled pyramid.

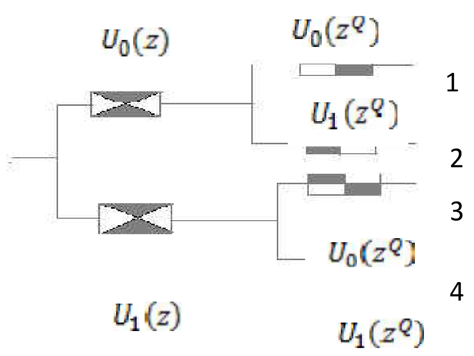


Fig. 5(b) The analysis part of an iterated nonsampled directional filter bank.

VI. EMBEDDING PHASE

The embedding process is done in the following steps at the sender side:

Step 1: The image selected is first decomposed using (NSCT). The result contains low pass coefficients and many high pass sub bands with same size .studied for the best coefficient suitable for embedding, high pass sub bands are chosen for embedding the data.

Step 2: The embedded message will be converted to ASCII code, so that it will be represented in binary format.

Step 3: embedding process steps:

a) The identified high pass sub band divided into blocks, and prepared that each block will hold one bit from the secret message.

b) Each bit (0 or 1) of the message to be embedded by adding its value to the value of block center.

m= d+ V_center (3) Where d is the bit to be embedded and V_center is the value of center of block .The process is continue until the whole data is embedded.

Step 4: Reconstruct the image from the processed coefficients to obtain stego_image.

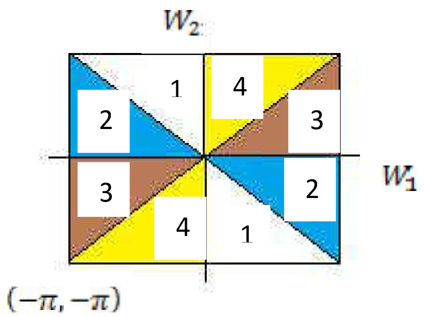


Fig.6. (b) Nonsubsampled DFB given in Fig.5.(b) .

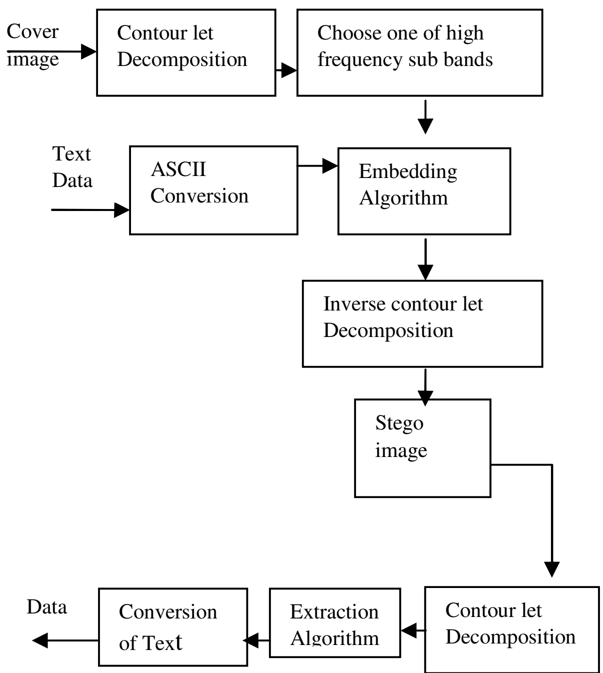


Fig. 7. Block diagram of proposed system

Step 5: Embed the coefficient level, block size and threshold value and coefficient level in stego_image using LSB method in its center.

VII. EXTRACTING PHASE

In Proposed system, steganography method is blind that mean the recipient must have the cover image to be able to extract embedded data. The embedding process is done in the following steps at the recipient side:

Step 1: extract the coefficient level, block size and threshold value from stego_image which were embedded using LSB method.

Step 2: The recipient received image (stego_image) and then, decomposed it using (NSCT).

Step 3: The cover image decomposed using (NSCT). Identified equivalent high frequency sub band S_cover;

Step 4: evaluate the different between S_cover and S_stego to get the difference matrix (D_matrix).

Step 5: divided (D_matrix) in step4 into blocks and then treat center of each block by comparing it with threshold value (V_center). The process is continued until the whole embedded data will extracted.

Step 6: the obtained binary sequence will be converted to ASCII code.

VIII. SIMULATION RESULTS

To evaluate the performance of the proposed embedding algorithm experiment was conducted using gray scale images with different size. Simulations were done using MATLAB. The image is composed to level one using NSCT. The result contains a low_ pass band and many high_ pass sub bands. One of the sub bands is chosen for data embedding related to its features. Proposed algorithm tested using different images and to embed data with those images and extract data without loss of information. Fig. 8(a) show the secret message (7225 char) that hiding in cover Image (pepper 512×512) shown in fig. 8(b) after converted it to binary sequence. Data hidden inside cover image known as Stego_image shown in fig 8(c). Table (1) and table (2)shows that decimal value of message and extracted message respectively.Table (3) shows that performance of cover image and stego_image is measured by using PSNR,MSE and Correlation coefficient.

We using contourlet a new data hiding technique for embedding text data in image..........……
…contourlet.

Fig.8. (a): The secret message

Correlation coefficient(R) between cover image and stego image can shows that how much the retrieval image close to the original one. From table3 we can analyze that inserting LSB of one bit to cover image, MSE can be reduced and increases the PSNR. If we insert all 8 bit to cover image, MSE can be increased and PSNR decreases.

TABLE 1: Decimal value of secret message

80	111	105	110
121	32	110	97
97	109	101	32
110	32	84	117

Fig.8. (b) cover image ;(c)stego_image

TABLE 2: Extraction of decimal value of secret message

91	111	107	110
127	38	110	103
107	111	103	41
110	39	85	127

TABLE 3: Performance of Stego Image and Cover Image

IMAGES	R(CORELLATION COEFFICIENTS)	LSB	PSNR(dB)	MSE
PEPPERS	1	1	62.58	0.144
	0.997	5	41.70	17.69
	0.776	8	20.96	2105
COINS	1	1	58.06	0.140
	0.996	5	35.95	22.85
	0.772	8	16.21	2149

IX. DISCUSSION

Image features like PSNR, MSE and Correlation are also analyzed between original image and retrieval one. PSNR stands for peak signal to noise ratio, which is a measure for image quality per pose. It is most easily defined using the mean squared error (MSE) which computed for two monochrome images i and j where one of them is considered as a noisy approximation of the other. Eq. (4) & (5) shows the MSE and PSNR evaluation.

$$PSNR = 10 \times \log\left(\frac{255^2}{MSE}\right) \quad (4)$$

$$MSE = \frac{1}{mn}\sum_{i=1}^{m}.\sum_{j=1}^{n}\left(\|i-j\|^2\right) \qquad (5)$$

$$R = \frac{\sum_{m}.\sum_{n}(A-\bar{A}).(B-\bar{B})}{\sqrt{\sum_{m}.\sum_{n}(A-\bar{A})^2\sum_{m}.\sum_{n}(B-\bar{B})^2}} \qquad (6)$$

Where $\overline{(A)}$ and $\overline{(B)}$ mean of the value in (A) and (B) computes the mean of the values in A and B respectively.

X. CONCLUSION

From the applied example on the proposed idea using the contourlet coefficient, shows that hiding data in the coefficients of the contourlet gives a robust technique for high security. Segmentation of the cover image in same block size achieved to reach high performance of data embedded, which goes to error free between the original message and the retrieved one when the cover image was segmented on 3×3 block size. In addition to the high performance on retrieving the message, the cover message was tested by evaluating the correlation factor, MSE and PSNR between the original cover and stego_cover.

REFERENCES

[1] Chin-Chen Chang, Tung-Shou Chen and Lou-Zo Chung , "A steganographic method based upon JPEG and quantization table modification", Intelligent Multimedia Computing and Networking ,Atlantic City, NJ, USA, Vol_ 141. Information Sciences 141 (2002) 123–138.

[2] Hedieh SAJEDI& Mansour JAMZAD, "Cont Steg: Contour let-Based Steganography Method ", Wireless Sensor Network, 3, 163-170 doi:10.4236/wsn.13022, published online (http://www.SciRP.org/ journal/wsn/).

[3] Sudeep P.V. & K.A. Navas &Sheeba V.S, " A Novel Data hiding Method in Spatial Domain", 10th National Conference on Technological Trends (NCTT09) IEEE (2008).

[4] Malini Mohan & Anurenjan P.R, "A New Algorithm for Data Hiding in Images using Contour let Transform", 978-1-4244-9477-4/11/$26.00 IEEE (2011).

[5] Dr. Diwedi Samidha and Dipesh Agrawa, "Random Image Steganography in Spatial Domain" 978-1-4673-5301-4/13/$ 31.00.IEEE (2013).

[6] Meaad M. Salih and Khalil Ibrahim AlSaif ," text embedding based on contour let transformation coefficients" International Journal of Information Technology and Business Management. 29th April 2013. Vol.12 No.1. ISSN 2304-0777 (2013).

[7] Mansi S. Subhedar and vijay H. Mankar,"Performance Evaluation of Image Steganography based on Cover Selection and Contour let Transform". International Conference on Cloud Technologies. 978-0-4799-2235-2/13 $26.00 © 2013 IEEE DOI 10.1109/CUBE.2013.39.

[8] Adel Almohammad Robert M. Hierons Gheorghita Ghinea," High Capacity Steganographic Method Based upon JPEG". The Third International Conference on Availability, Reliability and Security IEEE. 0-7695-3102-4/08 $25.00 © 2008 IEEE DOI 10.1109/ARES.2008.72

[9] Souvik Bhattacharyya, Avinash Prasad Kshitij, Gautom Sanyal," A Novel Approach to Develop a Secure Image based Steganographic Model using Integer Wavelet Transform" 2010 International Conference on Recent Trends in Information, IEEE. 978-0-7695-3975-1/10 $25.00 © 2010 IEEE DOI 10.1109/ITC.2010.68

[10] Arthur L.da Cunha, Jianping Zhou and Minh N. Do, "The Nonsubsampled Contour let Transform: Theory, Design, and Applications" 1057-7149/$20.00 © 2006 IEEE.

[11] Jianping Zhou, Arthur L. Cunha, and Minh N. Do," Nonsubsampled Contour let Transform: Construction and Application In Enhancement" This work was supported by the National Science Foundation under Grant CCR-0237633@2009 IEEE.

[12] Minh N. Do and Martin Vetterli, Fellow, IEEE, "The Contour let Transform: An Efficient Directional Multiresolution Image Representation" IEEE Transactions on image processing, vol. 14, no. 12, December 2005.

Pollination Based Optimization Algorithm in JPEG Images

Vikas Gupta
ECE, Punjab Technical University, Kapurthala, Punjab
vikas_gupta81@yahoo.co.in

Dhanvir Kaur
IT, Adesh Institute of Engineering & Technology, Faridkot, Punjab

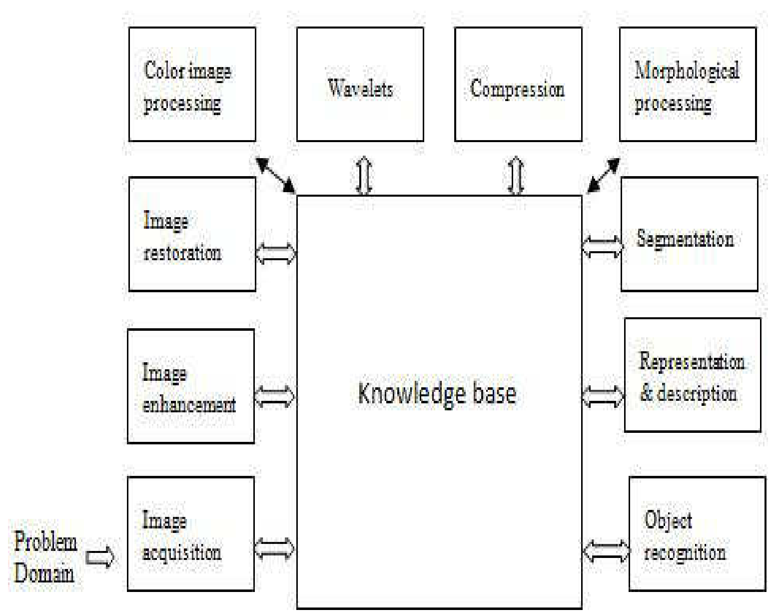
Rahul Malhotra
CT Institute of Technology & Research, Jalandhar

***Abstract*--JPEG is primarily a lossy method of compression. It used DCT which provides the high compressed image. PBO based JPEG compression scheme is used to compress the images. The use of this Algorithm reduces the size of images. Various parameters such as PSNR and EPI can be used for comparative study of the existing & the proposed techniques. The higher the PSNR value higher the quality of an image. The higher PSNR is obtained for compressed image by PBO based JPEG compression as compared to JPEG Compression.**

Keywords--JPEG, PBO, PSNR, EPI

I. INTRODUCTION

The theory of data compression becomes more and more significant for reducing the data redundancy to save more hardware space and transmission bandwidth. In computer science and information theory, data compression or source coding is the process of encoding information using fewer bits or other information-bearing units than a representation is not encoded. Compression is useful because it helps reduce the consumption of expensive resources such as hard disk space or transmission bandwidth. The basic goal of image data compression is to reduce the bit rate for transmission and storage while either maintaining the original qualit y or providing an acceptable fidelity. Uncompressed graphics, audio and video data require considerable storage capacity and transmission bandwidth. Despite rapid progress in mass storage, density, processor speeds, and digital communication system performance, demand for data storage capacity and data-transmission bandwidth continues to outstrip the capabilities of available technologies. Digital image processing is the use of computer algorithms to perform image processing on digital images. Interest in digital image processing methods stems from two principal application areas: improvement of pictorial information for human interpretation; and processing of image data for storage, transmission, and representation for autonomous machine perception. The steps used in image processing are shown in figure 1. The diagram does not imply that every process is applied to an image. Rather, the intention is to convey an idea of all the methodologies that can be applied to images for different purposes and possibly with different objectives. *The image acquisition* stage involves preprocessing, such as scaling. *Image enhancement* is the process of manipulating an image so that the result is more suitable than the original for a specific application. Basically, the idea behind enhancement techniques is to bring out detail that is obscured, or simply to highlight certain features of interest in an image. *Image restoration* is an area that deals with improving the appearance of an image. Color image processing is an area that has been gaining in importance because of the significant increase in the use of digital images over the Internet. Wavelets are the foundation for representing images in various degrees of resolution. *Wavelets* are the foundation for representing images in various degrees of resolution. Wavelet transforms are based on small waves, called wavelets, of varying frequency and limited duration



Morphological processing deals with tools for extracting image components that are useful in the representation and description of shape. *Segmentation* procedures partition an image into its constituent parts or objects. In general, autonomous segmentation is one of the most difficult tasks in digital image processing. A rugged segmentation procedure brings the process a long way toward successful solution of imaging problems that require objects to be identified individually. *Representation and description* almost always follow the output of a segmentation stage, which usually is raw pixel data, constituting either the boundary of a region or all the points in the region itself. *Description,* also called feature selection, deals with extracting attributes that result in some quantitative information of interest or are basic for differentiating one class of objects from another.

Recognition is the process that assigns a label to an object based on its descriptors. Knowledge about a problem domain is coded into an image processing system in the form of a *knowledge database.* Figure 2 shows the basic components comprising a typical general purpose system used for digital image processing. The function of each component is discussed in the following paragraphs, starting with image sensing. Image sensors required two elements to acquire digital images. The first is a physical device that is sensitive to the energy radiated by the object we wish to image. The second, called a digitizer, is a device for converting the output of the physical sensing device into digital form. Image processing hardware performs other primitive operations, such as arithmetic logic unit (ALU) that performs arithmetic and logical operations in parallel on entire images. The computer in an image processing system is a general purpose computer and can range from PC to a supercomputer. In dedicated applications, sometimes custom computers are used to achieve a required level of performance, but our interest here is on general purpose image processing systems. Software for image processing consists of specialized modules that perform specific tasks. A well designed package also includes the capability for the user to write code that, as a minimum, utilizes the specialized modules. Mass storage capability is a must in image processing applications. An image of size 1024×1024 pixels, in which the intensity of each pixel is an 8 bit quantity, requires one megabyte of storage space if the image is not compressed.

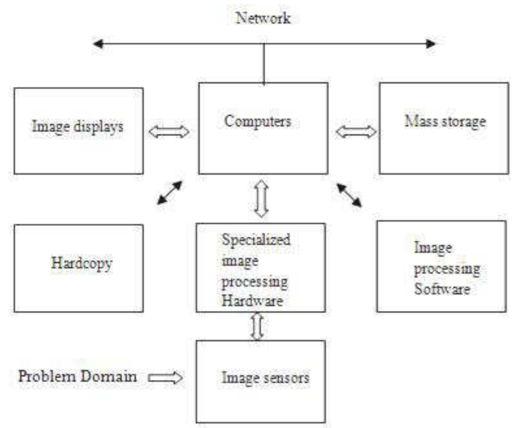


Fig. 2. Components of an Image processing System

Figure 2 illustrates the components involved in the image processing system. The figure defines how the image is displayed or stored using the components required for the purpose. Because of the large amount of data inherent of data inherent in image processing applications, the key consideration in image transmission is bandwidth. Image compression reduces the storage space required by an Image and the bandwidth needed when streaming that image across a network. The entire work of the paper has been categorized in a sequence of sections. Section 1 introduces the basic fundamentals of Image processing, Image compression and the techniques available for image compression. It also provides a view of the proposed algorithm to be implemented in this thesis research work. Section 2 provides a comprehensive research literature on the research carried out during the previous years. It opens a window to the grey areas where more work can be carried out to optimize the quality of image compression. Problem Formulation has been discussed in Section 3. The problem has been derived from the comprehensive literature discussed in Section 2. The research work implementation has been discussed in Section 4. This chapter provides information regarding the software and the technique used for carrying out the implementation of the research work. Section 5 provides the results or conclusions and finally the future scope.

II. LITERATURE SURVEY

Wide research has been carried out in the past on image processing. Hanhart (2013) described, "Subjective evaluation of HEVC intra coding for still image compression" demonstrates a significant improvement in compression efficiency of High Efficiency Video Coding (HEVC) as compared to H.264/MPEG-4 AVC, especially for video with resolution beyond HD, such as 4K UHDTV. One advantage of HEVC is the improved intra coding of video frames. Hence, it is natural to question how such intra coding compares to state of the art compression codecs for still images. This paper attempts to answer this question by providing a detailed analysis and performance comparison of HEVC intra coding with JPEG and JPEG 2000 (both 4:2:0 and 4:4:4 configurations) via a series of subjective and objective evaluations. The evaluation results demonstrate that HEVC intra coding outperforms standard codecs for still images with the average bit rate reduction ranging from 16% (compared to JPEG 2000 4:4:4) up to 43% (compared to JPEG). These findings imply that both still images and moving pictures can be efficiently compressed by the same coding algorithm with higher compression efficiency. Hurtik (2013) "Image Compression Methodology Based on Fuzzy Transform" the main objective is to develop an effective algorithm for image compression. They use both lossy and non-lossy compression to achieve best result. Their compression technique is based on the direct and inverse fuzzy transform (F-transform), which is modified to work with dynamical fuzzy partition. The essential features of the proposed algorithm are: extracting edges, automatic thresholding, and histogram adjustment. The article provides a comparison of our algorithm with the image compression algorithm (JPEG) and other existing algorithms based on fuzzy transform. Hagag (2013) corroborated in , "Multispectral image compression with band ordering and wavelet transforms" a new compression technique aiming at reducing the size of storage of multispectral images and maintaining at the same time the high-quality reconstruction is presented. An optimal multispectral band ordering process is applied before compression, and then, the dual-tree discrete wavelet transform is used in the spectral dimension, and the 2D discrete wavelet transform is used in the spatial dimensions. Finally, a simple Huffman coder is used for compression. Landsat ETM+ images are used for experimentations. Experimental results demonstrate that the proposed technique has better performance than JPEG, JPEG2000, SPIHT, and JPEG2000 with a 3D dual-tree transformation. Jeevan (2013) concluded in the paper, "Performance Comparison of DCT Based Image Compression on Hexagonal and Rectangular Sampling Grid" The advantages of processing images on hexagonal lattice are higher degree of circular symmetry, uniform connectivity, greater angular resolution, and a reduced need of storage and computation in image processing operations. In this work a comparison of DCT based Image compression on hexagonal and rectangular sampling grid is performed. DCT based image compression is performed on both rectangular domain and hexagonal domain using alternate pixel suppressal method.

III. PROBLEM FORMULATION

The research problem in this work is to study the JPEG image compression & to compare the results in terms of PSNR and EPI with the implementation of JPEG compression using Pollination Based Optimization (PBO). PBO based JPEG Compression is a new technique using which the results of compression are expected to be far better in comparison to the effective lossless Compression. The methodology focuses on the above objectives which are helpful in improving the compression parameters and are practically implemented using MATLAB 7.11.0 environment. In this methodology, Pollination based optimization algorithm is used to optimize the compression process. This algorithm provides better results as compared to previously implemented techniques.

IV. IMPLEMENTATION

As discussed in section 3, figure 3 shows the flow chart illustrating the sequence of steps involved in implementing the PBO Algorithm.

Step 1: Image Acquisition.

Step 2: Divide image into small segments through wavelets.

Step 3: Initialize PBO Parameters.

- a=1.2, A=0.9, D=1.2, N41.9, P=2,
- Number of Plants = 8,
- Number of weeks = 14,
- Number of seasons = 8, (number of iterations)
- Pollination weekly goal= [0.10 0.25 0.50 0.75 0.90 1.00]

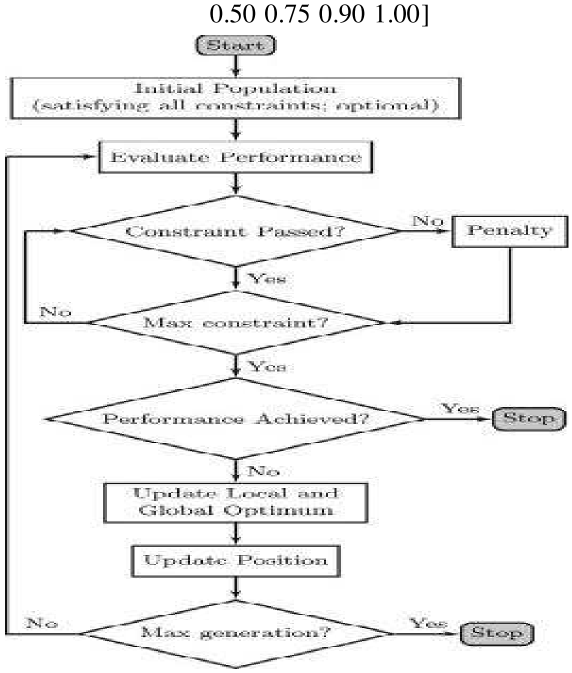


Fig. 3. Pollination based Optimization algorithm

Step 4: Randomly generate vectors.

- For season = 1 : number of seasons (iterations)
- For week = 1: number of weeks
- For k = 1: number of plants

Step 5: Evaluate Reproduction Vector:

$$R = \frac{(A \times D)}{(\alpha + A \times D)} + \frac{\left(\frac{a}{\alpha + A \times D}\right) \times N^P}{A^P + N^P} - C(N + D)$$

Step 6: Based on R, update number of seasons.

- Evaluate Error = Goal - R

Step 7: Based upon error update N, D, A

Step 8: Exit, if Error acceptable.

The following Figure 4 shows the flow chart of how PBO is used in the process of Image Compression. It entails all the steps from Image Acquisition up to mage Compression.

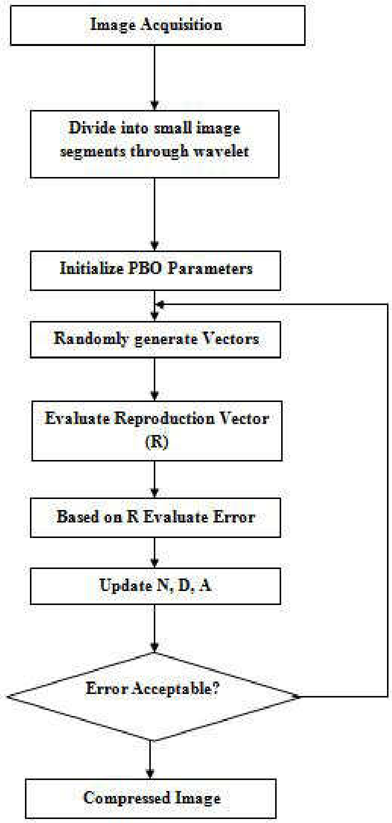


Fig. 4.Flow Chart of Image Compression using PBO

In this proposed work, a sample was gray scaled and converted into size (128×128). Then divide image into small segments using wavelets.

Table I. Comparative study between the JPEG Compression & PBO based JPEG Compression

Technique Used	PSNR	EPI
JPEG Compression	26.0047	7.74851
PBO Based JPEG Compression	28.0133	7.75742

Fig.5(a). Original Image

Fig.5 (b).Compressed Image

Above figures 5(a) & 5(b) show the original image & compressed image using PBO based JPEG compression technique.

Comparative study between the JPEG Compression & PBO based JPEG Compression for Image sample shown in Figure 5(a) has been shown in Table 1 as follows.

V. CONCLUSION & FUTURE SCOPE

The implementation of the PBO based JPEG obtains the higher PSNR value. The higher the PSNR value, higher the quality of an image. Higher PSNR has been obtained for compressed image by PBO based JPEG compression as compared to JPEG Compression. This shows that the PBO based JPEG Compression is better than JPEG technique for image compression.

In the present work we have implemented the PBO based JPEG compression quite successfully. Still there is some hope of improvement. If we can use more than two optimization algorithms together with more image samples for the compression, the results could have been better as compare to single optimization algorithm. So, future work could go on the direction of hybrid systems using more than one algorithm together.

REFERENCES

[1] Talukder, H, K., and Harada, K., "Haar Wavelet Based Approach for Image Compression and Quality Assessment of Compressed Image," International Journal of Applied Mathematics, vol.36, pp.1-8, 2007.

[2] Amar, Ben, C., and Jemai, O., "Wavelet Networks Approach for Image Compression," ICGST International Journal on Graphics, Vision and Image Processing, pp. 37-45, 2007.

[3] A, Khashman., and K, Dimililer., "Image Compression using Neural Networks and Haar Wavelet," WSEAS Trans Signal Processing vol.4, pp. 330-339, 2008.

[4] Senthilkumaran, N., and Suguna, J., "Neural Network Techniques for Lossless Image Compression using X-ray Images," International Journal of Computer and Electrical Engineering, vol.3, no.1, 2011.

[5] Kuther, Abood., Hayder., Aboud and Falih, H. A., "X-ray image compression using neural network," ISSN 2229-5518,vol.3, Issue 10, 2012.

[6] M, Sindhu., and R, Rajkamal., "Images and Its Compression Techniques," International Journal of Recent Trends in Engineering, vol.2, no. 4, 2009.

[7] Bhardwaj, Anuj., and Ali, Rashid., "Image Compression Using Modified Fast Haar Wavelet Transform," ISSN 1818-4952, vol.7, no.5, 2009.

[8] S, K, Ajay., Tiwari, Shamik., and Shukla, p., "Wavelet based Multi Class image classification using Neural Network," International Journal of Computer Applications, vol.37, no.4, 2012.

[9] Mhd, Nagaria, Baluram., Hashmi, Farukh., and Dhakad, Pradeep., "Comparative Analysis of Fast Transform for Image Compression for optimal Image Quality and Higher Compression Ratio," International Journal of Engineering Science and Technology (IJEST), vol.3, no.5, 2011.

[10] Sivanandam, N, S., Sumathi, S., and Deepa, N, S., "Introduction to Neural Networks using Matlab 6.0," McGraw Hill, 2012.

[11] Gonzalez. R., and woods, R., "Digital image processing" 3rd edition, Pearson education, 2011.

[12] Sonal., Kumar Dinesh., "A Study of Various Image Compression Technique," National Conference on Challenges & Opportunities in Information Technology, 2007.

[13] Hanhart, Philippe., Rerabek, Martin., and Korshunov, Pavel., "Subjective evaluation of HEVC intra coding for still image compression," Seventh International Workshop on Video Processing and Quality Metrics for Consumer Electronics - VPQM 2013, Scottsdale, Arizona, USA, January 30 - February 1, 2013.

[14] Hurtik, Petr., and Perfilieva, Irina., "Image Compression Methodology Based on Fuzzy Transform," International Joint Conference CISIS'12-ICEUTE´12- SOCO´12 Special Sessions, pp 525-532, 2013.

[15] Hugag, Ahmed., Amin, Mohamed., Fathi, E., Abd El-Samie., "Multispectral image compression with band ordering and wavelet transforms,"Signal, Image and

Video Processing, pp. 1863-1703, 2013.

[16] Mani, K, Jeevan., and A, Ravindran, Rahul., "Performance Comparison of DCT Based Image Compression on Hexagonal and Rectangular Sampling Grid," Journal of Image and Graphics, Vol. 1, No. 3, September 2013.

[17] Groach, Manik., and Garg, Amit, Dr., "DCSPIHT: Image Compression Algorithm," International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622, Vol. 2, Issue 2, Mar-Apr 2012, pp.560-567, 2012.

[18] Xu, Peng, Fei., Zhang, Bin, Hong., Wang, Feng, Xin., and Yu, Yong, Zheng., "Color Image Compression Using Block Singular Value Decomposition," Applied Mechanics and Materials, (Volumes 303 - 306), pp. 2122-2125, February, 2013.

Watered Area Detection from Satellite Image using Image Processing

Deep Singh Bedi
Department of Electronics Technology,
GNDU, Amritsar
d4bedi@gmail.com

Jaspreet Kaur
Department of Electronics Technology,
GNDU, Amritsar
gill_jas8@hotmail.com

Amandeep Singh
Department of Electronics Technology,
GNDU, Amritsar
amandeepsliet@gmail.com

***Abstract*–This paper deals with an intelligible method for segmentation of the watered area around the Indian peninsula. However, it can be applied to any satellite image displaying watered area and land area in proximity and also for its calculation. In the present day scenario, one of the major areas of environmental concern, is the climate change associated with sea level rise (SLR). The brunt of SLR on the coastal zone of any area can be estimated using this simple method of water detection and area calculation. Various segmentation techniques are applied for highlighting different areas of concern which are comparatively difficult to apply and include difficult operations and algorithms than the proposed one. This paper speaks about a very easy approach to highlight the water stretch alongwith a good estimate of the area calculation.**

Keywords*–*Sea level rise (SLR), Segmentation, Area calculation, Satellite image, watered area.

I. INTRODUCTION

Segmentation is the process of highlighting a particular area or a region of interest in a digital image. In segmentation an image is divided into regions that are homogeneous according to a given criteria. Various segmentation techniques are applied to highlight particular area of concern which include Edge Detection, Threshold, Histogram, Region based methods and Watershed Transformation [1]. Segmentation methods for color images and gray scale images are different, e.g., content based image retrieval [2]. An easy conversion whereas can be applied using MATLAB to convert a color image into a gray scale image using a single function. Segmentation techniques are used in various fields and is an important aspect of medical sciences, for example in diagnosing brain disease[3].Different threshold techniques can also be employed for image segmentation [4]. The segmentation of watered area is dealt with in this paper around the Indian peninsula in a very simple way. The SLR is an important matter of concern nationwide. The mean SLR trends along the Indian coasts are about 1.30 mm/year. It is estimated that between 1990 and 2100 the sea level could rise by 3.5 to 34.6 inches and would result to inundating valuable land and coastal communities and endangering wetlands. Khambat and Kutch in Gujarat, south Kerala and Mumbai and parts of the Konkan coast are the most vulnerable stretches along the western Indian coast (The report - India's Second National Communication to the United Nations Framework Convention on Climate Change). The climatic conditions are predicted using the mean SLR value and the possible increment in the frequency and intensity of coastal surges and cyclones. These threats cause unsustainable damage to life and property. They threaten life of people in vicinity to the coastal regions and thus, people migrate to other parts of the country. About one fourth of the Indian population living along the coastline is afflicted. The use of the modern technology has made it accessible to know the variations and the impact of such outcomes that cause severe threats, by using Satellite images. Satellite images have many applications in meteorology, agriculture, forestry, landscape, education and warfare. At a particular elevation from ground and particular spot, imaging can be done using satellite. Satellite imaging is also used in fields such as seismology and oceanography. Thus, it plays a crucial and indispensable role. Image pixel plays a significant part in studying the vast area of imaging and MATLAB is a very great platform to for image processing. The word pixel is formed by the synthesis of the two words: “pix” meaning picture and “el” standing for element. A pixel is considered to be the smallest single component of a digital image. It can be used as an element of measure. The number of pixels in an image is known as resolution. The more the resolution, the more accurate results can be drawn from the digital data encrypted in the image. The intensity of the image is also a major contributing factor.

II. LITERATURE REVIEW

Various methods are available related to image segmentation. A boundless literature is available and is tough to understand and analyze different segmentation techniques. Segmentation is a method that is very profitable in the field of medical sciences to diagnose the abnormalities related to images [6, 7]. The clustering methods have been discussed for medical image segmentation in particular for MR Images of brain and are successful in combining fuzzy c means and k-means to get novel fuzzy-k means algorithm [8]. Various methods are discussed in the field of image segmentation [1,9] and various surveys are conducted [6, 7, 10].

The area calculation alongwith the segmentation technique are also present which use different applications and different fields with more complex criteria for segmentation and area calculation [11,12].Convoluted methods are present in the extensive literature and include various arduous calculations and bulky algorithms.

A screen shot is shown in figure 2 which displays all the functions.

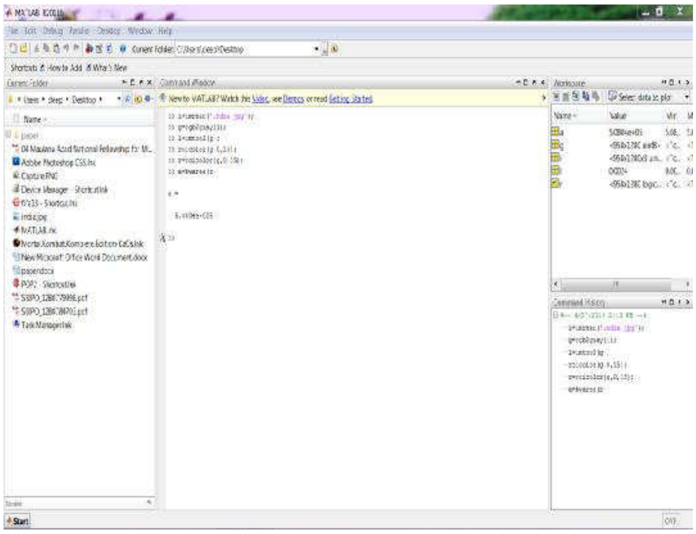


Fig 2. Desktop View of MATLAB

III. ETHODOLOGY AND PLANNING

A complete flow chart of the various steps involved in the proposal has been given below:

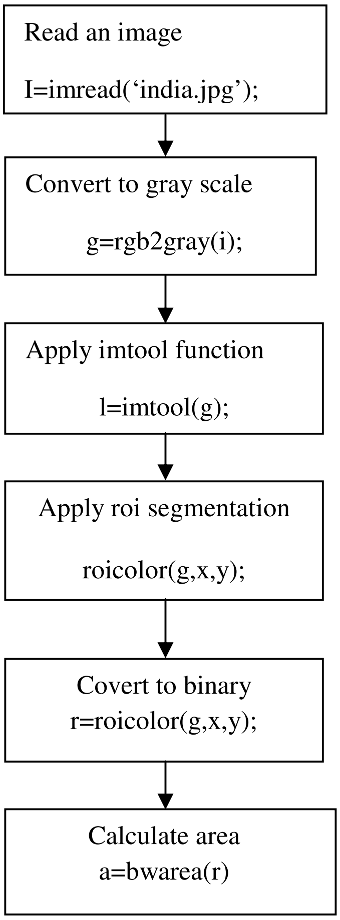


Fig 1. Flow Chart: A complete Process

A brief description can be given as:

1. Acquisition of a satellite image of the Indian peninsula
2. Conversion of the image into gray scale
3. Process the image in MATLAB platform and note the intensity value corresponding to the region of interest and also make a rough interpretation of the intensity value related to the other area leaving the watered area
4. Apply the roicolor operation to separate the particular intensity range as selected
5. Obtain the image in binary form
6. Calculate the area of the region of interest

After the application of all the steps including the segmentation of watered area of a satellite image and its calculation, a table is devised with different pixel values (All steps of algorithm are implemented on MATLAB 7.13.0.564(R2011b) [5] platform).

IV. SIMULATION AND RESULT

The same algorithm was applied on different images separating land and watered portion and was found effective to segment it from the rest of the area. The result of the calculation of the watered area also gives one of the simplest method of application with a single function use.

A. *Satellite image*

A satellite image can be obtained from the internet very easily. Satellite images with different pixel values and resolutions may be available. Selecting a good quality image depends on such factors. The satellite image of concern was downloaded from the internet. The image resolution of the used image is 1280×958, which means width is 1280 pixels and height is 958 pixels. The image is as shown below:

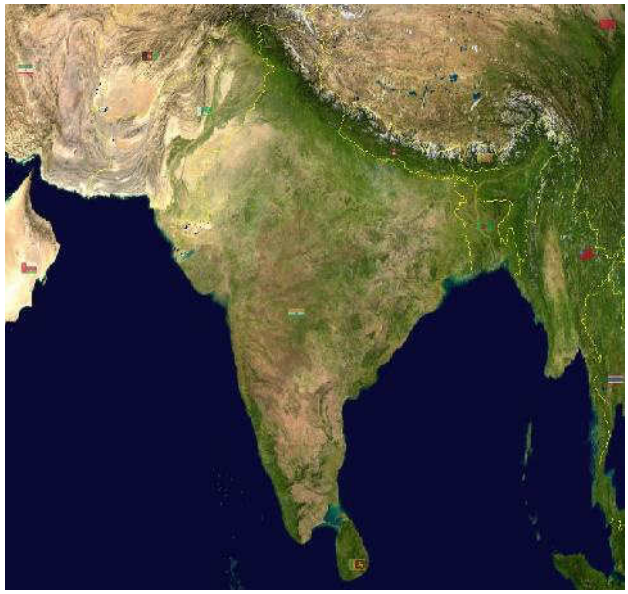


Fig 3. Satellite Image

i. *Converting image to gray scale*

The above satellite image is read in the MATLABand isthen converted to a gray scale image for further operations and processing. The gray scale output is a shown below:

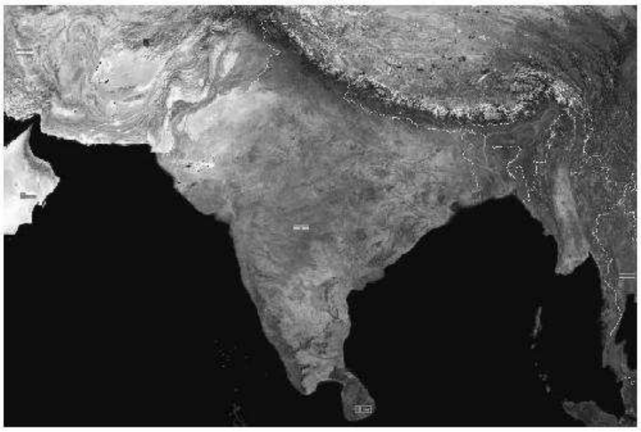


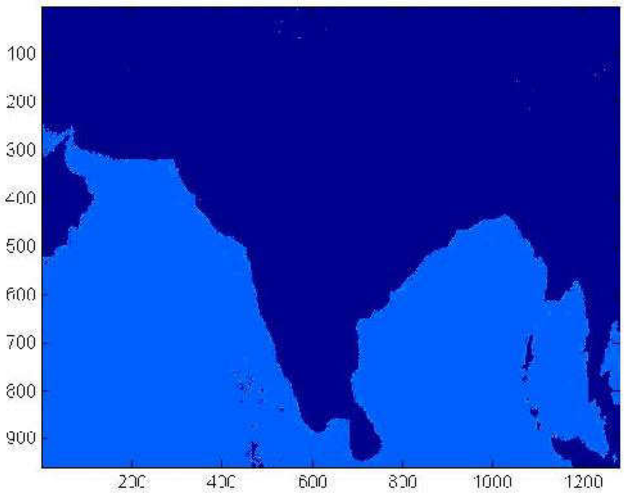
Fig 4. Satellite Image converted into Grey Mode

a. *Applying ROI and IMTOOL map*

The imtool function is applied on the converted gray scale image and the area of the pixel corresponding to the watered area is searched by moving the cursor on the water area.The pixel information as shown by imtool function on the imtool screen is on the left bottom corner of the image toolbox displayed. It shows the following

Pixel info: (X, Y) intensity

This is the main area of concern for processing of image. On moving, the cursor on the image, the X and Y display the position of the cursor and the intensity corresponds to the pixel value. The intensity corresponding to the water area is noted as displayed. In the above image, the intensity of water portion corresponds to 14. The intensity of the whole image is judged for reference by moving the cursor. In the above gray scale image all the intensity leaving the water portion is approximately above 40. Therefore reference is drawn for further operation to be done.



The roicolor operation is performed in the MATLAB which uses the range of intensity value to be highlighted in the image. The range is selected from x to y for obtaining a good boundary separation result and a sample image within this range is as shown in figure 5.

The light blue area segmented, is the watered area as can be clearly observed.

This roicolor image is automatically converted by the system into a binary image when a symbolic name is given to this image. The binary converted form is as shown in figure 6.

In the above shown image the white area is the region of concern that is the watered area.

b. *Quality Check:*

If we go for the quality measures of the proposed algorithm then we have to calculate various parameters like PSNR, MSE, L2RAT etc. The have been calculated for various threshold values and the obtained results are tabulated below:

From the above table it is very clear that all other parameters

are same but L2RAT gives best values corresponding to threshold value of 15. Same has been used in the processing of the satellite image.

Table I. Quality Deciding Parameters

S. no.	Threshold Value	PSNR	MSE	L2RAT
1	5	8.0287	1.0238e+004	5.9742e-009
2	10	8.0288	1.0238e+004	6.5238e-008
3	15	8.0335	1.0227e+004	4.0412e-005
4	20	8.0336	1.0226e+004	4.1056e-005

c. The image shown above is in binary form. The white pixel areas are known as the “on” pixel and the black pixel areas as “off” pixel. The area corresponding to the “on” area is calculated in MATLAB using a simple function which estimates thewatered area. The bwarea is the function used to estimate the area andis shown in table no. 1.

Table II. Bwarea

S.no	Image pixel least value (x)	Image pixel max value (y)	Area output
1.	0	15	5.0804e+005
2.	0	16	5.1130e+005
3.	0	17	5.1305e+005
4.	0	18	5.1413e+005
5.	0	19	5.1497e+005
6.	0	20	5.1575e+005

Average Computed Area = 5.1287e+005

B. CHALLENGES

The major challenge is related withobtaining the yearly satellite images if we are concerned with the SLR. Obtaining digital images of the same resolution, quality, and texture and with the same altitude while covering exactly same area can create issues as far as SLR is concerned.Study of SLR requires yearly review of satellite images or survey of coastal areasmore often and collection of data and its interpretation. Using this algorithm only an estimated area can be calculated which could account only for the SLR rather than aparticular value of increment.

C. CONCLUSION & FUTURE SCOPE

This work describes digital image processing technique to segment a watered area and calculate it. A very simple, efficient and less bulkier algorithmic approach is applied to its segmentation. The algorithm provides easy to apply approach. Very less number of steps are involved for the segmentation of the water body surrounding Indian peninsula. The calculation of areais also added by the use of a single function. However the all above is not perfect enough to get the 100% accuracy. So a lot of work can be done in the same such as using adaptive thresholding etc. to get superior quality output.

ACKNOWLEDGEMENT

The authors acknowledge the Department of Electronics Technology for the continuous support for the work by providing the excellent work environment and MATLAB Platform without which this work might not be a success.

REFERENCES

[1] Muhammad Waseem Khan "*A Survey:ImageSegmentation Techniques*"International Journal of Future Computer and Communication*, Vol. 3, No. 2, April 2014*

[2] M. Rehman, M. Iqbal, M. Sharif, and M. Raza, "*Contentbased image retrieval: survey*," World Applied SciencesJournal, vol. 19, pp. 404-412, 2012.

[3] M. Yasmin, M. Sharif, S. Masood, M. Raza, and S. Mohsin, "*Brain image reconstruction: A short survey*" World Applied Sciences Journal,vol. 19, pp. 52-62, 2012.

[4] Salem Saleh Al-amri, N.V. Kalyankar and Khamitkar S.D "*Image Segmentation by Using Thershold Techniques",* journal of computing, volume 2, issue 5, may 2010, issn 2151-9617

[5] Release notes for MATLAB R2011b, "*Summary of New features release notes for MATLAB R2011b"* Software Lab Manuals.

[6] Dr.S.V.Kasmir Raja, A.Shaik Abdul Khadir And Dr.S.S.Riaz Ahamed, "*moving toward region-based image segmentation techniques: a study*", Journal of Theoretical and Applied Information Technology ,vol. 5,no. 1,page no. 81-87.

[7] Dzung L. Pham, Chenyang Xu, and Jerry L. Prince, "*currentmethods inmedical image segmentation*", Annu. Rev. Biomed. Eng. 2000. Vol. 02 page no. 315–37.

[8] Ajala Funmilola A , Oke O.A, Adedeji T.O, Alade O.M, Adewusi E.A. "*Fuzzy k-c-means Clustering Algorithm for Medical Image Segmentation*" Journal of Information Engineering and Applications ISSN 2224-5782 Vol 2, No.6, 2012.

[9] A. M. Khan, Ravi.S" *Image Segmentation Methods: AComparative Study*"International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-4, September 2013.

[10] Nikhil R Pal and Sankar K Pal, " *A Review on ImageSegmentation Techniques*", Pattern Recognition, Vol 26, No. 29, 1993.

[11] Sanjay B. Patil,Dr. Shrikant K. Bodhe,"*BetelLeafArea Measurement Using Image Processing*", International Journal on Computer Science and Engineering (IJCSE), ISSN : 0975-3397, Vol. 3 No. 7 ,page no.2656-2660,July 2011.

[12] J.selvakumar, A.Lakshmi, T.Arivoli, "*BrainTumorSegmentation and Its Area Calculation in Brain MR Images using K-Mean Clustering and Fuzzy C-Mean Algorithm*", IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012) March 30, 31, 2012.

Hybridized VQ, HMM Technique for Telugu Isolated Word Recognition

P. Prithvi
Dept. of ECE
NIT
Warangal, India
prithvi@nitw.ac.in

D. Aditya Kumar
Dept. of ECE
NIT
Warangal, India
adityakumar12121@gmail.com

D. Medha
Dept. of ECE
NIT
Warangal, India
medhanitw14@gmail.com

T. Kishore Kumar
Dept. of ECE
NIT
Warangal, India
tkishore@nitw.ac.in

***Abstract*—In this paper we implemented telugu speech recognition by hybridizing VQ and HMM techniques in MATLAB. Telugu digits for 1 to 5 are the words of interest for recognition. Speech samples are collected for five telugu digits-okati, rendu, moodu, nalugu, aidu from 40 different speakers in a clean environment. Front end processing and MFCC feature extraction are applied to raw speech data. Data is divided into training and testing sets. Codebook of 50 vectors is generated by employing K-means algorithm on training set. Using VQ and codebook indices discrete observation sequences are obtained to train HMM for each word. Testing set is used to analyze performance of VQ-HMM recognition system. The implemented system resulted overall recognition rate of 72.2%.**

Keywords—Telugu words, Mel-Frequency Cepstral Coefficients, Vector Quantization, Hidden Markov Model, Baum-Welch algorith.

I. INTRODUCTION

As the speech recognition systems are encroaching rapidly into the human-machine communication services, ASR systems for local languages is the need of the hour in a multilingual country like India. Several algorithms have been proposed and implemented for speech recognition .Hidden Markov Models (HMMs) superseded other speech recognition models in many speech recognition tasks due to its ease of implementation. Traditional HMMs are widely implemented for speech recognition of different languages. A discrete observation HMM(DHMM) classifier is used for recognition of isolated words of interest. Restriction here is that classifier needs discrete observation symbols as input [1]. So the continuous valued feature extracted from acoustic signal must be quantized to employ discrete observation symbols. Vector quantization (VQ) is a technique where speech vectors are clustered to find out centroid's which constitutes codebook of required size. This codebook is used for quantization of training speech vectors. VQ itself can be used as classifier through indexing speech feature vectors with the help of codebook indices. Here an attempt is made to integrate VQ into HMM framework for Telugu word recognition. This looks forward to leverage advantages inherent in both classifier techniques.

II. APPROACH TO SPEECH RECOGNITION

A. *Various approaches in literature*

Different approaches to speech recognition by machine have been developed from an insight into speech production, perception and analysis. Hardware complexity of their implementations differs widely. Fundamental approaches to speech recognition are acoustic phonetic approach, pattern recognition approach and artificial intelligence approach. Out of them, the pattern recognition approach is preferred due to its robustness and ease of implementation. It exhibits invariance to different speech vocabularies, users, pattern comparison algorithms and decision rules. The choice of features, templates or models for reference patterns and methods used to create reference patterns and classify unknown test patterns are the factors that differentiate different pattern-recognition approaches

B. *Statistical pattern recognition approach*

A pattern obtained with 'd' features from the input data is represented as a point in 'd' dimensional vector space called feature space. Features are selectively chosen so that pattern vectors belonging to different categories of data fall into compact and disjoint regions in the feature space. The idea here is to classify the patterns into different classes by the use of decision boundaries. Two major strategies to reach this objective are density based approach and geometric approach. The 'd' dimensional features are assumed to have are assumed to have a conditional probability density function based on their class. In density based approach, first the estimation of these density functions is carried out. Bayes and maximum likelihood decision rule are used to define the decision boundary. Geometric based approach constructs the decision boundaries directly from optimization of certain cost functions. These two approaches are equivalent under certain assumptions of density functions [2].

C. *Relevance of Statistical pattern recognition approach to speech recognition*

Under this assumption a pre-processor can extract the feature vectors statistically representing acoustic parameters at regular intervals from a sampled speech waveform. Application of statistical pattern recognition approach [3] to speech recognition is to find the most likely word sequence or word given the acoustic data. If A is the acoustic evidence that is provided to the system and $W=w_1,w_2\ldots w_n$ is a sequence of words, then the recognition system must choose a word string or word that maximizes the probability that the word string or word was spoken given that the acoustic data was observed.

III. PRE-PROCESSING OF SPEECH SIGNAL AND FEATURE EXTRACTION

Speech signals are recorded using microphone with sampling rate of 8000 samples/sec. Then the signal is subjected to mean correction to remove any dc offset from microphone. Cropping of signal is performed using endpoint detection and MFCC feature vectors are extracted.

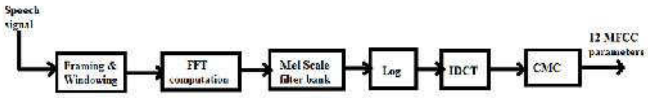


Fig. 1. MFCC Feature Extraction

D. Endpoint Detection

Accurate detection of speech endpoints in a recorded signal is crucial to reduce the size of training data samples. Efficiency of endpoint detection has considerable impact on recognizer's performance. Algorithm used for endpoint detection here is based on short term energy (STE) and zero crossing rate(ZCR).Signal is framed for every 80 samples with 50 percent overlapping of frames. Hamming window is applied to each frame before calculation of STE and ZCR. Absolute STE and ZCR are estimated on a frame basis. Lower and upper thresholds are fixed for STE based on which coarse search for speech boundaries is made. Then fine search of boundaries is executed using ZCR. This gives the starting and ending locations of speech within the signal.

E. Feature Extraction

Feature vectors are extracted from cropped speech signal by employing Mel-Frequency Cepstral analysis. MFCC [4] feature extraction has been extensively in speech recognition applications as it outperforms other parameter types. Signal is divided into frames of 80 samples with its first 20 samples overlapping with that of previous frame.

$$\left[F_{mel} = 2595\log_{10}(1+\frac{f(Hz)}{700})\right] \quad (1)$$

Hamming window is applied for each frame to taper down the signal discontinuities at the end [5].128-point DFT is calculated for each frame to get frequency spectrum. MFCC uses Mel scale shows high resemblance with the human auditory system. For frequencies less than 1 kHz it uses linear scale and beyond 1 kHz its uses logarithmic scale. Hence the acoustic frequency scale is mapped to perceptual frequency scale .The Mel-scale filter bank is generated with 27 triangular filters non-uniformly spaced along the frequency axis with a motive to modify the spectrum to human ear scale. MFCC parameters are computed by applying inverse DCT to the logarithm of the magnitude of the filter bank outputs. This gives 13 coefficients for each frame of speech signal. Cepstral Mean Correction (CMC) is applied to compensate for microphone offsets. Zero-order MFCC coefficient is excluded as average log energy measure is not reliable for speech recognition. Thus a matrix containing 12 MFCC parameters for each frame of signal is obtained from feature extraction. Using MFCC features gives higher recognition rates [6].

IV. VECTOR QUANTIZATION

Vector Quantization is employed to convert the continuous valued feature vectors into discrete observation symbols using codebook generated by K-means algorithm. The training dataset is divided into K clusters with each cluster having its own centroid which is found out by iteration to minimize the squared Euclidean distance parameter between clustered data and its chosen centroid. Thus codebook is generated for the training dataset and updated to make the data uncorrelated as far as possible.12 dimension vectors of all individual frames corresponding to training set are quantized to one of the centroid's and an observation symbol is assigned as index of the centroid. Invariable data values of MFCC vectors are discretized to reduce the level of data set as well the dimensionality by use of K means Algorithm. Testing set feature vectors make use of the code book to realize the observation symbol by comparing each frame feature vector with all codebook vectors such that it is replaced with the index of codebook vector whichever gives minimum distance. This generates an observation sequence for each training word with number of observation symbols equal to that of frames. Observation sequences corresponding to training dataset are used for HMM training.

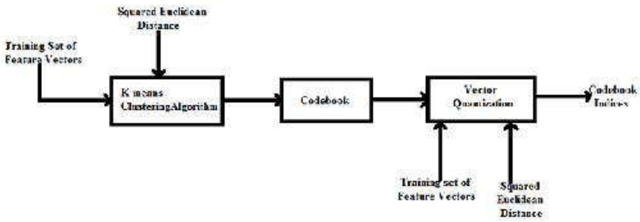


Fig.2. Codebook generation and Vector Quantization

V. DHMM FOR SPEECH RECOGNITION

Hidden Markov Model is a doubly embedded stochastic process with an underlying stochastic process and the sequence of observations can only be observed through another set of stochastic processes.HMM class of statistical speech signal models better accounts for non linear variability in speech waveform, thereby outperforming the deterministic models which exploit known specific properties of the signal [7]. HMM can be employed as the piecewise stationary model of a non-stationary speech signal. Further, it is then possible to estimate the parameters of this stochastic process in a well-defined linear model structure. The speech feature vectors (one feature vector per frame) obtained from speech signal is converted to the discrete observation sequence $O = O_1, O_2, \ldots O_T$. Thus, each word is modeled as a probabilistic model of some observable symbol sequence.

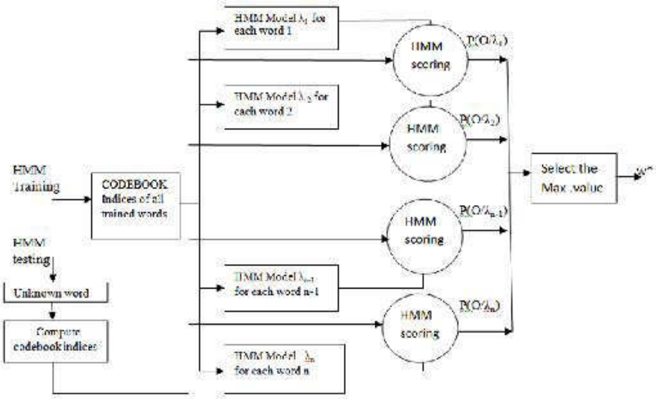


Fig. 3. HMM based isolated word recognition

Parameters of HMM:

- Number of states N in the model.
- Number of distinct observation symbols M per state.
- State transition probability distribution A.
- Observation symbol probability distribution B in a state.
- Initial state probability distribution .

To employ HMM for speech recognition certain computations like probability estimation, finding the optimal state sequence for a given observation sequence and parameter estimation must be carried out. Forward and backward procedures, viterbi algorithm and Maximum likelihood estimate give solutions to above computations.

A. Training the HMM classifier

Training the HMM [8] deals with its parameters estimation such that it encodes the features of particular word to be recognized and be able to recognize similar word utterance after training. There is no known analytical method to optimize HMM parameters in a closed form even though some popular methods are available. Here Baum-Welch algorithm, also known as Forward-backward re-estimation procedure is used to estimate model parameters (A, B,) such that likelihood of occurrence of training data given the model P(O/) is locally maximized. Initially parameters (A, B,) are chosen at random. Then iterative Baum-Welch algorithm [9] is applied to update the model parameters until desired maximization is reached. Thus, HMMs are generated for all words individually and their final model parameters after training phase are used during recognition phase.

B. HMM based Isolated word recognition

The task in terms of recognition aspect is to compute the Log-likelihood parameter for all HMM trained models. This is computed through Forward -Backward procedure which comes under the evaluation problem of HMM. Index of the model with maximum likelihood is identified as recognized word for given unknown word utterance [10].

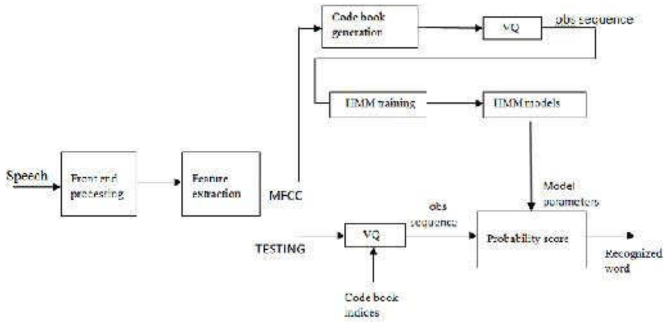


Fig.4 Block diagram of VQ-HMM IWR

VI. IMPLEMENTATION OF VQ-HMM FOR RECOGNITION OF TELUGU DIGITS

Two- thirds of data is taken for training phase and the rest is used during testing phase. The implementation is repeated by varying the codebook size and number of HMM states. From these, the parameters that resulted better recognition performance are identified. Eighty experiments are conducted to obtain overall classification results for testing data using VQ-HMM recognizer. The parameters used for implementation are

- Number of MFCC parameters : 12
- Codebook size : 50
- Number of states in DHMM : 8 Number of conducted experiments : 80

VII. EXPERIMENTAL RESULTS AND CONCLUSIONS

The VQ-HMM recognizer for telugu digits results in overall accuracy of 72.2%.Average recognition rates for each word are shown in Fig.5.Missed recognition cases can be possibly due to some inefficiency in endpoint detection and distortion due to vector quantization. Results indicate that there is further scope to improve the performance. First, the current database may be expanded to include additional speakers and the vocabulary size may be increased considering the human computer interface under consideration. In addition, a “clean” database needs to be collected inside a stringent noise controlled room. It is evident from the simulation results that there is certain error rate in recognition process. This can be attributed mainly due to the inherent background noise in the database collected, incorrect endpoint detection cases and distortion in information that may occur on account of vector quantization. But the recognition rates obtained are fair enough to be used for small scale applications. The results are comparable to that of the systems implemented for other Indian languages.

word	Average classification rate(%)				
	okati	rendu	moodu	nalugu	Aidu
okati	**73.64581**	7.3958	6.5625	0.8333	11.5625
rendu	6.3958	**71.3125**	8.2292	5.0000	7.0625
moodu	3.8542	5.3750	**74.4167**	7.8125	8.5417
nalugu	2.3958	2.2917	12.9167	**72.0833**	10.3125
aidu	10.5208	6.5625	0.2083	14.7708	**67.9375**

Fig. 5. Overall Classification rates

.Average recognition rates for 5 telugu digits with 8 state HMM (codebook size=50)

word	okati	rendu	moodu	nalugu	Aidu
Recognition Rate (%)	73.64	71.31	74.41	72.08	67.40

Average recognition rates for 5 telugu digits with 8 state HMM (codebook size=50)

word	okati	rendu	moodu	nalugu	Aidu
Recognition Rate (%)	72.30	69.31	70.54	73.08	63.96

Fig. 6. Average Recognition Rates

REFERENCES

[1] A Revathi and Y Venkataramani. “Speaker independent continuous speech and isolated digit recognition using vq and hmm.” In Communications Signal Processing (ICCSP), 2011 International Conference on, pages 198–202. IEEE, 2011.

[2] Anil K Jain, Robert P. W. Duin, and Jianchang Mao. Statistical pattern recognition A review. “Pattern Analysis and Machine Intelligence”, IEEE Transactions on, 22(1):4–37, 2000.

[3] MA Anusuya and SK Katti. “Front end analysis of speech recognition a review”. International Journal of Speech Technology, 14(2):99–145, 2011.

[4] Montri Karnjanadecha and Patimakorn Kimsawad. “A comparison of front-end analyses for thai speech recognition”. In INTERSPEECH, 2002.

[5] Joseph W Picone. “Signal modeling techniques in speech recognition”. Proceedings of the IEEE, 81(9):1215–1247, 1993.

[6] RLK Venkateswarlu, R Ravi Teja, and R Vasantha Kumari. “Developing efficient speech recognition system for telugu letter recognition”. In Computing, Communication and Applications (ICCCA), 2012 International Conference on, pages 1–6. IEEE, 2012.

[7] Lawrence Rabiner. “A tutorial on hidden markov models and selected applications in speech recognition.” Proceedings of the IEEE, 77(2):257–286, 1989.

[8] Lawrence R Rabiner and Biing-Hwang Juang. “Fundamentals of speech recognition,” volume 14. PTR Prentice Hall Englewood Cliffs, 1993.

[9] K-F Lee and H-W Hon. “Speaker-independent phone recognition using hidden markov models.” Acoustics, Speech and Signal Processing, IEEE Transactions on, 37(11):1641–1648, 1989.

[10] Fadhilah Rosdi and Raja N Ainon.” Isolated malay speech recognition using hidden markov models.” In Computer and Communication Engineering, 2008. ICCCE 2008. International Conference on, pages 721–725. IEEE, 2008.

Analysis of Denoisied Speech Signal for Different Threshold Selection Rules

Ranbir Kaur
Dept. ECE
Sri Guru Granth Sahib World University
Fatehgarh Sahib, India

Kamaljeet Singh Bhatia
Dept. of EIE
Sri Guru Granth Sahib World University
Fatehgarh Sahib, India.

Harsimrat Kaur
CTIMET, Jalandhar , India

***Abstract*—Wavelet transform is an efficient tool for analysing signal. In this paper analysis of two different wavelets i.e.db13 &sym13 is done for denoisied speech signal with three levels of decomposition using four different threshold rules i.e. Heursure, Rigsure, Minimaxi & Sqwtolog.**

Keywords*—SNR (Signal to noise ratio), MSE (Mean square error), DWT (Discrete wavelet transform), PSNR (Peak signal to noise ratio).

I. INTRODUCTION

Linear methods and non linear methods are used to denoise the signal. In linear method, Fourier transform is used to denoise signal, as compared to non linear method in which wavelet transform is used for denoising signal because of their ability of multiresoulation capability [1].Non stationary signal is represented by combing it with wavelet and mixed transform in mixed transform the signal is spilt into subbands which are further scaled for low-pass and high-pass signal filteration. To maximize the Signal to Noise ratio (SNR) of the reconstructed subband Adaptive algorithms are used in which subbands are finally combined to reconstruct the signal. It is verified that for a given number of coefficients, the proposed technique shows a higher SNR of the reconstructed signal than using mixed-transforms alone or the wavelet transform [2].Analysis of speech signal is done with wavelet thresholding technique and compared the results with Fouriertransform in terms of SNR, MSE, and Spectrogram& PRD .Results shows that the advanced thresholding method is better for denoising the speech signal as compared to conventional methods [3].Compartive analysis for different wavelets is done by using Squtlog thresholding technique to denoise the speech signal for soft thresholding .along with different levels of decompostion are also investigatedin this paper.The quality of speech signal is expressed in terms of peak signal to noise ratio[4].New denoising algorithm is proposed for speech signal with wavelet transform in which the implementation of two wavelets i.e. haar and daudechies is evaluated [5].Compartive analysis of various wavelets is done for speech signal using denoising technique and the effect of wavelet decomposition is also investigated. Universal threshold rule (Square root log) is used for selecting the value of threshold with soft threshold which can shrink coefficients .Peak Signal to Noise Ratio (PSNR) is calculated for checking the quality of denoised speech signal [6].The DWT is computed by passing a signal through a high-pass and a low pass filter . For each decomposition level, the high-pass filter produces the approximations coefficients and the complementary low-pass filter represents the scaling function produces the details coefficients. After decomposition, the signal can be reconstructed back by using Inverse Wavelet Transform Donoho and Johnston developed a universal thresholding rule which can effectively remove the Gaussian random noise [7].Squtlog thresholding technique is used to denoise the speechsignal for soft thresholding along with different levels of decomposition[8].The quality of speech signal is expressed in terms of peak signal to noise ratio.Hard thresholding can be described as the process of setting to zero the elements whose absolute values are lower than the threshold. Soft thresholding is called shrink or kill which is an extension of hard thresholding [9].

II. RESULT & DISCUSSION

Table I Shows decomposition values of Denoisied speech signal at level 1 for both soft & hard threshold .Decomposition values of Minimaxi and Sqtwolog threshold rules remains same for two wavelets i.e.db13 and sym13 but the values of decomposition for heursure & Rigrsure threshold rules differ

Table I

Type of wavelet		Heursure	Rigrsure	Minimaxi	Sqtwolog
db13	SOFT	0.3761	0.3600	3.1897	4.6036
	HARD	0.3310	0.3622	3.1897	4.6036
sym13	SOFT	0.3435	0.3235	3.1897	4.6036
	HARD	0.3232	0.3390	3.1897	4.6036

Graph 1 - Showed decomposition values of Denoisied speech signal at level 1. Soft1 & hard1 shows the decomposition values of speech signal for the wavelet db13. Soft2 & hard2 shows the decomposition values of speech signal for the wavelet sym13.Heursure threshold shows highest value of decomposition for denoisied speech signal.

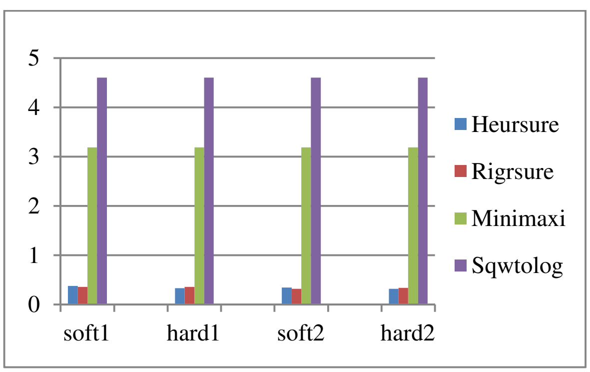


Fig.1.Decomposition Values at level 1

Table 2 - Shows decomposition values of Denoisied speech signal at level 2 for two different wavelet i.e.db13 & sym13 for both soft and hard threshold .The decomposition values for heursure & Rigrsure threshold rules increases at level 2 but the values of Minimaxi and Sqtwolog threshold rule remains same.

Table II

Type of wavelet		Heursure	Rigrsure	Minimaxi	Sqtwolog
db13	SOFT	0.5511	0.5125	3.1897	4.6036
	HARD	0.5400	0.5251	3.1897	4.6036
sym13	SOFT	0.4783	0.5114	3.1897	4.6036
	HARD	0.4931	0.5009	3.1897	4.6036

Graph 2 - Shows decomposition values of Denoisied speech signal at level 2. Soft1 & hard1 shows the decomposition values of speech signal for thewavelet db13. Soft2 & hard2 shows the decomposition values of denoisied speech signal for the wavelet sym13.

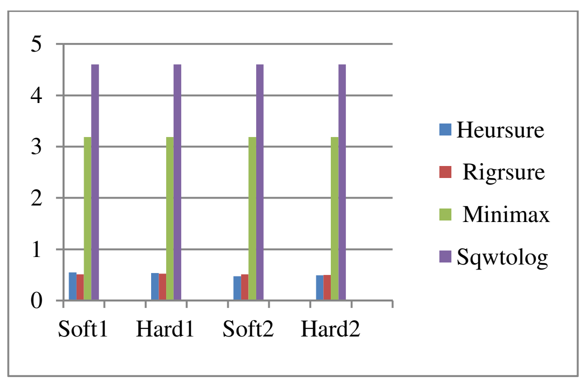


Fig.2. Decomposition Values at level 2

Table 3 -Shows decomposition values of Denoisied speech signal at level 3.Minimaxi and Sqtwolog decomposition values remains same but the values of heursure & Rigrsure threshold rule further increases at level 3 for both wavelets i.e.db13 and sym13.

Type of wavelet		Heursure	Rigrsure	Minimaxi	Sqtwolog
db13	SOFT	0.7636	0.8229	3.1897	4.6036
	HARD	0.8060	0.7910	3.1897	4.6036
sym13	SOFT	0.7790	0.7892	3.1897	4.6036
	HARD	0.8070	0.8018	3.1897	4.6036

Table III

Graph 3 - Shows decomposition values of Denoisied speech signal at level 2. Soft1 & hard1 shows the decomposition values of speech signal for the wavelet db13 and Soft2 & hard2 shows the decomposition values of speech signal for the wavelet sym13.Heursure threshold rule gives the highest value of decomposition .

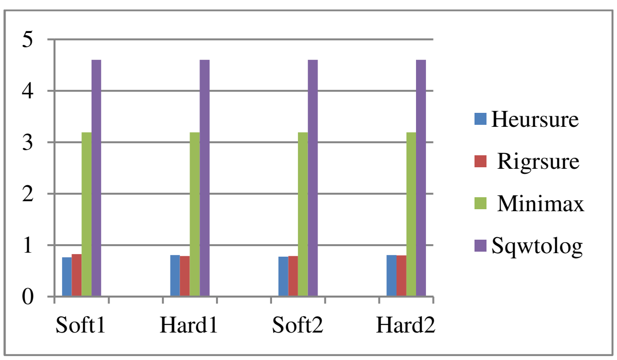


Fig.3. Decomposition Values at level 3

III. CONCLUSION

Denoising speech signal is decomposed at level 3 .The decomposition values for two different wavelet are calculated for different threshold i.e. Heurure, Rigsure, Minimaxi and Sqwtolog for denoisied speech signal .Decomposition value of db13 wavelet and sym 13 wavelet remains same for Minimaxi & Sqwtolog threshold rules but decomposition values for Heursure and Regsure increases from level 1 to level 3 .

REFERENCES

[1] Mallet, "A wavelet tour of signal processing" Academic Press, San Diego, USA, 1998.

[2] Wasfy B.Mikhael and Arun Ramaswamy, "An efficient representation of non stationary signal using mixed transform with application to speech", IEEE transaction on circuits and system: Analog and digital signal processing, Vol.42, No.6, pp.393-401, June1995.

[3] Puneet Arora(et.al) , "Comparative Analysis of Advanced Thresholding Methods for Speech-Signal Denoising",International journal of computer analysis,Vol.59 ,No.16, pp.28-32,December 2012

[4] A.K.Verma, "A Compartive analysis of wavelets in denoising of speech signals

[5] Dr.Mahesh S.Chaven (et.al), "Studies on implemention of wavelet for denoising speech signal", International journal on computer application, Vol.3, No.2, pp.1-7, June 2010.

[6] A.K.Verma(et.al) , "A comparative analysis performance analysis of wavelets in denoising of speech signals" ,International journal of computer applications ,pp.29-32 ,2012

[7] X.Zhou, C.Zhou and IJ.Kemp, "An improvement methodology for application of wavelet transform to partial discharge measuring denoising", IEEE transaction on dielectric and electric insulations, Vol.12, No.3, pp.586-594, June 2005

[8] Larbi Boubchir and, Bouaelm Boashash, "Wavelet denoising on MAP estimation using the BKF prior with application to image and EEG signal", IEEE transaction on signal processing, Vol.61, No.8, pp.1880-1894, April 2013.

[9] D.L.Donoho, "Denoising by soft thresholding", IEEE transaction on information theory, 41, No.3, pp.613-627.

Digital Image Processing by using Various Noise Models and Filtering Techniques

Jaspreet Kaur
Department of Computer science Engg.
CTITR, Maqsudan
Jalandhar, Punjab, India

Samreet Kaur
Department of Computer science Engg.
CTITR, Maqsudan.
Jalandhar, Punjab, India

Sarabjit Kaur
Department of Computer science Engg.
CTITR, Maqsudan.
Jalandhar, Punjab, India

Abstract—**The aim of the Digital Image Processing is to extract the significant features from the image in a desired manner. In this paper, noise reduction and image restoration is considered to improve the quality of the digital image. The efficient techniques are used to reduce the different types of noise. Noise can be generated during image transmission. The basic standard of the noise removal problem depends upon type of the noise through which the image is corrupted. In the field of reducing the noise from the image various linear or non-linear filtering techniques have been suggested. This paper introduces the different approaches for noise reduction and image enhancement.**

Keywords—Digital Image Processing, Noise Models, Filtering techniques

I. INTRODUCTION

Digital Image Processing processes the digital images. Digital Images deals with digital images with the help of digital computer. The digital image is the input to the computer system that system processes the image and gives the output of the image. Signal processing that deals with image analysis and processing of analog and digital signals and perform various actions such as storing, filtering and other operations. These signals include transmission signals, image signals. Digital image processing has controlled over analog image processing due to its applications. These images go through the different steps so as to obtain an image having good quality. The different steps used to produce an image are:

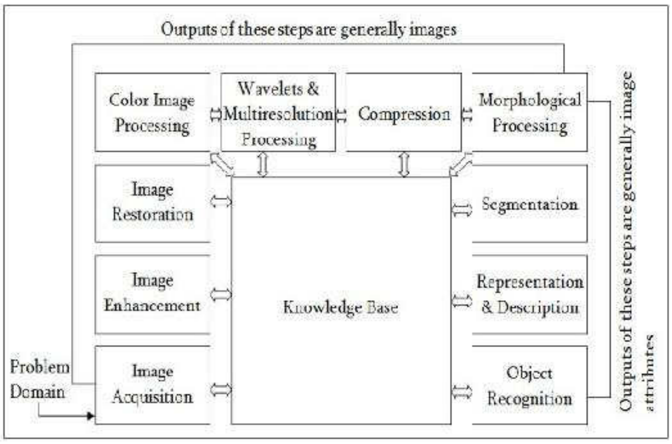


Fig 1. Stages in Digital Image Processing

A). Image Acquisitions

It involves preprocessing to images such as scaling. It is always the first step in image processing. If the image has been acquired, various methods of processing can be implemented on the image to perform the different tasks.

B). Image Enhancement

It involves manipulating an original image in such a way that the resultant image is more suitable than original one. It is easy to implement and supports many filters. Such as image sharpening, highlights the edges of image, improves the quality of a digital image.

C). Image Restoration

It includes various mathematical models to improve the appearance of image. The role of image restoration is to compensate for noises and data errors. To remove noises from the images through different filters comes under image restoration.

D) Color Image Processing

It involves the use of colored images through internet and applying different color models on them. It includes color formation.

E). Wavelets

It represent the images at various degree of resolution. This is useful in representation and reconstruction of the digital signals.

F). Compression

It is useful for storing the images. The various techniques which compresses the images to reduce the storage space.

G). Morphological Processing

It involves various tools which are useful in representing the shape of image. The morphological operations are completely depends on the ordering of the pixels value.

H). Segmentation

It involves partition of an image into various parts and then applying techniques on segmented parts to acquire more suitable image.

I). Representation and Description

It is used to represent the image that give quantitative information and manages the extracting attributes.

J) Object Recognition

It is the process which is used for identifying a specific object in digital image and assigning a label to the image. These are the various steps to represent the image. In this paper we only define the Image Restoration steps and describe various types of noises that occur in the image. It also gives description of various filters that helps in removing these noises.

II. ORIGIN OF NOISE

Noise is the variation of bits which is introducing by the image sensor due to the electronic fluctuations. As an image is acquired by digital camera, it is unable to use directly. Image noise can be occurred through number of reasons. When the image is examined from the photograph made on film, then film grain is a noise. Noise can be occurred due to electrical interference and electronic transmission of image data. The noise is random variations of intensity and brightness. Noise reduction is more important issue in image processing. There are various techniques for noise reduction which are based on mathematical analysis of the system used to create digital image.

III. NOISE MODELS

Noise can be modeled by a histogram as well as Probability Density Function [PDF]. It represents the unwanted information. During image acquisition, various factors such as environmental conditions may affect the performance of image. As an example, on acquiring images with charge-coupled device camera, intensity of light and sensor temperature is important factors that affect the amount of noise. During transmission, images are corrupted due to interference in the channel. The common types of noises are:-

- Gaussian Noise
- Salt-and-Pepper Noise
- Uniform Noise
- Gamma Noise
- Rayleigh Noise

A). Gaussian Noise or Amplifier Noise

Gaussian Noise describes statistical noise and its probability density function is equal to the normal distribution. This noise is also known as Gaussian distribution.

The PDF of a Gaussian Noise is specified by:-

$$p_G(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \quad (1)$$

B). Salt-and-Pepper Noise or Impulsive Noise

Salt-and-Pepper Noise is also referred to as shot noise and spike noise. If the image have dark pixels in bright regions and bright pixels in dark regions that means the image contains impulsive noise. This noise can be caused by bit errors in transmissions and eliminated by using interpolating around dark or bright pixels.

The impulse noise is defined as:-

$$p(z) = \begin{cases} P_a & \text{for } z = a \\ P_b & \text{for } z = b \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

C). Uniform Noise

Uniform Noise is also termed as Quantization noise. This type of noise can be caused by quantizing the pixels and which is also known as Uniform Distribution. The values of the gray levels are uniformly distributed. The different types of noise distribution can be generated by uniform noise. The uniform noise is used for image degradation.

The PDF of this noise is given by:-

$$p(z) = \begin{cases} \frac{1}{b-a} & \text{if } a \le z \le b \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

$$\mu = \frac{a+b}{2} \qquad \sigma^2 = \frac{(b-a)^2}{12}$$

D). Gamma Noise or Erlang Noise

Gamma Noise can be observed in laser imaging by using low-pass filtering.

The PDF of the Gamma Noise is given by:-

$$p(z) = \begin{cases} \frac{a^b z^{b-1}}{(b-1)!} e^{-az} & \text{for } z \ge 0 \\ 0 & \text{for } z < 0 \end{cases} \quad (4)$$

$$\mu = \frac{b}{a} \qquad \sigma^2 = \frac{b}{a^2}$$

E). Rayleigh Noise

This type of noise is helpful in qualifying the range and velocity imaging.

The PDF of Rayleigh Noise is given by:-

$$p(z) = \begin{cases} \frac{2}{b}(z-a)e^{-(z-a)^2/b} & \text{for } z \geq a \\ 0 & \text{for } z < a \end{cases} \tag{5}$$

$$\mu = a + \sqrt{\pi b/4} \qquad \sigma^2 = \frac{b(4-\pi)}{4}$$

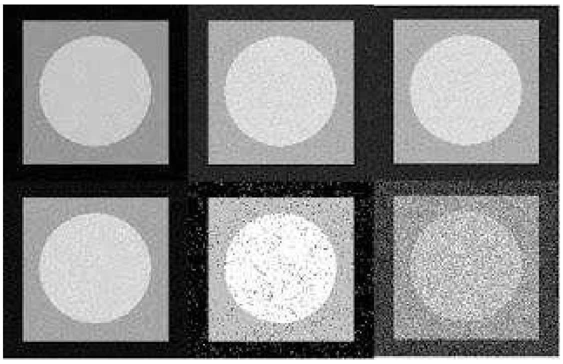


Fig 2 Different types of Noises

IV. FILTERING TECHNIQUES

Filtering is a principle operation in Digital Image Processing. It can be used for many tasks such as an image enhancement, noise reduction, smoothing and sharpening. Filtering is the standard process for modifying and enhancing an image. The concept of the filtering can be either in frequency domain or spatial domain. In the frequency domain filters, specified frequency components are accepting or rejecting. The spatial filter has neighborhood and predefined operation. The choice of filter is characterized according to the behavior of image data. Filters are used to remove the unwanted noise from digital image while holding the details of image and prevent the essential part of image processing. Filters can be classified into the two different categories:-

- Linear Filters
- Non-Linear Filters

A). Linear Filters

Linear Filters are applied by using convolution operations. In a linear filter, the output will be changed linearly with the change in input. The output image is represented by the weighted sum of the pixel value in the input image. The Gaussian and Average filters are used for this purpose. These filters have tendency to blur the sharp edges, destruct the lines and other finely details of images and perform hardly in the presence of the signal dependent noise.

B). Average Filter:

Avearge Filters are also known as Mean filters The value of the output pixel is the mean of its neighborhood.This filters performs on each pixel in an image by using square window which is surrounding each pixel.It is easy to implement and used for removing the noise from an image.

$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

Figure. 3 Example of Average Filter

1). Algorithm for Average Filter:-

- Place the window of size 3*3.
- Calculate the average value, sum up all the elements/pixels in window and divided by the number of elements.i.e.9
- Finally, the center value of the filter is replaced by the calculated average value.

C). Non-Linear Filters

Non-Linear Filters are order statistic filters, because it cannot be applied by convolution. The order statistic filter is a technique which arranges all the pixels in serial order and operates on small window. The position of the pixels can be characterized by its rank. It overcomes the shortcoming of linear filter by using Median Filter.

D). Median Filter

Median Filter is a non-linear filtering technique. In this type of filter, we do not replace the pixel value with the mean value; here we replace it with median value. The median can be calculated by first sorting all values of the pixels in ascending order and then replace its value with the middle value that is being calculated .It is quite easy to implement. These filter is more popular because it preserves the edges and very effective in the Impulsive Noise or Salt-and-Pepper Noise.

3	3	4
4	87	4
4	5	5

Fig 4 Example of Median Filter
Sorted pixel value:-
3,3,4,4,4,4,5,5,87
Median Filter Value=4

1). Algorithm for Median Filter:-

- Select the window of size 3*3 and study each pixel in the image.
- Sort all the pixels in increasing order according to their intensity values.

- Then replace the original value of pixel with the median value from the list.

V. CONCLUSION

In the image processing, enhancement is an essential task. This paper presents the several noise models and filtering techniques for de-noising the noise. The problem with average filter that it does not remove the image details whereas median filter have tendency to remove image details but its performance is unsatisfactory in single- dependent noise. The future research will be concentrated on other filtering techniques to inhibit noises

REFERENCES

[1]. Castleman, K.R.,"Digital Image Processing", Second Edition. Englewood Cliffs, New Jersy: Prentice-Hall, 1996.

[2]. Rafael C .Gonzalez and Richard E. woods, "Digital Image Processing", Pearson Education, Second Edition, 2005.

[3]. James Church, Dr. Yixin Chen, and Dr. Stephen Rice. " A Spatial Median Filter for Noise Removal in Digital Images" 2006

[4]. Charles Boncelet, "Image Noise Models" in Alan C.Bovik, Handbook of Image and Video Processing, 2005

[5]. Astola, J., Kuosmanen, P., "Fundamentals of Non-linear Digital Filtering", CRC Press, New York, Boca Raton, 1997.

[6]. I. Pitas and A.N.Venetsanopoulas,"Nonlinear Digital Filters: Principles and Applications, Boston, Kluwer, 1990.

[7]. R. Oten, R. de Figueiredo, "Adaptive Alpha Trimmed Mean Filters under Deviations from Assumed Noise Model", IEEE Trans. Image Processing, Vol. 13, No. 5, pp. 627-639, 2004,.

[8]. Piotr S. Windyaga, "Fast Impulsive Noise Removal", IEEE Trans. On Image Processing, Vol.1 0, No, 1, January 2001.

Segmentation of Document Images and Identification of Segmented Regions

Dharamveer Sharma
Dept. of Computer Science
Punjabi University
Patiala
dveer72@hotmail.com

Silky Arora
Dept. of Computer Science
Punjabi University
Patiala
arorasilky08@gmail.com

***Abstract*— Efficiency of a professional quality OCR depends on identifying the structural schema of the document. Document image segmentation is done to classify and analyze its contents. In this paper a segmentation technique is presented to segment a document image into its identifying regions. Image is iteratively decomposed into smaller blocks and then that block type is identified. This is implemented by combining two approaches- Run length approach and an iterative top down approach. This technique is applicable for documents in manhattan format and is script independent.**

Keywords— Document Image Segmentation, Horizontal and Vertical Profiles, Horizontal and Vertical Thresholds, OCR, Block type Identification.

I. INTRODUCTION

Millions of paper documents are being converted to scanned digital format for better access storage and retrieval. There can be a wide variety of documents such as books, journals, newspapers, magazines etc. and they can be in variety of languages. For performing OCR on a document, it is essential to know its structural schema. There may be figures, pictures, graphs of tables with in the document. These need to be segmented and identified. Document layout may also vary- Non- Manhattan layout and Manhattan layout as shown in fig. 1 and fig. 2 respectively. Layout analysis is the process of identifying layout structure of the page image. There can be physical layout analysis or logical layout analysis. During physical structure analysis the document image is divided into smaller geometrical regions that share same physical structure. While logical structure analysis works to identify headings, caption, titles etc. This paper deals with geometric structure analysis only. During analysis and interpretation of document images they can broadly be categorized as text components and non-text components. Non-text components can be pictures, graphs, tables etc. A complete understanding of the structure of a document is possible only with information about the contents of the document. Hence it is important to develop an integrated approach to document layout and structure analysis and various recognition modules.

Segmentation is done to analyze the layout and contents of the document. Up to now various strategies have been developed for segmentation. These segmentation strategies are based on broadly three types of approaches- Top Down Approach, Bottom Up approach, Hybrid Approach. In Top Down Approach for segmentation the whole document is taken at once and then divided recursively into smaller blocks until the block shares same kind of features. While Bottom Up approach starts from pixel level and starts from the pixel level and keep merging the components until when they share same kind of features. It is assumed here that noise removal and skew correction is already done on the image. The remaining paper has been divided as follows: section 2 covers work already done, section 3 presents objectives and scheme of the proposed work, section 4 highlights the results obtained and paper has been concluded in section 5, references are given at the end of the paper.


LIFESTYLE
The geek grew up


Grenzgänger zwischen Jazz und Klassik

Fig. 1. Non-Manhattan Layout

Fig. 2. Manhattan Layout

II. REVIEW OF LITERATURE

One of the earliest approaches of document analysis system was given by Wong et al. in 1982 [1]. For segmentation and classification of digital documents it uses run length smearing approach. By using regular features of text lines, a linear adaptive classification scheme discriminates text regions from others. A texture based document segmentation and classification approach [2] divides the

Formatted

Formatted: Font: Times New Roman, 10 pt

document image into three regions- graphics, text and space. The classification of groups is done using pre-learned heuristic rules. A top down segmentation technique named X-Y Cut algorithm [3] recursively subdivides the document image in rectangular boxes. This subdivision is represented using x-y tree using whole document as root node and all other nodes represent rectangular block of connected components. The basic criterion for subdivision is to calculate pixel-projection profiles in both horizontal and vertical directions. Run length smearing algorithm [4] is also based on top down segmentation approach. It takes whole binarized image as input and subdivides it into homogeneous areas. Threshold values are defined in this algorithm. White runs in the horizontal direction are replaced with black pixels, if they are less than threshold value, and black pixels remain untouched. This algorithm is applied in three steps-horizontally, vertically and results of both are combined by logical AND operator. Choosing threshold value in this approach is critical and depends upon the document class. There is another approach that does not depend on any threshold value, white space analysis algorithm [5]. It follows bottom up approach. White spaces around the component blocks are aggregated. Runs of white spaces greater than one-fifth of the page are identified in both horizontal and vertical directions. Thinning of lines is done using thinning algorithm and finally mesh is formed using horizontal and vertical lines. Hu et al. [6] does layout comparison in two steps. Different methods are used to compute the inter row gaps after segmenting the image into a grid of equal-sized cells. Kasturi et al. [7] describes various components of the document analysis system. It describes various building blocks of document analysis system irrespective of the problem domain or language of the document. Karim et al. [8] introduced an automatic segmentation approach which was presented in first International Newspaper segmentation contest. This approach divides the newspaper image into titles, horizontal and vertical lines, text regions and picture regions. Bottom up approach for document image segmentation is implemented by Docstrum algorithm [9]. It starts from pixel level information and finds k-nearest neighborhood pixels and start converging them. This approach is also threshold based, thresholds are based on properties of distance and angle of each connected components with k-nearest neighbors. There is one more approach that is based on bottom up approach called voronoi diagram based algorithm [10]. Shafait et al. [11] presented a paper discussing comparison of various segmentation approaches.

III. Proposed Work

The objective of our work is to segment document images recursively using horizontal and vertical thresholds until the resulting regions share same physical layout. After segmentation identification of each rectangular block is done using horizontal and vertical histogram values for each of the block.

The technique presented in this paper for segmentation of document images combines two approaches- one is Run Length Smearing algorithm and other is recursive top down approach. Microsoft visual studio 2010 is used to implement the proposed technique. Run Length Smearing algorithm is implemented for identifying heading rows of the document. In this approach horizontal run lengths are drawn for the image by merging two black pixels which are less than a threshold value apart. Thus a continuous stream of black pixels is formed in horizontal direction. The top heading rows are differentiated from rest of the image by selecting first block formed by horizontal run lengths, in case of single top heading row and first two blocks in case top heading has two rows. In this way heading rows can be identified and separated from the image.

After separating heading rows, rest of the image is segmented using recursive top down approach. In recursive top down approach rest of the image is taken as input and then recursively divide it into smaller regions until each region has same physical characteristics. There are two parts of this approach- first the image is divided into its identifying regions and second each region is recognized, whether it is picture or text or graph etc.

Recursive top down segmentation is implemented using horizontal and vertical projection profiles and is implemented recursively. Also this technique works differently for columnar and non- columnar images. Top down segmentation technique works as follows-

1. Get full image except heading rows as first rectangle R.
2. Create Horizontal Projection Profiles (HPPs) and find horizontal gaps more than or equal to horizontal threshold. Divide the rectangle R into smaller rectangles by creating dissection points using these horizontal gaps.
3. If number of rectangles formed using HPPs is more than 1 then create Vertical Projection Profiles (VPPs) for each of rectangle and find vertical gaps more than or equal to vertical threshold. Using these vertical gaps, final rectangles are formed.
4. If number of rectangles using HPPs is 1 then create Vertical Projection Profiles (VPPs) for that rectangle and find vertical gaps more than or equal to vertical threshold. Using these vertical gaps, final rectangles are formed. If number of rectangle formed using VPPs is more than one then create HPPs for each rectangle and horizontal gaps more than or equal to horizontal threshold are identified. Final rectangles are created using HPPs.

In this way whole image rectangle is segmented using top down approach recursively. This algorithm runs for rectangles at each level until they cannot be further divided using horizontal or vertical thresholds.

Till now a number of smaller rectangular blocks are formed, now type of rectangular blocks will be identified. Following series of steps are required to perform this identification-

1. Draw horizontal histograms for each rectangular region.
2. If for a rectangular region at regular gaps horizontal histogram values are not equal to or slightly greater than 0 then the rectangular region is picture.
3. If for a rectangular region at regular gaps horizontal histogram values are equal to or slightly greater than 0 then the rectangular region is text block or table.
4. To further differentiate between text and tables we will draw vertical histograms for both. If at regular gaps vertical histogram values are equal to or slightly greater than 0 then the region is table otherwise it is text.

Fig. 3, 4 and 5 show flowchart representations of the above algorithms proposed. While fig. 3 represents the isolation of heading rows, fig 4 and fig 5 represents segmentation of document image using recursive top down approach and identification of segmented blocks by identifying gaps in horizontal and vertical histogram values respectively.

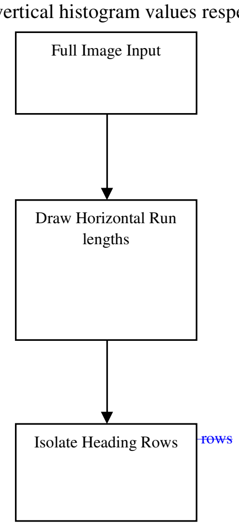


Fig. 3. Isolation of heading rows

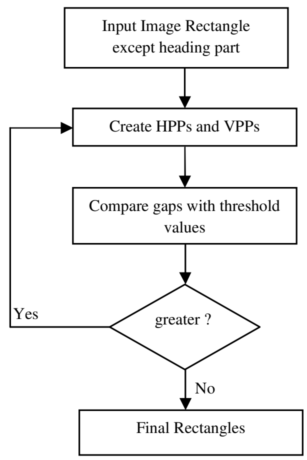
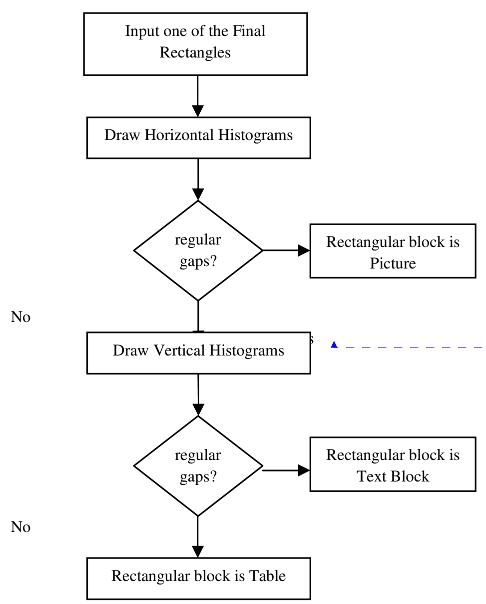


Fig. 4. Recursive Top Down Approach for Segmentation



Formatted: Indent: First line: 0.5"

Formatted: Font: Calibri, 11 pt

Yes

Fig. 5. Identification of each Segmented Rectangular Block

IV. ~~IV.~~ RESULTS

First the heading rows are distinguished from rest of the image using horizontal run lengths. The technique presented in this paper works for any number of heading rows. Fig. 6 shows Image example with single heading row and fig. 7 shows result after isolation of heading row. Similarly fig. 8 and fig. 9 show image example and result for double heading rows. The image examples are collected by scanning newspaper articles. For identification of segmented regions, horizontal and vertical histograms are drawn and their values are compared. Fig. 10 shows resulting image after horizontal histograms are drawn for each of the rectangular block. Similarly fig. 11 shows vertical histograms for each of the segmented region. Histograms are drawn for the image in shown in fig. 6. The pixel co-ordinates where histogram values are equal to or slightly greater than zero represents gaps in the content of rectangular block. From the variation in values of horizontal and vertical histograms, it can be concluded that the rectangular block is text block, image or table.

Fig. 6. Example Image Fig. 7. Image after Isolation of Single Heading Row

Fig. 8. Example Image Fig. 9. Image after Isolation of ~~Double Heading Row~~ Double Heading Row

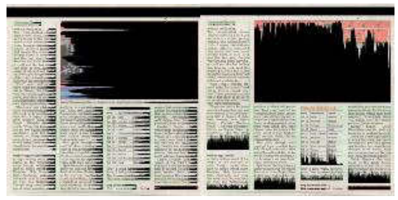


Fig. 10. Horizontal histograms of each block Fig. 11. Vertical Histograms of each block

V. ~~V.~~ CONCLUSION

In this paper a segmentation technique is developed which is different from classical segmentation approaches. An optimized framework is developed to get close solution to the problem. As the heading rows are distinguished from rest of the image, an optimized threshold value can easily be calculated using horizontal and vertical histogram values for each of the identifying region separately.

REFERENCES

[1] K.Y. Wong, R.G. Casey and F.M. Wahl, “Document analysis system”, IBM Journal of Research and Development, vol. 26, no. 6, pp. 647-656, 1982.

[2] M.W. Lin, J.R. Tapamo and B. Ndovie, “Texture-based Method for Document Segmentation and Classification”, South African Computer Journal, vol. 36, pp. 49-56, 2006.

[3] B. Kruatrachue, N. Moongfangklang and K. Siriboon, “Fast Document Segmentation Using Contour and X-Y Cut Technique”, International Journal of Computer, Information science and Engineering, vol. 1, no. 5, 2007.

[4] N. Priyadharshini and M. S. Vijaya, “Genetic Programming for Document Segmentation and Region Classification Using Discipulus”, International Journal of Advanced Research in Artificial Intelligence (IJARAI), vol. 2, no. 2, 2013.

[5] R. Garg, G. Harit and S. Chaudhury, “A hierarchical analysis scheme for robust segmentation of Document Images using white-spaces”, in proceedings of 1st National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics, Gujarat, India, 2008.

[6] J. Hu, R. Kashi and G. Wilfong, “Document Classification using Layout Analysis”, in proceedings of International Workshop on Document Analysis and Understanding For Document Database, Florence, Italy, pp. 556–560 , 1999.

Formatted: Indent: Left: 0.5"

Formatted: Indent: First line: 0.5"

Formatted: Indent: Left: -0.01", Hanging: 0.26", Numbered + Level: 1 + Numbering Style: I, II, III, … + Start at: 1 + Alignment: Right + Aligned at: 0.83" + Indent at: 0.95"

Formatted: Indent: Left: -0.01", Hanging: 0.26", Numbered + Level: 1 + Numbering Style: I, II, III, … + Start at: 1 + Alignment: Right + Aligned at: 0.83" + Indent at: 0.95"

Formatted: Centered

Formatted: Centered

[7] R. Kasturi, L. O'Gorman and V. Govindaraju, "Document Image Analysis: A Primer", Sadhana, vol. 27, no. 1, pp. 3–22, 2002.

[8] K. Hadjar, O. Hitz and R. Ingold., "Newspaper Page Documentation using a Split and Merge Approach", in proceedings of 6th International Conference on Document Analysis and Recognition, pp. 1186–1189, Seattle, Washington, USA, 2001.

[9] L. O'Gorman, "The Document Spectrum for Page Layout Analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no.11, pp. 1162–1173, 1993.

[10] K. Kise, A. sato and M. Iwata, "Segmentation of Page Images Using the Area Vornoi Diagram", computer Vision and Image Understanding-Special Issue on Document Image Understanding and Retrieval, vol. 70, issue 3, pp. 370-382, 1998.

[11] F. Shafait, D. Keysers and T. M. Breuel, "Performance Comparison of Six Algorithms for Page Segmentation", in proceedings of 7th International Association of Pattern Recognition (IAPR) Workshop on Document Analysis System,pp.368–379,2006.

Recognition of Isolated Machine Printed Characters of Gurmukhi Script Captured through Mobile Camera

Dharamveer Sharma
Department of Computer Science
Punjabi University
Patiala
dveer72@hotmail.com

Shilpa Arora
Department of Computer Science
Punjabi University
Patiala
arorashilpa69@yahoo.com

***Abstract*—This paper presents an analysis of applying different classifiers for recognition of isolated machine printed characters of Gurmukhi script. It has been assumed that the characters are pre-segmented and noise free. This forms the basis for recognition of machine printed text of Gurmukhi script, which may even include text extracted from natural scene images or sign boards. The whole process consists of two stages. In first stage the features which uniquely identify the character are extracted using zoning. In second stage classifiers KNN and SVM with different kernels are used. Probability outputs are used to classify characters. The KNN classifier has achieved higher accuracy as compared to SVM. But time taken by SVM is lesser than time taken by KNN.**

Keywords—Gurmukhi script, svm, isolated characters, classification, probability.

I. Introduction

The detection and extraction of text regions in an image is a well-known research area in the computer vision. Text embedded in images contains large quantities of useful semantic information, which can be used to fully understand images. However, since the text data may be embedded in an image in different font styles, orientations, colours, and against a complex background, the problem of extracting the candidate text region becomes a challenging task. Also, in some images the observed characters may have widely different sizes; it is therefore difficult to extract all text areas from the image using only a single method. Recognition of Indian language scripts is a challenging problem. Optical Character Recognition (OCR) is the process of converting scanned images of machine printed or handwritten text into a computer processing format.Work for the development of complete OCR systems for Indian language scripts is major field of research. Research in the field of recognition of Gurmukhi script faces major problem mainly related to the unique characteristics of the script like connectivity of characters on the headline, characters in a word present in both horizontal and vertical directions, two or more characters in a word having intersecting minimum bounding rectangles along horizontal direction, existence of a large set of visually similar character pairs, multi-component characters, touching characters which are present even in clean documents and horizontally overlapping text segments. The machine printed isolated Gurmukhi characters for recognition is considered. The work here is to design model using training data which can classify the unknown pattern based on that model. For training purposes 7257 isolated Gurmukhi characters images of different font styles with different camera resolutions are collected. The rest of the paper is organized like section 2 covers review of literature, section 3 elaborates objectives, assumptions and the proposed technique, section 4 covers results and discussion, section 5 concludes the paper and references are given at the end.

II. Review of Literature

A lot of research has been done on the recognition of isolated Gurmukhi characters, resulting in number of different techniques. Sharma and Lehal [1**]**, presents technique which segments the words in an iterative fashion by focusing on presence of headline, aspect ratio of characters and vertical and horizontal projection profiles. The proposed approach of segmentation can be used for handwritten text of Indian language scripts like Devanagari, Bangla etc. having structural feature similar to Gurmukhi script. Lehal and Chandan [2], presents the complete Machine-Printed Gurmukhi OCR System .This is the first time that a complete multi-font and multi-size OCR system for Gurmukhi script has been developed. It has been tested on good quality images from books and laser print outs .This paper addresses the problems in the various stages of the development of a complete OCR for Gurmukhi script and discusses potential solutions. The overall system design of the Gurmukhi OCR system developed and implemented. As with most of the OCR systems, there are five main processing stages: Digitization, Pre-processing, Segmentation, Recognition and Post-processing. Ohya et al. [3], presents a method for recognizing characters in scene images. An image segmentation method based on local thresholding is applied and an evaluation of the gray-level differences between regions and their back ground is taken. Tou and Gonalez [4], designed handwritten character recognition system making use of topological feature extraction and multilevel decision making. By properly specifying a set of easily detectable topological features, the system classifies the handwritten characters into primary categories. The recognition system consists of two stages, with the final classification accomplished by a secondary stage that performs local analysis on the characters in each primary category. The recognition system consists of two stages: global recognition, followed by local recognition. Jhajj and Sharma [5], presents a system to recognize isolated handwritten Gurmukhi characters. Zoning and two classification methods, k-nearest neighbor, SVM (support vector machines) have been used and compared. Paper work is done by classifying image area in two classes, text and non text using SVM (support vector machine). Kaur and Josan [6], identified the features and train a model based on the feature vector which is then used to classify text and non text area in an image. This algorithm is in sensitive to text orientation the output of the text extraction algorithm is can be fed to an OCR system to recognize the contained information. The results obtained on varied set of images are compared with respect to precision and recall rates. This work is based on feature extraction and classification using SVM. Assumption is made that images are of good quality and free from noise. Lee et al. [7], propose a new methodology for character segmentation and recognition which makes the best use of the characteristics of gray scale images. In the proposed methodology, the character segmentation regions are determined by using projection profiles and topographic features extracted from the gray-scale images. Then a nonlinear character segmentation path in each character segmentation region is found by using multi-stage graph search algorithm. Finally, in order to confirm the nonlinear character segmentation paths and recognition results, recognition-based segmentation method is adopted. Through the experiments with various kinds of printed documents, it is convinced that the proposed methodology is very effective for the segmentation and recognition of touched and overlap characters. Segmentation and recognition may be improved by using contextual information. The proposed method used only the character recognition results in the recognition-based character segmentation step .However, the contextual knowledge could be used to reject the mis-recognized characters and find another path which has correct segmentation and recognition results.

III. PROPOSED TECHNIQUE

In this paper the system to recognize isolated Gurmukhi characters has been developed. The isolated 7257 character images are used as input. Zoning method has extracted the features and based on that features the KNN and SVM classifier has been used to classify a character. The code has been implemented in Microsoft Visual Studio 2010. The flowchart of proposed technique is as follow:

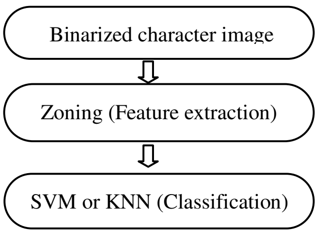


Fig. 1. Flowchart of proposed technique

A. *Objectives of Research*

The objectives of the proposed study are outlined as follows:

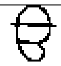
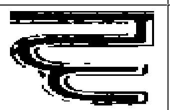
- To develop a software module to recognize Gurmukhi isolated machine printed characters.
- To analyze the different approaches of classification.

B. *Assumptions*

There are following assumptions while developing the system

- Only middle zones (41) characters are considered.
- The character images are noise free.
- Images are pre- segmented binarzied images.
- The images are collected using mobile phone cameras of different resolutions.
- The characters are in different font styles.
- Skewed images are also collected.
- There is no benchmark database for Gurmukhi script. Therefore own database has been used. Some samples of images which have been clicked by different mobile cameras along with their size are shown in following table.

TABLE I. SOME SAMPLES OF IMAGES

 (14*17)	 (24*29)	 (27*31)
 (14*17)	 (24*29)	 (27*31)
 (14*17)	 (24*29)	 (27*31)
 (75*101)	 (72*101)	 (95*110)
 (76*63)	 (33*36)	 (113*99)

C. *Feature Extraction*

The frame containing the character is divided into non-overlapping zones and the densities of object pixels in each zone are calculated. Density is calculated by finding the number of object pixels in each zone and dividing it by total number of pixels. Densities are used to form a representation. For binary images, value of each pixel is either 1 or 0. Object pixels are black pixels.

D. Classification

Two different classification techniques are as follows:

1) K-Nearest Neighbour: The K nearest neighbour method is a non-parametric classification method, which is simple but effective in many cases. It only needs reference data points for both classes. A test sample is then attributed the same class label as the label of the majority of its K nearest (reference) neighbors.The Euclidean distance is calculated between the test point and all the reference points in order to find K nearest neighbours, and then rank the obtained distances in ascending order and take the reference points corresponding to the k smallest Euclidean distances.The different values of k i.e. 1,3,5 are used.

2)Support Vector Machine: Support vector machine is a machine learning technique. An SVM classifier discriminates two classes of feature vectors by generating hyper-surfaces in the feature space, which are optimal in a specific sense that is the hyper-surface obtained by the SVM optimization is guaranteed to have the maximum distance to the nearest training examples, the support vectors. On the binary linear separable case, SVM determines the optimal hyper-plane through maximizing the margin between the separating hyper-plane and the data, subjecting to the constraint of classifying the training samples correctly. SVM can be extended easily to the nonlinear classifier by projecting the data into a high dimensional feature space. Support vector machine is a very flexible technique for classification. There are various parameters that can be set during training. LIBSVM has been used to classify data. LIBSVM is an integrated software for support vector classification.

IV. RESULTS AND DISCUSSIONS

The 7257 images of Gurumukhi characters, 177 images of each character have been used. The system is analyzed with KNN and different types of kernels of Nu-SVC type of SVM. The recognition accuracy is obtained by dividing the correctly recognized characters to total number of character images which are present in the testing database. For analysis two different methods are used, results of both these methods are shown separately.

- First, the same training and testing data has been considered.
- Second, the separate training and testing data has been considered.

A. Same Training and Testing Data with SVM Classifier

The system is trained with 7257 images and tested with the same number of images i.e. with 7257 .The recognition accuracy is given in the following table. The linear kernel classification gives the best result out of all other combinations.

TABLE II. RECOGNITION ACCURACY OF SAME TRAINING AND TESTING DATA WITH SVM

Feature extraction	SVM kernel	Correctly recognized	Recognition accuracy	Time elapsed(in seconds)
zoning	Linear	7173	98.84%	1486
zoning	polynomial	7024	96.78%	1626

The confusion matrix for the linear kernel function is given in the following table. The character ਝ is recognized with least accuracy. The character ਬ character is getting confused maximum number of times (10.61%) with ਥ.

TABLE III. CONFUSION MATRIX OF SAME TRAINING AND TESTING DATA WITH LINEAR KERNEL OF SVM

Character	Confused with character		
ੳ	ੳ 100%		
ਅ	ਅ 100%		
ੲ	ੲ 100%		
ਸ	ਸ 99.115%	ਮ 0.884%	
ਹ	ਹ 99.115%	ੳ 0.884%	
ਕ	ਕ 100%		
ਖ	ਖ 100%		
ਗ	ਗ 98.2%	ਲ਼ 0.884%	ਗ਼ 0.884%
ਘ	ਘ 100%		
ਙ	ਙ 100%		
ਚ	ਚ 97.34%	ਦ 2.654%	
ਛ	ਛ 100%		
ਜ	ਜ 100%		
ਝ	ਝ 87.23%	ਙ 3.191%	ਡ 9.96%
ਞ	ਞ 97.34%	ਵ 2.654%	
ਟ	ਟ 100%		
ਠ	ਠ 100%		
ਡ	ਡ 98.2%	ਝ 1.76%	
ਢ	ਢ 99.115%	ੳ 0.884%	
ਣ	ਣ 99.115%	ਟ 0.884%	
ਤ	ਤ 100%		
ਥ	ਥ 94.69%	ਧ 0.884%	ਬ 4.42%
ਦ	ਦ 100%		
ਧ	ਧ 100%		
ਨ	ਨ 100%		
ਪ	ਪ 100%		
ਫ	ਫ 99.11%	ੳ 0.884%	
ਬ	ਬ 89.38%	ਥ 10.61%	
ਭ	ਭ 99.115%	ੳ 0.884%	
ਮ	ਮ 100%		
ਯ	ਯ 100%		
ਰ	ਰ 100%		

Character	Confused with character		
ਲ	ਲ 99.115%	ਲ਼ 0.884%	
ਵ	ਵ 99.115%	ਜ਼ 0.884%	
ੜ	ੜ 99.101%	ਤ 0.89%	
ਸ਼	ਸ਼ 100%		
ਖ਼	ਖ਼ 99.115%	ਖ 0.884%	
ਗ਼	ਗ਼ 98.23%	ਗ 1.76%	
ਜ਼	ਜ਼ 100%		
ਫ਼	ਫ਼ 100%		
ਲ਼	ਲ਼ 99.115%	ਲ 0.884%	

The probability outputs of SVM are used for classification. In the following tables output is shown for two Gurmukhi characters. The index varies from 0 to 40 it is corresponding to the 41 middle zone characters. The probability variable stores the probability outputs of the SVM. The character is recognized as that character which has the highest probability output from all other 41 values. For example for the character ਝ the first three higher probability SVM outputs are shown in following table. The probability value is much higher so ਝ is correctly classified as ਝ.

TABLE IV. PROBABILITY OUTPUTS FOR ਝ

Index	Probability values
13(ਝ)	0.94851609894939037
21(ਥ)	0.0086607432694186109
17(ਡ)	0.0055631967751457586

Out of 41 values the first three higher probability values for the ਥ character is shown in the table. ਥ is wrongly recognized as character ਬ as higher probability value is assigned to classify a character. Instead of recognizing the character corresponding to the highest probability value, the Corpora analysis can be done and bigrams can be analyzed to correctly recognize a character.

TABLE V. PROBABILITY OUTPUTS FOR ਥ

Index	Probability values
27(ਬ)	0.48052184097681205
21(ਥ)	0.20764565668149301
13(ਝ)	0.038142150341352113

B. Same Training and Testing Data with KNN Classifier

The KNN classifier with different values of k is used on this 7257 dataset of images the results are shown in a following table.

TABLE VI. RECOGNITION ACCURACY OF SAME TRAINING AND TESTING DATA WITH KNN

Feature extraction	K	Correctly recognized	Accuracy	Time elapsed(in seconds)
Zoning	1	7257	100%	3005
Zoning	3	7257	100%	3372
Zoning	5	7257	100%	3506

C. Separate Training and Testing Data with SVM Classifier

The system is trained with 7257 images and tested with the different dataset of images of 4182 images Tthere are 102 images for each character. The results for SVM classifier are shown in following table. The recognition accuracy is obtained by dividing correctly recognized images with total number of images present in testing database i.e. dividing by 4182.

TABLE VII. RECOGNITION ACCURACY OF SEPARATE TRAINING AND TESTING DATA WITH SVM

Feature extraction	SVM kernel	Correctly recognized	Recognition accuracy	Time elapsed(in seconds)
zoning	linear	3966	94.83%	719
zoning	polynomial	3599	86.05%	945

The confusion matrix for the linear kernel function is given in the following table. The character ਛ is recognized with least accuracy. Many characters like ਅ, ਘ, ਟ, ਝ, ਪ, ਲ, ਭ have been recognized with 100% accuracy. The characters ਸ਼, ਖ਼, ਜ਼, ਫ਼, ਲ਼ are getting confused with similar corresponding characters ਸ, ਖ, ਜ, ਫ, ਲ.

TABLE VIII. CONFUSION MATRIX OF SEPARATE TRAINING AND TESTING DATA WITH LINEAR KERNEL OF SVM

Character	Confused with character				
ੳ	ੳ 100%				
ਅ	ਅ 100%				
ੲ	ੲ 97.08%	ਦ 0.97%	ਮ 1.94%		
ਸ	ਸ 97%	ਮ 0.95%	ਸ਼ 0.95%	ਖ 0.95%	
ਹ	ਹ 99.02%	ਰ 0.97%			
ਕ	ਕ 99.01%	ਤ 0.98%			
ਖ	ਖ 99.01%	ਪ 0.98%			
ਗ	ਗ 98.01%	ਚ 0.99%	ਗ਼ 0.99%		
ਘ	ਘ 100%				
ਙ	ਙ 100%				
ਚ	ਚ 99.01%	ਢ 0.98%			
ਛ	ਛ 53.39%	ਚ 45.6%	ਜ 0.097%		
ਜ	ਜ 97.05%	ਜ਼ 2.95%			
ਝ	ਝ 99.02%	ਚ 0.97%			
ਞ	ਞ 94.23%	ਧ 0.96%	ਵ 4.8%		
ਟ	ਟ 100%				
ਠ	ਠ 95.09%	ਗ0.98%	ਤ 0.98%	ਨ 2.94%	
ਡ	ਡ 98.03%	ਝ 0.98%	ਙ 0.98%		
ਢ	ਢ 90.19%	ਚ 5.88%	ਦ 0.98%	ਫ 2.94%	
ਣ	ਣ 98.04%	ਟ 0.96%			
ਤ	ਤ 98.05%	ਚ 0.97%	ਜ 0.97%		

Character	Confused with character				
ਥ	ਥ 93.20%	ਧ 0.97%	ਕ 0.97%	ਢ 0.97%	ਬ 3.9%
ਦ	ਦ 89.1%	ਕ 2.97%	ਜ 6.93%	ਥ 0.99%	
ਧ	ਧ 86.13%	ਪ 0.99%	ਯ 12.87%		
ਨ	ਨ 85.43%	ਠ 14.5%			
ਪ	ਪ 100%				
ਫ	ਫ 89.32%	ਚ 1.94%	ਢ 0.97%	ਰ 2.9%	ਫ਼ 4.8%
ਬ	ਬ 100%				
ਭ	ਭ 100%				
ਮ	ਮ 99.02%	ਸ 0.98%			
ਯ	ਯ 99.02%	ਧ 0.98%			
ਰ	ਰ 91.34%	ਗ 4.80%	ਚ 1.92%	ਦ 0.96%	ਫ਼ 0.96%
ਲ	ਲ 100%				
ਵ	ਵ 80.76%	ਘ 2.88%	ਢ 1.92%	ਫ਼13.46%	ਭ 0.96%
ੜ	ੜ 100%				
ਸ਼	ਸ਼ 94.77%	ਸ 5.82%			
ਖ਼	ਖ਼ 96.07%	ਖ 3.92%			
ਗ਼	ਗ਼ 100%				
ਜ਼	ਜ਼ 88.23%	ਜ 11.7%			
ਫ਼	ਜ਼ 85.43%	ਚ 10.6%	ਫ 2.9%	ਵ 0.97%	
ਲ਼	ਲ਼ 96.7%	ਲ 1.29%	ਢ 2.59%		

D. *Separate Training and Testing Data with KNN Classifier*

The KNN classifier with training and testing data different has been used in this section. The results are shown in a table for different values of K. Recognition accuracy is obtained by dividing correctly recognized images with total number of images i.e. 4182.

TABLE IX. RECOGNITION ACCURACY OF SEPARATE TRAINING AND TESTING DATA WITH KNN

Feature extraction	K	Correctly recognized	Accuracy	Time elapsed (in seconds)
Zoning	1	4041	96.62%	1410
Zoning	3	4036	96.51%	1439
Zoning	5	4003	95.71%	1571

V. CONCLUSION AND FUTURE SCOPE

In this paper the isolated Gurmukhi pre segmented noise free images have been used as input. Zoning method is used to extract the features. KVM and SVM classifiers are used. When training and testing data are same or when training and testing data are different, in both cases the highest accuracy is achieved with KNN classifiers. Speed is better in SVM classifier as compared to KNN. The work presented in this paper can be further extended by taking all upper zone, lower zone and other half characters. Only 41 middle zone characters in this research. The probabilities found in the SVM classifier can be further used to achieve higher accuracy in recognition of text from images. Using probabilities the wrongly classified characters can be corrected in post processing.

REFERENCES

[1] D. Sharma and G. S. Lehal, “An Iterative Algorithm for segmentation of Isolated Handwritten Words in Gurmukhi Script”, in proceedings of the 18th International Conference on Pattern Recognition (ICPR), vol. 2, 2006, pp. 1022-1025,.

[2] G. S. Lehal and Chandan Singh, “A Complete Machine Printed Gurmukhi OCR System”, Vivek, vol. 16, 2006. pp. 10-17,

[3] J.Ohya, A. Shio, S. Akamatsu,” Recognizing Characters in Scene Images”, IEEE transactions on pattern analysis and machine intelligence. vol. 16, no. 2, 1994, pp. 214-220,

[4] J. T. Tou and R. C. Gonzalez, “Recognition of Handwritten Characters by Topological Feature Extraction and Multilevel Categorization”,Computers, IEEE Transactions, vol. C-21, 1972, pp. 776 – 785,

[5] P. Jhajj, D. Sharma, “Recognition of Isolated Handwritten Characters in Gurmukhi Script”, International Journal of Computer Applications, vol. 4, no. 8, 2010, pp. 9-17.

[6] S. Kaur and G.S Josan, “Gurmukhi Text Extraction From Image Using Support Vector Machine (SVM) ”, International Journal of Engineering Research and Applications (IJERA) ISSN : 0975-5462 vol. 3,no. 4, 2011,pp. 9-15,

[7] S.W. Lee, D. Lee, H. Park, “A new methodology for gray-scale character Segmentation and recognition”, IEEE transaction on pattern analysis and machine intelligence,vol. 31, 1996,pp. 1264-1271,

Robust Threshold Estimation for Image Denoising using Wavelet Transforms

Amandeep Singh
Department of ECE
CTIEMT
Jalandhar, Punjab
amanjot1988@gmail.com

Alok Kumar
Department of ECE
CTIEMT
Jalandhar, Punjab
Kalok2705@gmail.com

Amandeep singh
Department of ECE
CTIEMT
Jalandhar, Punjab
Er.amansaini88@gmail.com

Vishav Kapoor
Department of ECE
CTIEMT
Jalandhar, Punjab
vishavkapoor55@gmail.com

***Abstract*--This paper deals with the threshold estimation method for image denoising in the wavelet transform domain. The proposed technique is based upon the discrete wavelet transform analysis where the algorithm of wavelet threshold is used to calculate the value of threshold. The proposed method is more efficient and adaptive because the parameter required for calculating the threshold based on sub band data. The threshold value is computed by x 2w0 / where x is the scale parameter which depends upon the sub band size and number of decomposition and w0 is the noise variance estimation. are the wavelet coefficient variance estimation in various sub bands. Experimental results on several test images are compared with popular denoise technique from three aspects (PSNR, RMSE and CoC).**

Keywords--Wavelet Thresholding, Image Denoising

I. INTRODUCTION

An image is a two dimension function f(x,y), where x and y are plane coordinates, and the amplitude off at any pair of coordinates(x, y) is called the gray level or intensity of the Image at that point. There are two types of images i.e. gray scale image and RGB image. An image is often corrupted by noise in its acquition and transmission. Basically image noise is unwanted fluctuations. There are different types of image noise present in the image like Gaussian noise, salt and pepper noise, speckle noise, shot noise, white noise [1]. Image denoise is used to remove the additive noise while retaining as much as possible the important signal features. There is various noise reduction techniques which are used for removing the noise. Most of the standard algorithm used to denoise the noisy image and perform the individual filtering process. Denoise generally reduce the noise level but the image is either blurred or over smoothed due to losses like edges or lines. In the recent years there has been a fair amount of research on wavelet thresholding and threshold section for image denoising [3], because wavelet provides an appropriate basis for seprating noisy signal from the image signal. Wavelet transform is good at energy compaction, the small coefficient are more likely due to noise and large coefficient due to important signal feature [8]. These small coefficients can be thresholded without affecting the significant features of the image.

The wavelet transform (WT) is a powerful tool of signal processing for its multiresolutional possibilities. Unlike the Fourier transform, the wavelet transform is suitable for application to non-stationary signals with transitory phenomena, where frequency response varies in time [2]. The wavelet coefficient represents a measure of similarity in the frequency content between a signal and a chosen wavelet function [2]. These coefficient are computed as a convolution of the signal and the scaled wavelet function, which can be interpreted as a dilated band pass filter because of its band pass like spectrum [5]. By wavelet analysis from a signal at high scales, extracted global information called approximations, and at two scales, extracted fine information called details. The discrete wavelet transform (DWT) requires less space utilizing the space saving coding based on the fact that wavelet families are orthogonal or biorthogonal bases, and thus do not produce redundant analysis. The discrete wavelet transform corresponds to its continuous version sampled usually on a dyadic grid, which means that the scales and translations are power of two [5]. Thresholding is a simple non-linear technique, which operates on one wavelet coefficient at a time. In its most basic form, each coefficient is thresholded by comparing against threshold. If the coefficient is smaller than threshold then it set to be zero; otherwise it is kept or modified. We replace the small noisy coefficient by zero and inverse wavelet transform on the result may lead to reconstruction with the essential signal characteristics and with less noise. Since the work of Donoho and Johnstone [4], [9], [10], there has been much research on finding thresholds, however few are specifically designed for images.

II. IMAGE DENOISING

In image processing, wavelets are used for instance for edges detection, watermarking, texture detection, compression, denoising, and coding of interesting features for subsequent classification

A. Wavelet Denoising Principle

i). Image denosing based on the wavelet transform is mainly completed by wavelet thresholding in wavelet domain [12]. The processing of image denosing in wavelet domain can be considered as an optimal estimation to the input image with noise data using the threshold. The wavelet thresholding for image denoising involves the following steps [6]: 1. The wavelet coefficients can be obtained by using the wavelet decomposition on the input image with noise. w = W (X + N)

ii). The optimal estimations are acquired by modifying the wavelet coefficients based on a rule of wavelet threshold. Where is the optimal estimation of the wavelet coefficients; (w) is a wavelet threshold function; is the threshold.

iii). The denoised image can be got by wavelet inverse transform on the modified wavelet coefficients. Based on the above analysis, obviously, there are two problems existing in this research of the wavelet thresholding denoising on image: a: The choice of the wavelet thresholding function b: The choice of the wavelet threshold Our research revolves around the above two main questions. To compute the two-dimensional DWT of an image, we decompose the approximations at level j to obtain four matrixes of coefficients at level j +1.

III. SOFT AND HARD THRESHOLDING

Signal denoising using the DWT consists of the three successive procedures, namely, signal decomposition, thresholding of the DWT coefficients, and signal reconstruction. Firstly, we carry out the wavelet analysis of a noisy signal up to a chosen level N. Secondly, we perform thresholding of the detail coefficients from level 1 to N. Lastly, we synthesize the signal using the altered detail coefficients from level 1 to N and approximation coefficients of level N [2]. However, it is generally impossible to remove all the noise without corrupting the signal. As for thresholding, we can settle either a level-dependent threshold vector of length N or a global threshold of a constant value for all levels. According to D. Donoho's method, the threshold estimate for denoising with an orthonormal basis is given by where the noise is Gaussian with standard deviation of the DWT coefficients and L is the number of samples or pixels of the processed signal or image. This estimation concept is used by Matlab. From another point of view, thresholding can be either soft or hard. Hard thresholding zeroes out all the signal values smaller than . Soft thresholding does the same thing, and apart from that, subtracts from the values larger than . In contrast to hard thresholding, soft thresholding causes no discontinuities in the resulting signal. In Matlab, by default, soft thresholding is used for denoising and hard thresholding for compression [2].

IV. WAVELET THRESHOLD

The traditional wavelet threshold is Donoho wavelet threshold which is shown as formula

$$\gamma = \sqrt{2\sigma} * 2\log N \tag{1}$$

Where s is the estimate variance, N is the number of the wavelet coefficients. However, this wavelet threshold has a fatal shortcoming. Because this threshold value is proportional with the signal size logarithm's square root, therefore, when N is big, the threshold value will tend to reset the while wavelet coefficients, as a result, the wavelet filter will degenerate as the degradation of low-pass filter. And there is a serious "overkill" wavelet coefficients tendency.

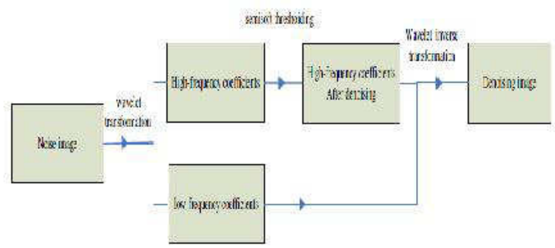


Fig.1 Wavelet Threshold Technique

Due to the above deficiency of Donoho's threshold, the more effective Bayes wavelet threshold is used to optimize the performance of the wavelet threshold denoising algorithm in this study with assumption of the wavelet coefficients obeying a distribution. The threshold According to the Bayes Rule is shown formula s are various thresholds in various sub bands. D is the noise variance estimation. s are the wavelet coefficients variance estimations in various sub-bands. Because the main information of the noise images concentrates on the wavelet high-frequency (low-scale) sub band, the median of the magnitudes of the all diagonal details coefficients in the wavelet low-scale sub-band is used to calculate the value of the noise variance estimation, which is showed as formula Where w are the wavelet high-frequency diagonal coefficients. The wavelet coefficients variance estimations in various sub-bands can be calculated as formula: - Where n are the variance estimations of the coefficients in various sub-bands. mxn is the number of the corresponding coefficient wn in different sub-bands. The wavelet bayes threshold considers the distribution characteristics of the wavelet coefficients, hence the properties of this threshold is more excellent than the traditional wavelet threshold. Based on the above section the improved method based on the wavelet thresholding function has many advantages and can obtain the great result.

V. IMAGE DENOISE ALGORITHM

This section describes the image denoising algorithm, which achieves near optimal semi soft threshholding in the wavelet domain for recovering original signal from the noisy one. The algorithm is very simple to implement and computationally more efficient. It has following steps:

- Perform multiscale decomposition of the image corrupted by guassian noise using wavelet transform.
- Estimate the noise variance
- For each level, compute the scale parameter
- Compute the standard deviation
- Compute threshold
- Apply semi soft thresholding to the noisy coefficients.
- denoise high frequency coefficient
- merge low frequency coefficient with denoise high frequency coefficient
- Invert the multiscale decomposition to reconstruct the denoised image.

VI. PARAMETRIC DESCRIPTION

A. Algorithm for Peak Signal to Noise ratio (PSNR)

Step1: Difference of noisy image and noiseless image is calculated using imsubract Command.

Step2: Size of the matrix obtains in step 1 is calculated.

Step3: Each of the pixels in the matrix obtained in step issquared.

Step4: Sum of all the pixels in the matrix obtained in Step3 is calculated. Step5: (MSE) is obtained by taking the ratio of value obtained in step 4 to the value obtained in the Step2

Step6: (RMSE) is calculated by taking square root to the value obtained in Step5.

Step7: Dividing 255 with RMSE, taking 1log base 10 and multiplying with 20 gives the value of PSNR.

B. Algorithm for Correlation of Coefficient (Coc)

Step1: Mean of the noiseless image and noisy image are calculated.

Step2: Mean of the noiseless image is subtracted from each of the pixel in the noiseless image resulting in a matrix.

Step3: Similarly the mean of noisy image is subtracted from each of the pixels in the noise image resulting in a matrix.

Step4: Values obtained in Step2 and Step3 are multiplied.

Step5: Sum of all the elements in the matrix obtained in Step4 is calculated. Step6: Square of all the elements of the matrix obtained in Step2 is calculated and sum of this squared matrix is determined.

Step7: Similarly square of all the elements of the matrix obtained in Step3 is calculated and sum of the elements of this squared matrix is also determined.

Step8: Values obtained in Step6 and Step7 are multiplied and its square root is taken.

Step9: Ratio of the value obtained in Step5 to the value obtained in Step8 is calculated.

C. Algorithm for Root Mean Square Error (RMSE)

Step1: Difference of noisy image and noiseless image is calculated using imsubract command.
Step2: Size of the matrix obtains in Step1 is calculated.
Step3: Each of the pixels in the matrix obtained in step is squared.
Step4: Sum of all the pixels in the matrix obtained in step 3 is calculated. Step5: (MSE) is obtained by taking the ratio of value obtained in Step4 to the value obtained in the step 2.
Step6: RMSE) is calculated by taking square root to the value obtained in Step5.

VII. 2-D DECIMATED DISCRETE WAVELET TRANSFORM IMAGE DENOISING USING THRESHOLD APPROACHES

You can generate MATLAB code to reproduce GUI-based 2-D decimated wavelet denoising at the command line. You must perform this operation in the Wavelet 2-D – – De-noising tool.

You must first denoise your image before you can enable the File > Generate Matlab_Code (Denoising Process) operation.

- Enter wave menu at the MATLAB command prompt.
- Select **Wavelet 2-D**.
- Load the Noisy SinSin example indexed image. Using the default biorthogonal wavelet and level 3 decomposition, click **De-noise**.
- In the Select **thresholding method** drop-down menu, select the default Fixed form threshold and **soft** options. Use the default Unscaled white noise. Set the thresholds by level for the horizontal, diagonal, and vertical coefficients as follows:
- Level 3 — 4

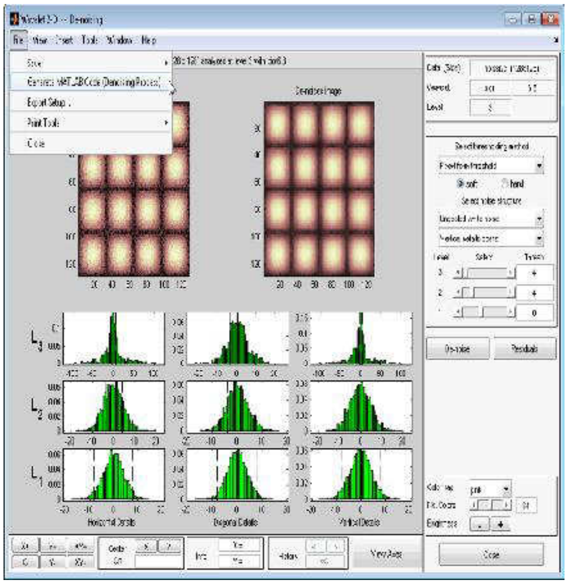


Fig.2 Shows Denoising process is performed via software

- Level 2 — 4
- Level 1 — 8
- Enter these thresholds for the horizontal, diagonal, and vertical coefficients.
- Select De-noise.
- Generate the MATLAB code with File > Generate Matlab Code (Denoising Process).

A). The operation generates the following MATLAB code.

function [XDEN,cfsDEN,dimCFS] = func_denoise_dw2d(X)

FUNC_DENOISE_DW2-D Saved Denoising Process.

X: matrix of data

XDEN: matrix of denoised data
cfsDEN: decomposition vector (see WAVEDEC2)
dimCFS: corresponding bookkeeping matrix
Analysis parameters.
----
wname = 'bior6.8';
level = 3;
Denoising parameters.
----
meth = 'sqtwolog';
scal_OR_alfa = one;
sorh = 's';   % Specified soft or hard thresholding
thrParams = [...

8.00000000   4.00000000   4.00000000 ; ...
8.00000000   4.00000000   4.00000000 ; ...
8.00000000   4.00000000   4.00000000   ...];

roundFLAG = true;

Denoise using CMDDENOISE.

[coefs,sizes] = wavedec2(X,level,wname);
[XDEN,cfsDEN,dimCFS] wdencmp('lvd',coefs,sizes, ...
wname,level,thrParams,sorh);
if roundFLAG , XDEN = round(XDEN); end
if isequal(class(X),'uint8') , XDEN = uint8(XDEN); end

B.) Save func_denoise_dw2d.m in a folder on the MATLAB search path, and execute the following code.

load noissi2d.mat;
noissi2d = X;
[XDEN,cfsDEN,dimCFS] func_denoise_dw2d(noissi2d);

C). Save your denoised image in a folder on the MATLAB search path as denoisedsin.mat.

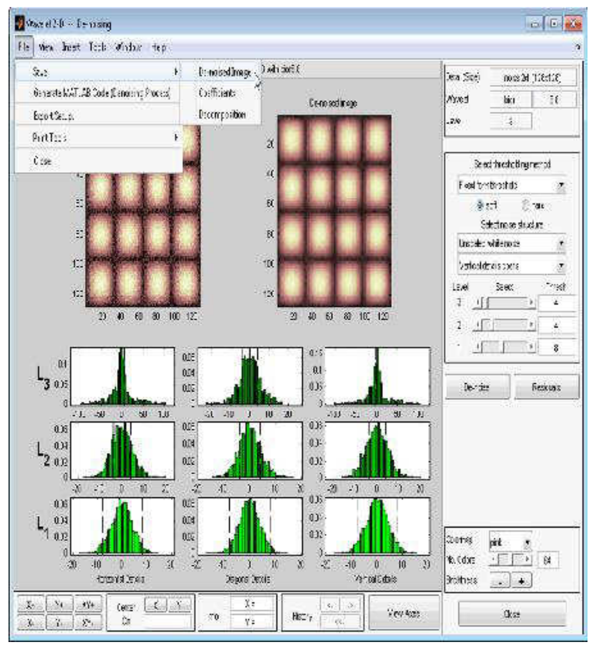


Fig.3 Shows to save Denoising process performed via software

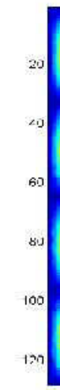
Load the denoised image in the MATLAB workspace. Compare the result with your generated code.

load denoisedsin.mat;

% denoised image loaded in variable X%

subplot(121);
imagesc(X); title('Image denoised in the GUI');
subplot(122);
imagesc(XDEN); title('Image denoised with generated code');

% Norm of the difference is zero



norm(XDEN-X,2)

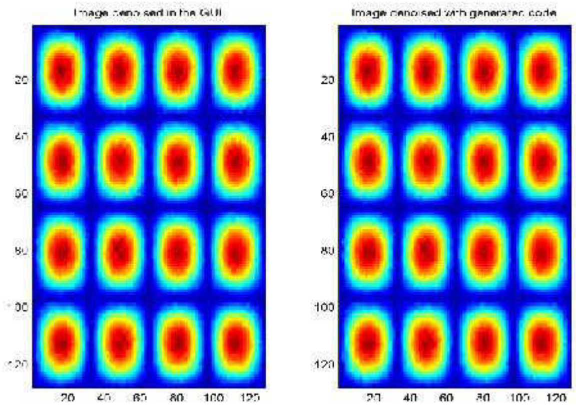


Fig.4 Shows the Image denoising with GUI and with generated codes

VIII. CONCLUSION

In this paper, a simple and sub band semi soft threshold method is proposed to address the issue of image recovery from its noisy counterpart. It is based on the discrete wavelet transform and Gaussian distribution modeling of subband coefficients. The image denoises algorithm uses semi thresholding to provide smoothness and better image details preservation. The wavelet semi soft thresholding denoise algorithm produce overall better PSNR, MSNR and COC result compared with other traditional denoise approaches.

REFERENCES

[1] Kazubek, M , "Wavelet domain image de-noising by thresholding and Wiener filtering".Signal Processing Letters IEEE, Volume: 10, Issue: 11, pp 265 Vol.3, Nov. 2003.

[2] G. Oppenheim J. M. Poggi M. Misiti, Y. Misiti." Wavelet Toolbox", The MathWorks, Inc.,Natick, Massachusetts 01760, April 2001.

[3] S. Grace Chang, Bin Yu and M. Vattereli, "Adaptive Wavelet Thresholding for Image Denoising and Compression", IEEE Trans. Image Processing, vol. 9, pp. 1532-1546, Sept. 2000.

[4] D.L. Donoho and I.M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage", Journal of American Statistical Assoc., Vol. 90, no. 432, pp 1200-1224, Dec. 1995.

[5] C. Valens.,"A really friendly guide to wavelets" eBook, 2004

[6] D.L.Donono, I.M.Johnstone.,"Adaptive to unknown smoothness via wavelet shrinkage",Journal of the American Statistical Association., Vol.90, No.432, pp.1200-1224, 1995.

[7] Chen juan, Chen qian-hui, Shi lu-huan," Edge detection technology in image tracking",Chinese Journal of Optics and Applied Optics., Vol.2. No.1, pp.46-53.february 2009.

[8] Maarten Jansen, "Noise Reduction by Wavelet Thresholding", Springer– Verlag New York, Inc.-2001.

[9] D.L. Donoho and I.M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage Biometrica", Vol. 81, pp. 425-455,1994.

[10] D.L. Donoho and I.M. Johnstone, "Wavelet shrinkage: Asymptopia", Vol. 57, no. 2, pp. 301-369, 1995.

[11] Xu Bao Guo,Wng ji.,"A new adaptive edge detection algorithm based on mathematic morphology" Journal of China Application, Vol.29, No.4, pp.997-1000,2012

A New Method for Salt and Pepper Noise Reduction for Gray Scale Images

Mandeep Singh
Dept. of Computer Applications
CT Group of Institutions, Shahpur, Jalandhar
mandeepdalli@gmail.com

Amarjit Singh
Dept. of Computer Applications
CT Group of Institutions, Shahpur, Jalandhar
amarjit.rhce@gmail.com

Vikas Attri
Dept. of Computer Applications
CT Group of Institutions, Shahpur, Jalandhar
vikas_apj@hotmail.com

Shival Chadda
Dept. of Computer Applications
CT Group of Institutions, Shahpur, Jalandhar
shival.chadda@gmail.com

***Abstract*- Digital images can be corrupted by additive impulse noise due to errors in the communication channels during transmission or due to faulty sensors during image acquisition. Image noise is a fluctuation of pixel values. Impulse noise removal in image processing is an important pre-processing step which involves the removal of salt and pepper noise from digital images to restore the images. In salt and pepper noise, the noisy pixel value will be either 0 or 255. Salt and pepper noise involves changing a part of the pixel values with random ones[1]. The median filter is one of the most popular nonlinear filter for removing salt and pepper noise because of its good denoising power and computational efficiency. The main aim of Image removal is to remove noise, preserve edges with good quality image.The performance of each algorithm is compared by computing Peak Signal to Noise Ratio (PSNR) and Mean Squared error (MSE)[4]. In case of image denoising methods, the characteristics of the degrading system and the noises are assumed to be known beforehand. Basically noise removal from the images is done with the help of various filters such as mean filter, geometric filter, harmonic mean filter, contra harmonic filter, median filter, max and min filter and alpha trimmed filter.**

Keywords: Tolerance Value, Peak Signal to Noise Ratio (PSNR), Arithmetic Mean Filtering (AMF), Mean Square Error (MSE).

I. INTRODUCTION

The field of image processing focuses on automating the process of gathering and processing visual information. The process of receiving and analyzing visual information by digital computer is called *digital image processing[1]*. An image may be described as a two-dimensional function Z.

$$Z = f(x, y)$$

where x and y are spatial coordinates. Amplitude of f at any pair of coordinates $(x; y)$ is called intensity Z or gray value of the image. When spatial coordinates and amplitude values are all finite, discrete quantities, the image is called digital image.

II. EXISTING MEAN FILTERING TECHNIQUES

To recover the image from its noise there exits many mean filtering techniques which are application oriented. Some filtering techniques have better effects than the others according to noise category. Mean filtering techniques are described below.

A. *Arithmetic mean filter*

This is the simplest of the mean filters[5]. Let S_{xy} represent the set of coordinates in a rectangular sub image window of size m X n, centered at point (x, y).The arithmetic mean filtering process computes the average value of the corrupted image g(x, y) in the area defined by S_{xy}.The value of the restored image at any point (x, y) is simply the arithmetic mean computed using the pixels in the region defined by S. In other words

$$\hat{f}(x, y) = \frac{1}{mn} \sum_{(s,t) \in S_{xy}} g(s,t)$$

This operation can be implemented using a convolution mask in which ail coefficients have value 1/mn. Mean filter simply smoothes local variations in an image. Noise is reduced as a result of blurring.

B. *Contra harmonic mean filter*

The contra harmonic mean filtering[5] operation yields a restored image based on the expression

$$\hat{f}(x, y) = \frac{\sum_{(s,t) \in S_{xy}} g(s,t)^{Q+1}}{\sum_{(s,t) \in S_{xy}} g(s,t)^{Q}}$$

where Q is called the order of the filter. This filter is well suited for reducing or virtually eliminating the effects of salt-and-pepper noise. For positive values of Q, the filter eliminates pepper noise. For negative values of Q it eliminates salt noise. It cannot do both simultaneously. Note that the contra harmonic filter reduces to the arithmetic mean filter if Q = 0, and to the harmonic mean filter if Q = - 1

C. Median filter

The best-known order-statistics filter is the median filter[4], which, as its name implies, replaces the value of a pixel by the median of the gray levels in the neighborhood of that pixel:

$$\hat{f}(x, y) = \underset{(x, y) \in S_{xy}}{median} \{g(s, t)\}$$

The original value of the pixel is included in the computation of the median. Median filters are quite popular because, for certain types of random noise, they provide excellent noise-reduction capabilities, with considerably less blurring than linear smoothing filters of similar size. Median filters are particularly effective in the presence of both bipolar and unipolar impulse noise.

D. Max and min filters

Although the median filter[5] is by far the order-statistics filter most used in image processing. it is by no means the only one. The median represents the 50th percentile of a ranked set of numbers, but the reader will recall from basic statistics that ranking lends itself to many other possibilities. For example, using the 100th percentile results in the so-called max filter given by:

$$\hat{f}(x, y) = \underset{(s, t) \in S_{xy}}{\max} \{g(s, t)\}$$

This filter is useful for finding the brightest points in an image. Also, because pepper noise has very low values, it is reduced by this filter as a result of the max selection process in the sub image area S. The 0th percentile filter is the Min filter.

$$\hat{f}(x, y) = \underset{(s, t) \in S_{xy}}{\min} \{g(s, t)\}$$

III. PROPOSED APPROACH

To avoid the problems that are visible by the existing mean filtering techniques, we have developed an algorithm, which is a modification of AMF. Our proposed algorithm is described below:

A. Arithmetic Mean Filtering Technique based on tolerance

Salt and Pepper noise is considered as the extreme case among all types of noises. To recover the images affected by this noise, we have developed a technique called AMFTT[4]. In this technique our main concern is to use the Arithmetic Mean Filtering Technique efficiently to recover from Salt and Pepper noise. We know that for Salt and Pepper noise the pixel value of the noisy image is converted to 0 and 255. When we use Arithmetic Mean Filtering Technique we take 3 × 3 windows and find out the Arithmetic Mean and in this case all the 9 pixels of this 3 × 3 window are used to calculate the Arithmetic Mean. But to calculate mean using the extreme value, provide us with erroneous result in our technique. To avoid this effect we ignored the pixel of value 0 and 255 while calculating the mean. But it may be the case that the pixels of the 3× 3 window represent a black or white object. Hence the pixels are not affected by the noise rather the original values 0 and 255. To deal with this situation, we consider one pixel and a sub window of size 3×3 around that pixel and find out Arithmetic Mean from the pixels of the sub window ignoring the pixels with the maximum (255) and minimum (0) value. If the number of pixels is less than 3 out of 9 (window size m×n) adjacent pixels, we use the traditional Arithmetic Mean Filtering Technique. Otherwise we use the calculated mean found by this technique.

In this technique we have also used a threshold called Tolerance[4]. If the difference between the calculated Arithmetic Mean (excluding the pixels with gray level 0 or 255) and the intensity of the considered pixel is greater than the Tolerance, we replace the intensity of the considered pixel by the Arithmetic Mean. Otherwise the intensity of the considered pixel is unchanged.

B. Proposed Algorithm

1. For q = 0 to 2 repeat steps 2 to 5
2. For each pixel p in the image do

   Take a sub window of size m x n (3 X 3) around the pixel.
3. In the sub – window of 3 X 3 size, count number of 0's and 255.
4. If count is less than 5 (count<5)
   a. Find out the median of 9 pixels of the sub – window

      Else

   b. Find out mean of 9 pixel of sub – window.
5. Repeat steps 2 to 4 for the entire image.

C. The significance of the Tolerance value

For Salt and Pepper noise the value of the distorted pixel is 0 or 255. So we find a significant difference between the mean and the value of the distorted pixel. Replacing only the distorted pixel will provide us with better result than replacing all the pixels. Our Tolerance value ensures that only the distorted pixels are replaced.

If we take Tolerance value 0, it will provide same result as Arithmetic Mean Filtering Technique. If we take small Tolerance value such as 5 or 10, then not only the distorted pixels but also the other pixels are replaced. If we increase the Tolerance value then for Salt and Pepper noise PSNR increases. If we take a very large Tolerance value (like 65 or greater than 65), some distorted pixels are not replaced which decreases the PSNR for Salt and Pepper noise[4].

From our experiments this technique produces a very good result for Salt and Pepper noise when the Tolerance value is 60. Considering a moderate Tolerance value such as 30 to 35, will provide better result than that of minimum Tolerance value (e.g. 5 to 10), but it will provide lower performance than taking high Tolerance value 60.

D. Significance of the Proposed Technique

For Salt and Pepper noise the PSNR obtained by the proposed technique is much higher than that of all other mean filtering techniques and the image is free from blurring effect. If we use a suitable Tolerance value like 60 or around 60, the best result is achieved. If we increase the Tolerance value from that level, the PSNR and image quality both will be decreased.

IV. SIMULATION

A. PSNR Vs Tolerance value

For the simulation purpose we have used, Microsoft Visual C++ and Paint Shop Pro. In our proposed approach, the PSNR of the filtered image varies with the Tolerance value. Whenever we take a smaller Tolerance value we obtain comparatively lower PSNR. The PSNR Value increases along with the increase of the Tolerance value up to a certain level. Then the PSNR decreases again. From table1 we can see, when the Tolerance value is 0 we get PSNR[4] 24.5621 dB, PSNR value increases from that stage over the sample image. For the Tolerance value 60 we obtain the highest PSNR 25.4341. After that PSNR decreases again.

(a) (b) (c)

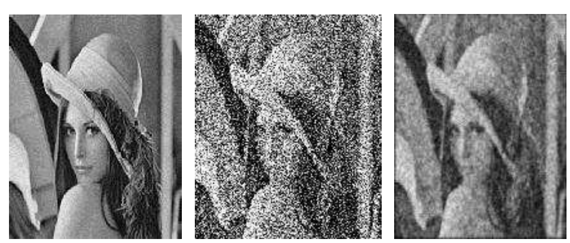
(d) (e)

Restoration of original image 4.1(a) Original image, (b) Corrupted image with Salt & pepper noise (30%), (c) Image recovered by Mean Filter technique (d) Image recovered by Median Filter technique (e) Image recovered by Proposed Technique

(a) (b) (c)

(d) (e)

Restoration of original image 4.2 (a) Original image, (b) Corrupted image with Salt & pepper noise (70%), (c) Image recovered by Mean Filter technique (d) Image recovered by Median Filter technique (e) Image recovered by Proposed Technique



(a) (b) (c)

(d) (e)

Restoration of original image 4.3 (a) Original image, (b) Corrupted image with Salt & pepper noise (80%), (c) Image recovered by Mean Filter technique (d) Image recovered by Median Filter technique (e) Image recovered by Proposed Technique

PSNR, MSE, MAE Value Result:

Table 4.1: PSNR for various filters

Image Noise(%)	Mean Filter	Median Filter	Proposed Algorithm
10%	25.14	30.81	30.84
20%	26.41	32.70	32.76
30%	24.90	29.52	29.72
40%	18.60	19.72	19.73
50%	26.51	33.12	34.03
60%	22.25	25.68	25.71
70%	22.47	24.45	24.59
80%	21.24	21.43	21.68

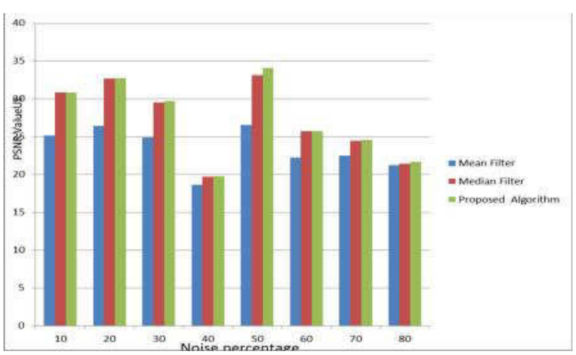


Figure4.5: Graph PSNR vs Noise percentage

Table 4.2: MSE for various filters

Image Noise (%)	Mean Filter	Median Filter	Proposed Algorithm
10%	199.06	53.85	53.53
20%	148.50	34.89	34.40
30%	210.30	72.46	69.32
40%	896.63	692.28	690.28
50%	145.01	31.65	25.65
60%	386.69	175.60	174.29
70%	368.06	233.29	225.67
80%	488.56	467.43	441.30

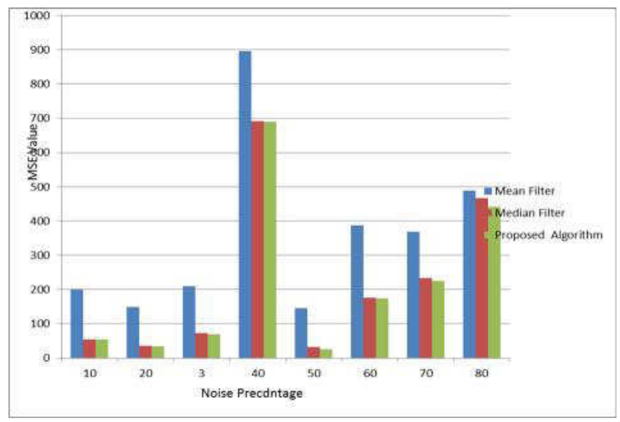


Figure 4.6: Graph MSE vs Noise percentage

Table 4.3: MAE for various filters

Image Noise(%)	**Mean Filter**	**Median Filter**	**Proposed Algorithm**
10%	7.95	3.13	3.13
20%	7.50	2.97	2.88
30%	10.2	4.08	4.04
40%	19.75	13.18	13.20
50%	9.28	2.1	1.98

60%	13.6	6.84	6.82
70%	14.11	8.67	8.55
80%	16.82	14.35	13.93

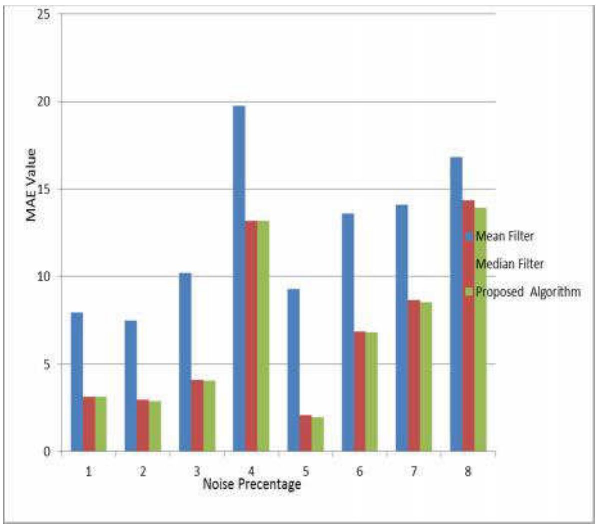


Figure 4.7 Graph MAE vs Noise percentage

V. CONCLUSION AND FUTURE PLAN

In this paper we have developed a new filtering technique, which is better than the existing Mean Filtering Techniques and in some cases Median Filtering Technique for Salt and Pepper noise. The proposed algorithm is a simple and fast method to remove the salt and pepper noise in digital images. The performance of the new filter is tested for various noise corrupted gray scale images. The restoration results of the proposed filter are compared with standard methods and existing method. The performance of the standard methods is poor in terms of noise smoothening and detail preservation but the proposed algorithm removes salt and pepper noise with edge preservation for low to high salt and pepper noise corrupted images. The PSNR, MSE and MAE values of various methods are shown in table 4.1, table 4.2, and table 4.3 and graphically in figure 4.5, figure 4.6 and figure 4.7. In our implementation the PSNR value of proposed method is greater than existing method and MSE value is less than the existing method. We have successfully proposed a new algorithm to remove salt and pepper noise from the gray scale images. This algorithm achieves better results than other existing methods. Our results are compared with the Existing method's results, which show that our method is more efficient and fast than Existing methods.

In future the method can be modified to use for color images to remove salt and pepper noise. Also the proposed method can be further modified to achieve better result and better edge preservation[2].

The proposed algorithm result can be shown in the following example. If you are using a Median Filter then value will be same on this example it's a drawback for the Median filter Technique Noisy Image Segment Proposed result.

$$\begin{pmatrix} 0 & 50 & 0 \\ 30 & 0 & 40 \\ 0 & 60 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 50 & 0 \\ 30 & 20 & 40 \\ 0 & 60 & 0 \end{pmatrix}$$

0	0	0	0	0	30	40	50	60

Figure 4.12: showing the sorted list of the 3×3

REFERENCES

[1] Alok Singh, Umesh Ghanekar,Chakresh Kumar, Ghanendra Kumar (2011), " An Efficient Morphological Salt-and-Pepper Noise Detector" Int. J. Advanced Networking and Applications 873 Volume: 02, Issue: 05, Pages: 873-875.

[2] Chih-Yu Hsu , Ta-Shan Tsuib, Shyr-Shen Yub, and Kuo-Kun Tseng(2011)," Salt and Pepper Noise Reduction by Cellular Automata", International Journal of Applied Science and Engineering

[3] Yu-Mei Huang, Michael K. Ng, and You-Wei Wen (2009), "Fast Image Restoration Methods for Impulse and Gaussian Noises Removal", ieee signal processing letters, vol. 16, no. 6.

[4] Krishnan Nallaperumal, Justin Varghese,Saudia, Santhosh.P.Mathew, K.Krishnaveni, and Annam, (2006), "Adaptive Rank-ordered Switching Median Filter for Salt & Pepper Impulse Noise Reduction".

[5] Krishnan Nallaperumal, Justin Varghese, K.Krishnaveni, Santhosh. P. Mathew (2006), "An efficient Switching Median Filter for Salt & Pepper Impulse Noise Reduction ",Digital Information Management International Conference.

Accuracy Enhancement Technique for Pseudo Zernike Moments

Monika Badhan
Dept. of ECE
Sri Guru Granth Sahib World University
Fatehgarh Sahib, India

Kamaljit Singh Bhatia
Dept. of ECE
Sri Guru Granth Sahib World University
Fatehgarh Sahib, India

Rahul Upneja
Dept. of ECE
Sri Guru Granth Sahib World University
Fatehgarh Sahib, India

Harsimrat Kaur
Dept. of ECE
CTIEMT
Jalandhar, India

Abstract **Pseudo Zernike moments (PZMs) suffer from geometric error and numerical integration error. Geometric error is caused when a square image is mapped into a unit circular disc which can not exactly match the square domain. Numerical integration error is caused when the integration is approximated by zeroth order summation. We propose methods which reduce these errors. The propose methods are similar to the numerical integration technique which divides the pixel into subparts. The effect of the improved accuracy in the pseudo Zernike moments is also analyzed for reconstruction and numerical stability. While improving the accuracy, the speed becomes slow. Therefore, an attempt is also made to enhance the speed.**

Keywords Pzeudo Zernike moments; Zernike moments; Orthogonal fourier mellin moments.

I. INTRODUCTION

Pseudo Zernike moments (PZMs) are one of the most prominent orthogonal rotation invariant moments (ORIMs), which are known as one of the best region based shape descriptor [1]. Due to the orthogonal property, PZMs exihibit many useful characteristics such as they possess minimum information redundancy and immunity to noise image. PZMs are defined over a unit circle due to its polar representation. PZMs have more low order moments than other ORIMs for the same order, therefore, they are less sensitive to image noise [2]. Also it provides better image representation than Zernike moments (ZMs) and Orthogonal fourier mellin moments (OFMMs) for small scale images. These features of PZMs make them suitable for image processing and pattern recognition applications such as face recognition [3], content based image retrieval [4], image watermarking [5], audio watermarking [6], map matching [7], palmprint verification [8], edge detection [9] and image reconstruction [10].

PZMs suffer from various errors which are generated during mapping and computation process. The most prominent errors in PZMs are geometric error and numerical integration error. These errors have been analyzed in detail by Liao and Pawlak [11] for ZMs and Singh and Upneja [1] for all ORIMs. They observe that geometric error occurs when a square or rectangular image is mapped into a circular region. This mapping allows only those pixels to take part in moment computation whose center fall inside the unit circle, while discarding other pixels. For the computation of moments, we approximate the circular region by grids, which results in a zig zag pattern as shown in the Fig.1. The second error, that is, numerical integration error is caused when double integration in moment calculation is approximated by zeroth order summation. An approach for the removal of numerical integration error in case of ZMs has been proposed by Xin et al. [12] in which ZMs are calculated in polar coordinates. But their approach requires the conversion of square image pixels into the circular domain resulting in circular grid. Another approach is introduced by Wee and Paramesran [13], in which an outer circle as shown in Fig.1(c), encloses the square image

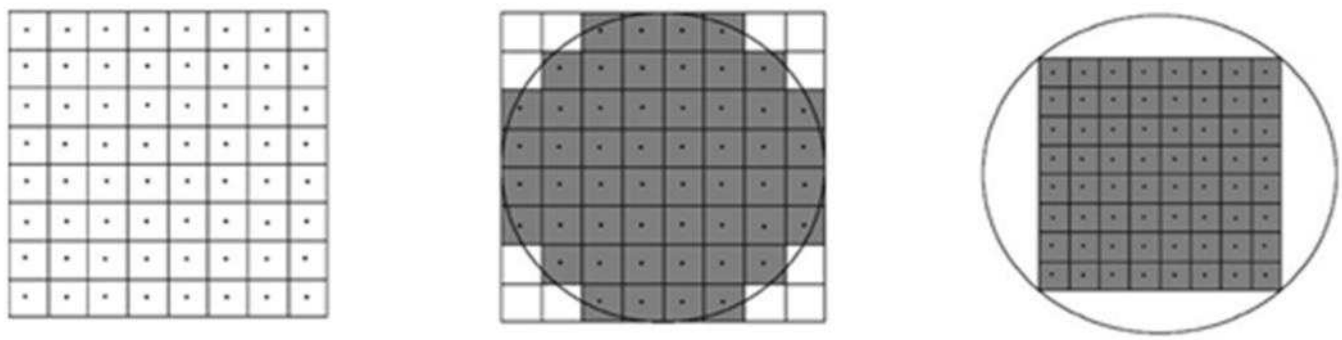


Fig.1. (a) 8x8 original image grid (b) inner circle approximated by square grid (c) outer circle containing sqaure image.

completely to remove the geometric error. Numerical integration error is removed with the help of relationship between geometric moments and Zernike moments after accurately computing the geometric moment. Numerical stability is another major cause of concern for higher order moments in PZMs. Researchers [14, 15], found that, numerical stability can be inproved by using reccursion method. Singh and upneja [2], develop a method for PZMs to reduce the geometric error and numerical integration error using Gaussian numerical integration. However, the method is efficient to reduce both the errors, but it requires the Gaussian points and weights for different order of numerical integration. To overcome this problem, we develop a technique to reduce numerical integration error and geometric error simultaneously by dividing a pixel into k × k sub regions. This method reduces geometric error and numerical integration error simultaneously. We got the motivation for this method from the research of Liao and Pawlak [16] for ORIMs.

The rest of the paper is organized as follows. Section 2 gives the mathematical framework of PZMs. In Section 3, we introduce our purposed scheme for the computation of PZMs. Section 4 devoted for the experimental work. Conclusions are given in Section 5.

II. PSEUDO ZERNIKE MOMENTS

Pseudo Zernike moments are defined inside the unit circle. The two-dimensional Pseudo Zernike moment of order p with repetition q of an image function $f(x, y)$ is given by,

$$A_{pq} = \frac{p+1}{\pi} \iint_{x^2+y^2 \le 1} f(x, y) V_{pq}^{*}(x, y)\, dx\, dy \tag{1}$$

Where, $V_{pq}^{*}(x, y)$ is the complex conjugate function of the pseudo-Zernike polynomials (*PZPs*), and $V_{pq}(x, y)$ is defined as

$$V_{pq}(x, y) = R_{pq}(x, y) e^{jq\theta} \tag{2}$$

Where, $r = \sqrt{x^2 + y^2}$, $j = \sqrt{-1}$, p is a non-negative integer, $|q| \le p$, $\theta = \tan^{-1}(y/x)$, and Radial polynomial for PZMs is given by

$$R_{pq}^{P}(r) = \sum_{s=0}^{p-|q|} \frac{(-1)^s (2p+1-s)!\, r^{p-s}}{s!\,(p+|q|+1-s)!\,(p-|q|-s)!} \tag{3}$$

It is difficult to find the exact solution of double integration given in Eq. (1). Therefore, the computation of *PZMs* is normally carried out using the zeroth order approximation, which leads to the following form

$$\hat{A}_{pq} = \frac{p+1}{\pi} \sum_{\substack{i=0 \\ x_i^2+y_k^2 \le 1}}^{N-1} \sum_{k=0}^{N-1} f(x_i, y_k) V_{pq}^{*}(x_i, y_k)\, \Delta x_i \Delta y_k \tag{4}$$

The point (x_i, y_k) describes the center of the pixel (i, k) and is given by

$$x_i = \frac{2i+1-N}{D}, \quad y_k = \frac{2k+1-N}{D}, \quad i, k = 0, 1, \ldots, N-1 \tag{5}$$

and

$$\Delta x_i = \Delta y_k = \frac{2}{D} \tag{6}$$

where

$$D = \begin{cases} N & \textit{for inscribed circular disk},\ \textit{Fig}.1(b) \\ N\sqrt{2} & \textit{for circular disk containing the image},\ \textit{Fig}.1(c) \end{cases}$$

For the maximum moment order $p_{\max}$, the image function can be reconstructed as

$$\hat{f}(x_i, y_k) = \sum_{p=0}^{p_{\max}} \sum_{q=-p}^{p} A_{pq} V_{pq}(x_i, y_k), \quad i, k = 0, 1, \ldots, N-1 \tag{7}$$

The reconstruction error, MSRE is determined by

$$MSRE = \frac{\sum_{i=0}^{N-1} \sum_{k=0}^{N-1} \left[f(x_i, y_k) - \hat{f}(x_i, y_k) \right]^2}{\sum_{i=0}^{N-1} \sum_{k=0}^{N-1} \left(f(x_i, y_k) \right)^2}, \quad x_i^2 + y_k^2 \le 1 \tag{8}$$

III. PROPOSED METHOD

PZMs are more robust to image noise than other orthogonal rotation invariant moments. Particularly, at low order moments, they exhibit less sensitivity to noise. The geometric errors and numerical integration error are the major errors involved in the computation of moments that severely affects the image moments and hence the reconstructed image is not of good quality. So we proposed a method, in which we divide a pixel into $n \times n$ sub regions with the same weights and include only those subparts whose center fall inside the unit disc.In our method we take different moment order to obtain the reconstruction error. The proposed methods reduce the white spot at the center of the unit disc and increase the accuracy of the reconstructed images. The moments are stable for very high values of p_{max}. Lesser the reconstruction error, better the reconstructed image is. This fact can be seen when we compare our proposed method with the traditionl methods, we found that our proposed method gives better reconstructed image with minimum reconstruction error.

IV. EXPERIMENT ANALYSIS

For our experiments we have chosen a 64×64 size gray scale image as shown in Fig.2. The performance of the reconstructed image is compared with the traditional method. The image is reconstructed for different values of p_{max} =10, 20, 30, 40, 50, 60 and 70 using inner circle mapping. The reconstruction ability of various moments is computed by simply visually inspecting the image and by computing reconstruction error.

Fig.3 shows the reconstructed images using the traditional and proposed methods. Column 2 of Fig.3 shows the reconstructed images using traditional method. The intensity values of reconstructed images are deviated highly from the original image. The reconstruction error at vicinity of the

Fig.2. Original image of size 64×64

center of the circle increases as the moment order p_{max} increases. The result of Column 3 to Column 6 represents our proposed method using the different values of n. It is seen that there is little improvement in the reconstructed image as the white spot reduces as compare with traditional method using $n=3$. However, the other values of n give the better image reconstruction quality. The results of the reconstruction using n=4 and n=5 are far better than the traditional method. Lesser the reconstruction error, better is the quality of reconstructed image. To observe this trend Fig. 4 is drawn for different values of p_{max} usig traditional

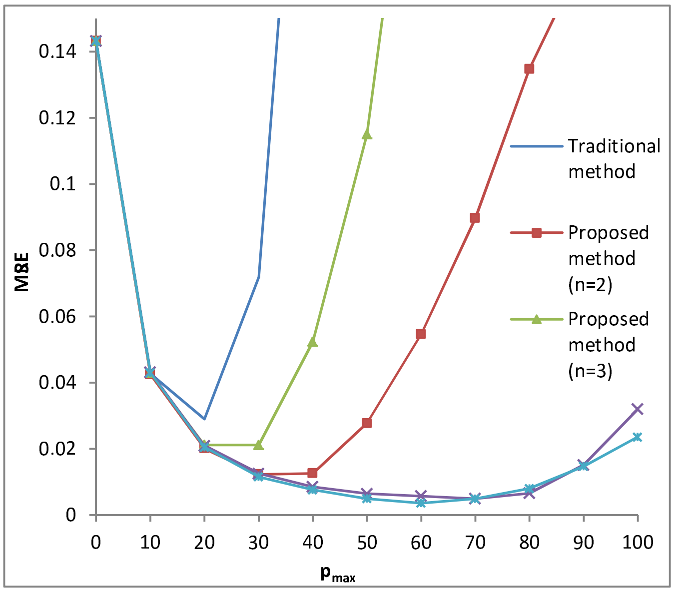


Fig.4. The reconstruction error (MSRE) as a function of order of moments for 64×64 pixels benchmark image.

and proposed method. It is also shown in figure very clearly that, as we move from lower order of moments to high order moments, the reconstruction error keeps on increasing, especially in the case of traditional method, whereas, it remains low in the case of proposed method. It is clearly visible form the Fig.4 that the proposed method is stable for high order of moments as compared to the traditional method. The traditional method is stable when the value of $p_{\max} > 20$. However, proposed method using n=4 and n=5 are stable upto $p_{\max} = 80$. The reconstruction error keeps on decreasing for high value of $p_{\max}$ using proposed method which gives the better image reconstruction quality.Therefore, It is observed that we get better result of reconstructed image quality with the help of proposed method.

V. CONCLUSION

The PZMs are one of the effective region based shape descriptor. They possess minimum information redundancy and immunity to image noise because of presence of orthogonal and complete basis function. They exhibit better image representation than other orthogonal rotation invariant moments for small scale images. The properties of PZMs suffer from various errors such as geometric error, numerical integration error. Therefore, we develop a proposed method in which we divide a complete pixel into sub regions with the same weights. The proposed method is stable for very high values of p_{max}. This method not only reduces above errors but also enhances the accuracy of the reconstructed image. Therefore the proposed method is very useful in the digital image processing applications where higher order moments are required.

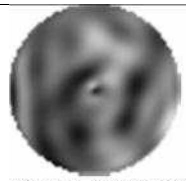
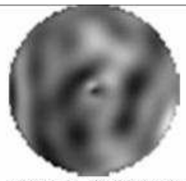
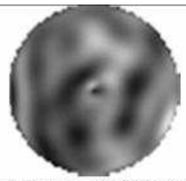
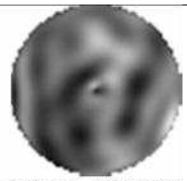
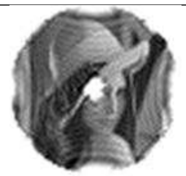
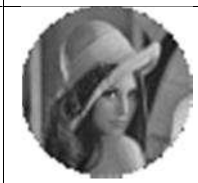
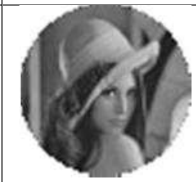
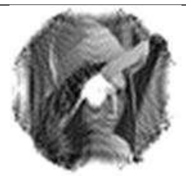
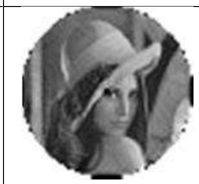
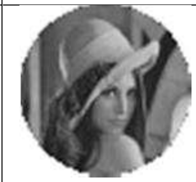
Methods / p_{max}	*Traditional method*	*Proposed method (n=2)*	*Proposed method (n=3)*	*Proposed method (n=4)*	*Proposed method (n=5)*
10	 MSRE=0.042740	 MSRE=0.042467	 MSRE=0.042903	 MSRE=0.043191	 MSRE=0.042971
20	 MSRE=0.028961	 MSRE=0.020119	 MSRE=0.021207	 MSRE=0.020950	 MSRE=0.020475
30	 MSRE=0.071915	 MSRE=0.012270	 MSRE=0.021128	 MSRE=0.012539	 MSRE=0.011514
40	 MSRE=0.283445	 MSRE=0.012562	 MSRE=0.052355	 MSRE=0.008505	 MSRE=0.007589
50	 MSRE=0.690505	 MSRE=0.027717	 MSRE=0.115027	 MSRE=0.006455	 MSRE=0.004967
60	 MSRE=1.305849	 MSRE=0.054672	 MSRE=0.238636	 MSRE=0.005716	 MSRE=0.003594
70	 MSRE=2.682446	 MSRE=0.089695	 MSRE=0.499003	 MSRE=0.004929	 MSRE=0.004881

Fig. 3. Reconstructed Image 0f 64×64 size grey image using proposed method.

REFERENCES

[1] C. Singh, R. Upneja, “Error Analysis in the Computation of Orthogonal Rotation Invariant Moments”, J Math Imaging Vis, vol. 49, issue 1, pp. 251-271, 2014.

[2] C. Singh, R. Upneja, “Accurate calculation of high order pseudo-Zernike moments and their numerical stability”, Digital Signal Processing, vol. 27, pp. 95-106, 2014.

[3] Y. H. Pang, A. B. J. Teoh, D. C. L. Ngo, “A discriminant pseudo-Zernike moments in face recognition”, Journal of Research and Practice in Information Technology vol. 38, issue 2, pp. 197–211, 2006.

[4] X. Y. Wang, Y. J. Yu, H. Y. Yang, “An effective image retrieval scheme using color, texture and shape features”, Comput. Stand. Interfaces, vol. 33, issue 1, pp. 59–68, 2011.

[5] X.Y. Wang, L.M. Hou, “A new robust digital image watermarking based on pseudo-Zernike moments”, Multidimensional System Signal Processing, vol. 21, issue 2, pp. 179–196, 2010.

[6] X.Y. Wang, T.X. Ma, P.P. Niu, “A pseudo-Zernike moment based audio watermarking scheme robust against desynchronization attacks”, Computer Electr. Eng., vol. 3, pp. 425–443, 2011.

[7] W. Cao, J. Zhao, W.C. Meng, R.S. Li, “Map matching algorithm according to pseudo Zernike moments”, Second WRI World Congress on Software Engineering, pp. 21–24, 2010.

[8] Y. H. Pang, T. B. J. Andrew, N. C. L. David, F.S. Hiew, “Palmprint verifications with moments”, J. WSCG, vol. 12, pp. 1–3, 2003.

[9] A. Kaur, C. Singh, “Sub-pixel edge detection using pseudo Zernike moment”, International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 4, issue 2, pp. 107–118, 2011.

[10] C. Singh, S. Pooja, R. Upneja, “On image reconstruction, numerical stability, and invariance of orthogonal radial moments and radial harmonic transforms”, Pattern Recognition and Image Analysis, vol. 21, issue 4, pp. 663–676, 2011.

[11] S. X. Liao, M. Pawlak, “On the accuracy of Zernike moments for image analysis”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, issue 12, pp. 1358–1364, 1998.

[12] Y. Xin, M. Pawlak, S. X. Liao, “Accurate calculation of moments in polar co-ordinates”, IEEE Transactions on Image Processing, vol. 16, pp. 581–587, 2007.

[13] C. Y. Wee, R. Paramseran, “On the computational aspects of Zernike moments”, Image and Vision Computing, vol. 25, pp. 967–980, 2007.

[14] C. Singh, E. Walia, S. Pooja, R. Upneja, “Analysis of algorithms for fast computation of pseudo Zernike moments and their numerical stability”, Digital Signal Processing, vol. 22, issue 6, pp. 1031–1043, 2013.

[15] C. W. Chong, P. Raveendran, R. Mukundan, “An efficient algorithm for fast computation of pseudo-Zernike moments”, IJPRAI, vol. 17, pp. 1011–1023, 2003.

[16] X. Wang, S. Liao, “Image reconstruction from Orthogonal Fourier Mellin moments”, ICIAR , pp. 687–694, 2013.

To Study the Effect of Fog, Snow and Rain Attenuation on FSO Link

Saurabh Dhawan
Communication System, GNDU, Amritsar
saurabhdhawan29@gmail.com

Mandeep Singh
Communication System, GNDU, Amritsar
mandeepmatharu99@yahoo.com

***Abstract*— Now a days, the wireless technology has become popular and receive growing acceptance as a Wireless Access system. FSO is a technology that uses the visible and infrared (IR) light propagating through the atmospheric channel to transmit information. The FSO communications are attracting attention as the contemporary technology to solve the last mile bottleneck issues in local area access networks due to their high bandwidth, low cost implementation in a non-licensed spectrum, relatively lower power consumption and immunity and security compared with RF technologies. As we know that the free space optics is highly affected by the various atmospheric conditions which cause the degradation of the performance of FSO link. This paper presents the study of the effect of the various atmospheric conditions on the FSO link. This paper gives the detail about the FSO communication systems and the components used in the FSO communication System. This paper also shows the effect of the fog, snow and rain over the FSO links. This paper provides the analysis of the various fog models such as Kim Model,Kruse Model and Al Naboulsi et al Model. The analysis of the wet and dry snow is also presented in this paper. These effects are represented by the help of matlab simulation.**

Keywords: FSO (Free Space Optics), LED, LD, FP, DFB, Attenuation.

I. INTRODUCTION

The Free Space Optical (FSO) communication is also known as Wireless Optical Communication.During recent years, there has been growing interest in free-space optical (FSO) communications.FSO is a technology that uses the visible and infrared (IR) light propagating through the atmospheric channel to transmitinformation. The FSO communications are attracting attentionas the contemporary technology to solve the last milebottleneck issues in local area access networks due to their highbandwidth, low cost implementation in a non-licensedspectrum, relatively lower power consumption and immunityand security compared with RF technologies. FSO is basically the same as fiber optic transmission. The difference is that the laser beam is collimated and sent through atmosphere from the transmitter, rather than guided through optical fiber [1]. The main cause of FSO attenuation is due to the presence of instable atmosphere because atmospheric channel is not ideal..However, the constitution of the atmosphere, particularly, aerosols (fog, smoke, dust) have similar particle size distributions comparedwith optical wavelengths in FSO. This can potentially result inscattering and absorption of visible and IR optical beams, thusdegrading the FSO link performance and its availability [2-3]. Long-range links uses to connect building-to-building, ground-to-aircraft, or ground-to-satellite. Short-range links often use infrared or visible light-emitting diodes (LEDs) and can be used indoors for data communications or outdoors for vehicular communication.

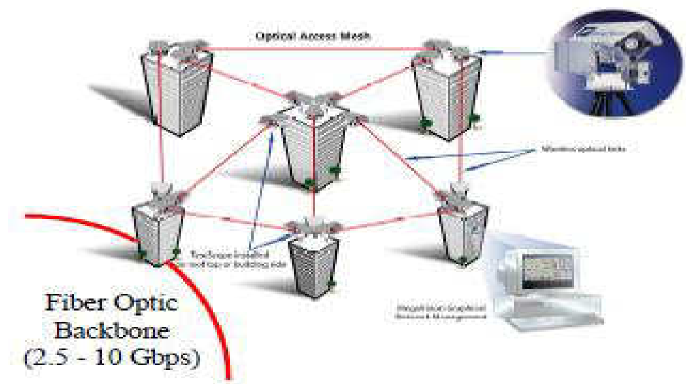


Fig .1. FSO laser communication links

In order to provide solution to last mile problem we can use free space laser communication in mesh network to get the high bandwidth quickly to the customers.

II. FSO COMMUNICATION SUBSYSTEM

FSO communication is a line of sight technology that uses laser beam for sending the very high bandwidth digital data from one point to another through atmosphere. This can beachieved by using a modulated narrow laser beam lunched from a transmission station totransmit it through atmosphere and subsequently received at the receiver station. FSO system is illustrated as:

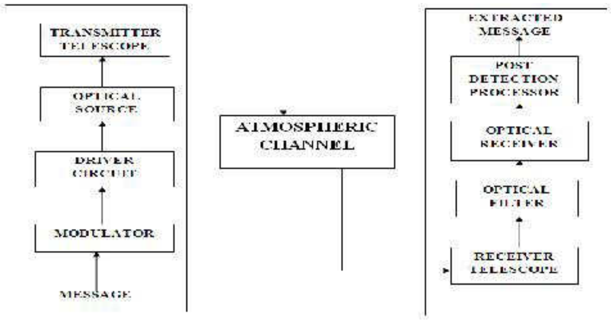


Fig. 2. FSO communication subsystem

A. *Transmitter*:

Transmitter transforms the electrical signal to an optical signal and it modulates the laser beam to transfer carrying data to the receiver through the atmosphere channel. Thetransmitter consists of four parts as shown in Fig. 2: laser modulator, driver, optical sourceand transmit telescope.

i. *Laser modulator*: Laser modulation means the data were carried by a laser beam. The modulation technique can be implemented in following two common methods: internal modulation and externalmodulation [4]. Internal modulation is a process which occurs inside the laser resonator and it depends on the change caused by the additive components and change the intensity of the laser beam according to the information signal. External modulation is the process which occurs outside the laser resonator and it depends on both the polarization phenomena and the refractive dualism phenomenon.

ii. *Driver*: Driver circuit of a transmitter are used to convert electrical signal to an optical signal by varying the current that is flow through the light source.

iii. *Optical source:* Optical source may be a laser diode (LD) or light emitting diode (LED), which used to convert the electrical signal to optical signal. A laser diode is a device that produces optical radiation by the process of stimulated emission. A laser diode emits light that is highly monochromatic and very directional i.e why laser diode has a narrow spectral width and small output beam angle divergence. LDs produce light waves with a fixedphase relationship between points on the electromagnetic wave. There are two commontypes of laser diode: Nd:YAG solid state laser and fabry-perot and distributed-feedback laser (FP and DFB) [5]. The selection of a laser source for FSO applications depends on various factors.. Typically the factors that impact the use of a specific light source include the following [6]:

- Cost and availability of components
- Transmission power and lifetime
- Modulation capabilities
- Eye safety

iv. *Transmitter telescope*: It is used to collects, collimates and directs the optical signal towards the receiver telescope i.e at the opposite end of the channel.

B. *FSO channel:*

For FSO links, the propagation medium is the atmosphere. The atmosphere may be regarded as series of concentric gas layers around the earth. Three principal atmosphericlayers are defined in the homosphere [7], the troposphere, stratosphere and mesosphere.These layers are differentiated by their temperature gradient with respect to the altitude. InFSO communication, we are especially interested in the troposphere because this is wheremost weather phenomena occur and FSO links operate at the lower part of this layer [7]. The atmosphere is primarily composed of nitrogen (N2, 78%), oxygen (O, 21%), and argon(Ar, 1%), but there are also a number of other elements, such as water (H2O, 0 to 7%) andcarbon dioxide (CO, 0.01 to 0.1%), present in smaller amounts. There are also small particlesthat contributed to the composition of the atmosphere; these include particles (aerosols) suchas haze, fog, dust, and soil [8]. Propagation characteristics of FSO through atmosphere drastically change due tocommunication environment, especially, the effect of weather condition is strong. Due to the presence of atmospheric obstacles received signal power fluctuates. The main effects on optical wireless communication are absorption, scattering, and scintillation [9].

C. *Receiver*

The receiver optics consists of five parts as shown in Fig. 2: receiver telescope, optical filter, detector, amplifier and demodulator.

i. *Receiver telescope:* The receiver telescope collects and focuses the incoming optical radiation on to the photo detector. It should be noted that a large receiver telescope aperture is desirable because it collects multiple uncorrelated radiation and focuses their average on the photo detector [10].

ii. *Optical filter:*By introducing optical filters that allow mainly energy at the wavelength of interest to impingeon the detector and reject energy at unwanted wavelengths, the effect of solarillumination can be significantly minimized.

iii. *Detector:* The detector is a semiconductor devices which converts the light energy into an electrical signal. The diodes are generally reverse-biased and capacitive charged [11].The two most commonly used photodiodes are the pin photodiode and the avalanche photodiode (APD) because they have good quantum efficiency and are made ofsemiconductors that are widely available commercially [12].

III. VARIOUS FOG ATTENUATION MODELS OF FSO

A. *Basics Of Fog Formation*

Fog can be described as a cloud of small particles of water,smoke, ice or combination of these near the earth surfacethereby scattering the incident light and hence reducing thevisibility [13]. Fog forms when the difference betweentemperature and dew point are (5°F) 3 °C, or less and water vapours in the air begin to condense into liquid water whilerelative humidity reaches to 100%. The formation of a foglayer occurs when a moist air mass is cooled to its saturationpoint (dew point). This cooling can be the result of Advection fog), radiation fog , frontal fog and upslope fog. Valley fog forms as aresult of air being radiatively cooled thereby becoming denserthan its surroundings and starts going down the slope creatinga pool of cold air at the valley floor resulting in fog formation,if the air gets cold enough to reach its dew point. The mostcommonly encountered fog types in nature are Advection and Radiation fogs.

B. *Fog Attenuation Model:*

Fog is the most critical attenuation factor among all atmospheric attenuation factors of FSO. Mie scattering is used to calculate attenuation in case of fog droplets. However, it requires information of various fog parameters such as particle size distribution, refractive index, particle size, etc. which may not be readily available at a particular location of installation. Moreover it involves complex computations [14]. Another way is to use visibility data to predict specific attenuation. This model Kruse, Kim, Al Naboulsi [15-17] use this approach and predict specific attenuation using visibility. The specific attenuation for Kruse is given by:

$$\gamma(\lambda) = \frac{3.912}{v}\left(\frac{\lambda}{550(nm)}\right)^{-q} \quad (1)$$

The coefficient q depends on

$$q = \begin{cases} 1.6, & v > 50km \\ 1.3, & 6km < v < 50km \\ 0.585v^{\frac{1}{3}}, & v < 6km \end{cases} \quad (2)$$

Where V is visibility in km and   is wavelength in nm and $\gamma(\lambda)$ represents specific attenuation. Kim et al reject such wavelength dependent attenuation for low visibility in dense fog. The q variable in eq. (1) for Kim model is given by:

$$q = \begin{cases} 1.6, & v > 50km \\ 1.3, & 6km < v < 50km \\ 0.16v + 0.34 & 1km < v < 6km \\ v - 0.5, & 0.5 < v < 1km \\ 0, & v < 0.5km \end{cases} \quad (3)$$

Al Naboulsi et al. [17] have provided relations to predict fog attenuation by characterizing advection and radiation fog separately. Al Naboulsi provide the advection fog attenuation coefficients as

$$\gamma_{ADV}(\lambda) = \frac{0.11478\lambda + 3.8367}{v} \quad (4)$$

Radiation fog is related to the ground cooling by radiation. Al Naboulsi provides the radiation fog attenuation coefficients as:

$$\gamma_{RAD}(\lambda) = \frac{0.18126\lambda^2 + 0.13709\lambda + 3.7502}{v} \quad (5)$$

IV. SNOW ATTENUATION IN FSO

Snow causes the scattering of the light wave and the laser beam power is attenuated due to the reduction in received signal strength. When received signal level decreases either complete link failure or bit errors occurs[18]. The amount of light attenuation is proportional to number and size of fog, rain and snow particles [19-20]. Since the size of snowflakes are generally larger than rain drops, the received signal strength fluctuation will be larger[21] and can cause link failure[22-23]. Whenever the laser beam passes through the snow then the received signal power will depend upon the area of the snow particle and separation from the transmitter as well as on the location of snow particle relative to the beams cross section.The FSO attenuation due to snow is calculated for both dry and wet snow. If snow rate in mm/hr is represented by S then specific attenuation in dB/km is given by [24]

$$\alpha_{snow} = a.S^b \quad (5)$$

If   is the wavelength, a and b are as following for dry snow

$$a = 5.42 \times 10^{-5} \times \lambda + 5.4958776, b = 1.38 \quad (6)$$

The same parameters for wet snow are given as follows

$$a = 1.023 \times 10^{-4} \times \lambda + 3.7855466, b = 0.72 \quad (7)$$

V. RAIN ATTENUATION IN FSO

Rain is formed by water vapor contained in the atmosphere. It consists of water dropletswhose form and number is variable in time and space. Their form depends on their size: they are considered as spheres until a radius of 1 mm and beyond that as oblate spheroids: flattened ellipsoids of revolution.

Scattering due to rainfall is called non-selective scattering, this is because the radius ofraindrops (100 – 1000 µm) is significantly larger than the wavelength of typical FSOsystems. The laser is able to pass through the raindrop particle, with less scattering effectoccurring. The haze particles are very small and stay longer in the atmosphere, but therain particles are very large and stay shorter in the atmosphere. This is the primaryreason that attenuation via rain is less than haze. An interesting point to note is that RFwireless technologies that use frequencies above approximately 10 GHz are adverselyimpacted by rain and little impacted by fog. This is because of the closer match of RFwavelengths to the radius of raindrops, both being larger than the moisture droplets in fog . The rain scattering coefficient can be calculated using Stroke Law see Eq. (8)[25]:

$$\beta_{rainscat} = \pi a^2 N_a Q_{scat} \left(\frac{a}{\lambda}\right) \quad (8)$$

Where:

a= is the radius of rain drop, (cm)

N_a= is the rain drop distribution, (cm^{-3})

Q_{scat}= is the scattering efficiency.

The rain drop distribution N_a can be calculated using equation:

$$N_a = \frac{R}{1.33(\pi a^2) v_a} \quad (9)$$

Where:

R= is the rain fall rate, (cm/sec).

v_a= is the limit speed precipitation.

It is given as:

$$v_a = \frac{2a^2 \rho g}{9\eta} \quad (10)$$

Where:

ρ= is water density, ($\rho = 1g/cm^3$)

g= is the gravitational constant, ($g = 980cm/sec^2$).

η= is viscosity of air, ($\eta = 1.8 \times \frac{10^{-4}g}{cm}.sec$).

The rain attenuation can be calculated by using Beer's law as:

$$\tau = exp(-\beta_{rainscat} L) \quad (11)$$

VI. MATLAB SIMULATION

Matlab Simulation is carried out in order to predict the effect of various weather conditions on the optical communication System which is shown below:

A. *Kim Model:*

The fig 3 presents the variation of the atmospheric attenuation with the visibility distance in km for Kim Model. It is clear from the fig 3 that according to the Kim model the attenuation decreases with increase in the wavelength. and also it shows that attenuation doesn't have the wavelength dependency for the visibility distance greater then 10 km i.e for the visibility distance greater then 10 km this model doesn't give accurate results.

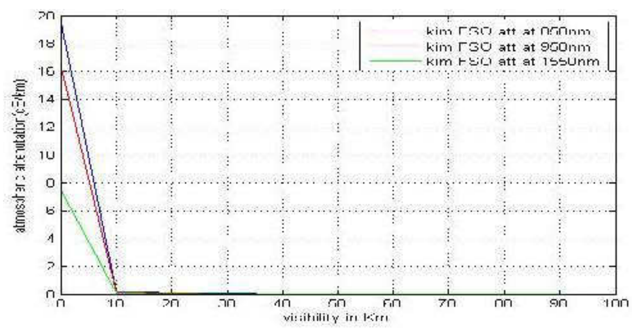


Fig. 3. Attenuation for Kim Model

B. *Kruse Model:*

The fig 4 presents the variation of the atmospheric attenuation with the visibility distance in km for Kruse Model. It can be observed from the fig 4 that specific attenuation for longer wavelengths is less. At extremely low visibility, the specific attenuation at 850 nm and 950 nm is almost 20 dB/km more than the specific attenuation at 1550 nm. This shows that under dense fog with extremely low visibility, 1550 nm surpasses the performance of 850 nm and 950 nm.

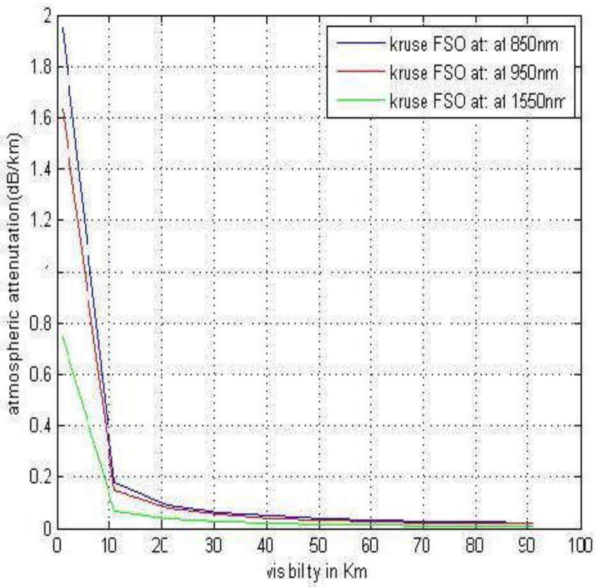


Fig .4. Attenuation for Kruse Model

C. *Al Naboulsi et al Model*:

The Al Naboulsi et al model describe two types of the fogs that effect the communication through the free space i.e Advection and Radiation fog. The fig 5,6,7 presents the variation of both types of fog with the visibility range for the three different wavelengths i.e at 850 nm,950 nm,1550 nm. It can be concluded from the figures 5,6,7 that Al Naboulsi model also does not exhibit any wavelengths dependent atmospheric attenuation.

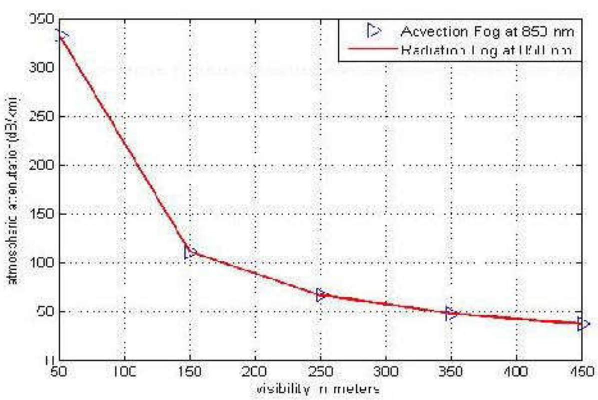


Fig. 5. Fog attenuation at 850 nm

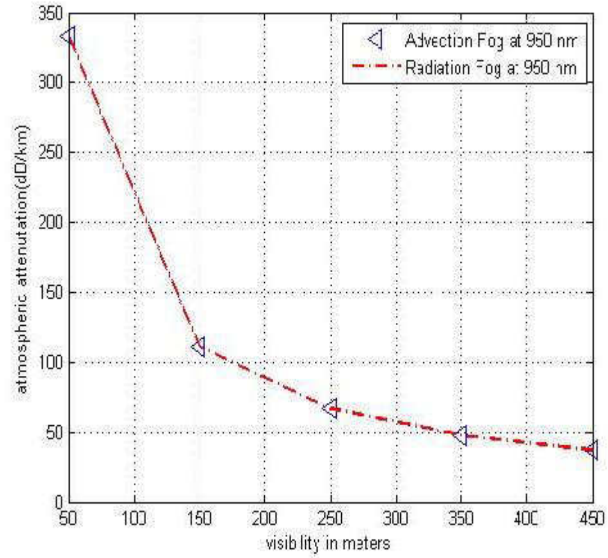


Fig. 6. Fog attenuation at 950 nm

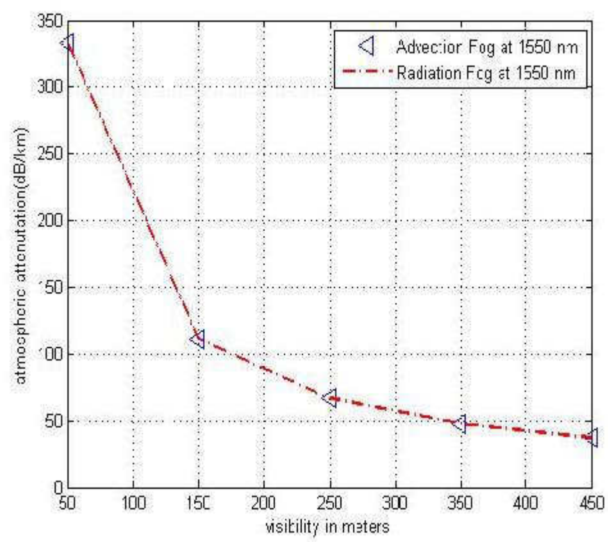


Fig .7. Fog attenuation at 1550 nm

D. *Snow Attenuation*:

The matlab simulation for both the dry and wet snow is shown below.. The fig 8 and fig 9 represents the variation of the attenuation caused by the dry snow with the snow rate and wavelength .From the fig 8 it is clear that the specific attenuation for the dry snow increases with the increase in the snow rate and the specific attenuation is wavelength independent. Whereas from Fig 9 it can be concluded that the for particular value of the snow rate the specific attenuation for the dry snow remain almost constant for the various values of the wavelength.

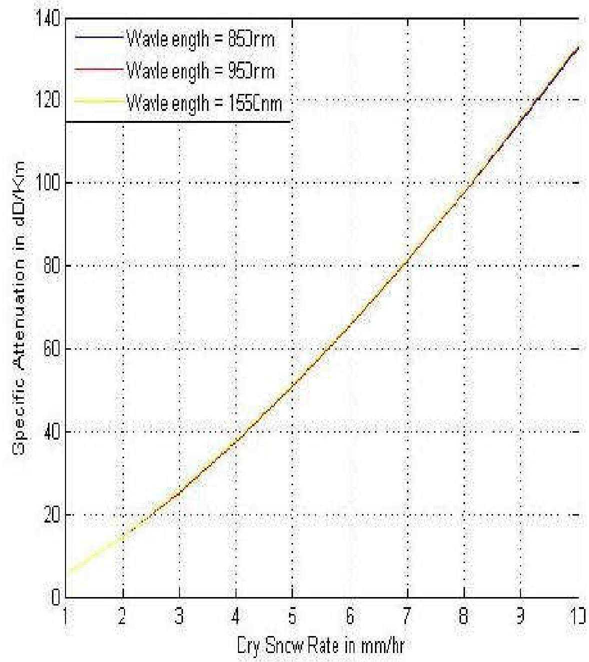


Fig .8. Dry snow attenuation

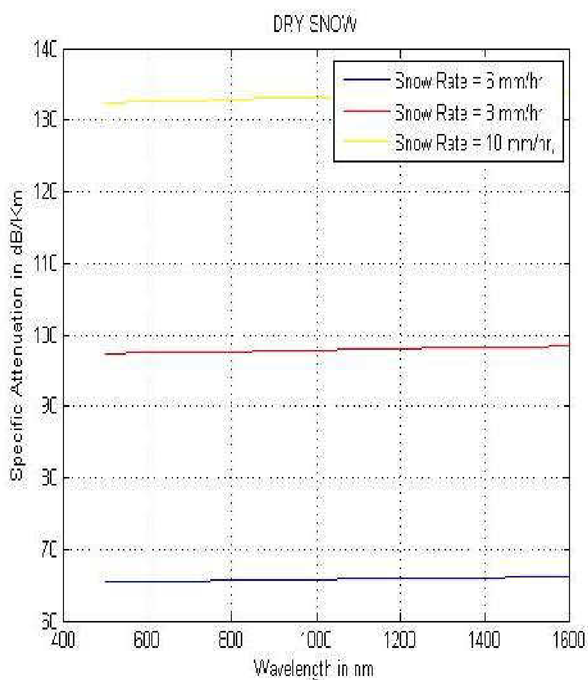


Fig .9. Dry snow attenuation

The fig 10 and fig 11 represents the variation of the attenuation caused by the wet snow with the snow rate and wavelength .From the fig 10 it is clear that the specific attenuation for the wet snow increases with the increase in the snow rate and the specific attenuation is wavelength independent. Whereas from Fig 11 it can be concluded that the for particular value of the snow rate the specific attenuation for the dry snow remain almost constant for the various values of the wavelength.

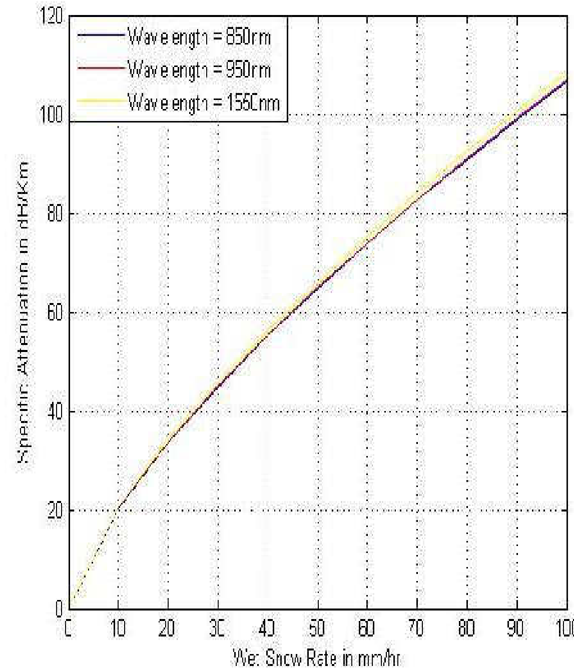


Fig .10. Wet snow attenuation

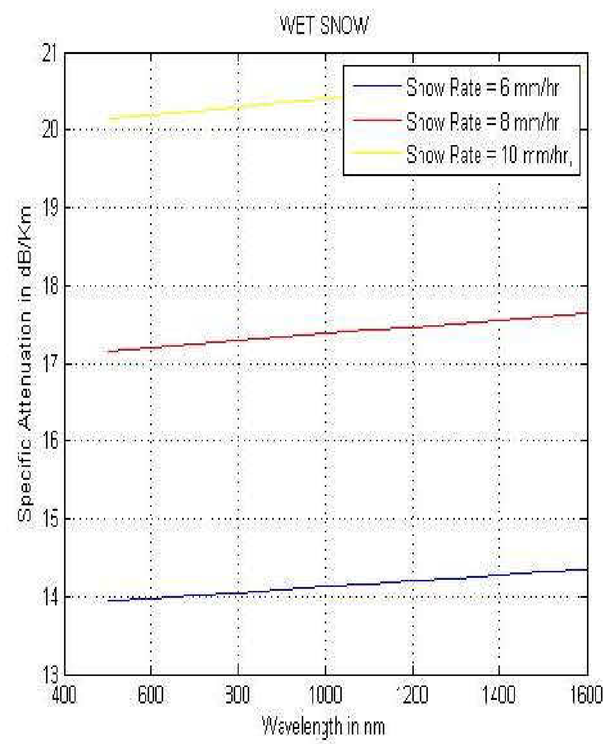


Fig .11. Wet snow attenuation

E. Rain Attenuation:

Fig 12 shows the variation of the specific attenuation experienced by the optical wavelength s for the various values of the rain rate. Also rain doesn't show the wavelength dependent attenuation.

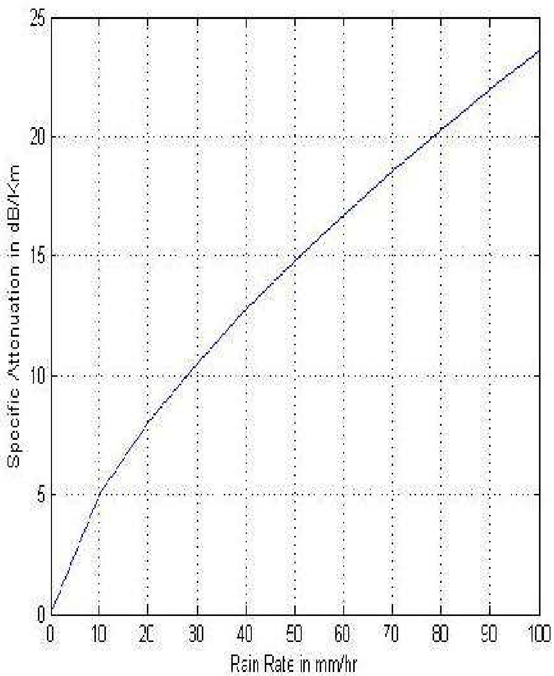


Fig .12. Rain Attenuation

VII. CONCLUSION

The study of the effect of the various atmospheric conditions on the FSO link is very important in order to predict the efficiency of the FSO communication system. The study of these effects is done on the basis of the attenuation coefficient. Attenuation is defined as the reduction in the received power as transmitted through space. The effect of the fog is studied with the help of the various models such as the kim model, kruse model and Al Naboulsi et al Model. The effect of snow is analyzed by considering the dry and wet snow. It has been observed that fog attenuation doesn't have the wavelength dependency for the visibility distance greater than 10 km. For snow attenuation it can be concluded that it increases with the increase in the snow rate and the rain attenuation also increases with the increase in the rain rate. This study has practical importance in designing of wireless network in various terrain and environment condition.

REFERENCES

[1] Weichel H., "Laser Beam Propagation in the Atmosphere", SPIE, Optical EngineeringPress, Vol. TT-3, 1990.

[2] J. Perez, Z. Ghassemlooy, S. Rajbhandari, M. Ijaz, and H. Lee-Minh, "Ethernet FSO Communications Link Performance Study Under a Controlled Fog Environment," *IEEE Communications Letters,* pp. 1-3,2012.

[3] R. N. Mahalati and J. M. Kahn, "Effect of fog on free-space optical links employing imaging receivers," *Journal of Optics Express,* vol. 20, pp.1649-1661, 2012.

[4] K. Majumdar, J. C. Ricklin, "Free Space Laser Communications Principles and Advances", Springer ISBN 978-0-387-28652-5, 2008.

[5] H. Hemmati, "Near-Earth Laser Communications", California, Taylor & Francis Group,Book, LLC, 2008.

[6] H. Willebrand and B. S. Ghuman, "Free-Space Optics Enabling Optical Connectivity inToday's Networks", SAMS, 0-672-32248-x, 2002.

[7] B. Olivieret., "Free-Space Optics, Propagation and Communication", Book, ISTE,2006.

[8] Roberto Ramirez-Iniguez, Sevia M. Idrus and Ziran Sun, "Optical Wireless Communications IR for Wireless Connectivity", Taylor & Francis Group, Book, CRC Press, 2007.

[9] N. Araki, H. Yashima, "A Channel Model of Optical Wireless Communications duringRainfall", IEEE, 0-7803-9206, 2005.

[10] Z. Ghassemlooy and W. O. Popoola , "Terrestrial Free Space Optical Communication" ,Optical Communications Research Group, NCR Lab., Northumbia University,Newcastle upon Tyne, UK, 2010.

[11] Nadia B. M. Nawawi, "Wireless Local Area Network System Employing Free SpaceOptic Communication Link", A Bachelor Degree thesis, May 2009.

[12] Delower H. and Golam S. A., "Performance Evaluation of the Free Space Optical (FSO)Communication with the Effects of the atmospheric Turbulence", A BachelorDegree thesis, January 2008.

[13] S. S. Muhammad, B. Flecker, E. Leitgeb and M. Gebhart, "Characterization of fog attenuation in terrestrial free space links", Journal of OE 46(4) 066001 June 2007.

[14] Nadeem, F., Leitgeb, E., Awan, M. S., Kandus, G., "Optical Wavelengths Comparison for Different Weather Conditions", International Workshop on Satellite and Space Communication (IWSSC), 2009.

[15] Kruse, P., McGlauchlin, L., McQuistan, R., "Elements of Infrared Technology: Generation, transmission and detection". John Wiley & Sons, 1962.

[16] Kim, I. I., McArthur, B., Korevaar, E., "Comparison of laser beam propagation at 785 nm and 1550 nm in fog and haze for optical wireless communications," Proc. of SPIE, 4214, 2001, pp. 26-37.

[17] Al Naboulsi, M., Sizun, H.,deFornel, F.,"Fog attenuation prediction for optical and infrared waves," Journal of SPIE, Optical Engineering, 43, 2004, pp. 319-329.

[18] M. Akiba, W. Wakamori, S. Ito, "Measurements of optical propagation characteristics for free space optical communication during rain fall", IEICE Trans. Commun. E87-B, pp. 2053-2056, (2004).

[19] M. Achour, "Simulating atmosphere free space optical propagation part I, rainfall attenuation", Proc. SPIE 4635, pp. 192-201 (2002).

[20] I .I. Kim, B. McArthur, E. Korevaar, "Comparison of laser beam propagation at 785 nm and 1550 nm in fog and haze for optical wireless communication, Proc. SPIE 4214, pp. 26-37 (2001).

[21] M. Akiba, K. Ogawa, K. Walkamori, K. Kodate, S. Ito, " Measurement and simulation of the effect of snow fall on free space optical propagation", Applied Optics Vol. 47 No. 31, pp. 5736-5743 (2008).

[22] S. E. Yuter, D. E. Kingsmill, L. B. Nance, M. Loffler-Mang, " Observations of precipitation size and fall speed characteristics within coexisting rain and wet snow ", J. Appl. Meteorol. 45, pp. 1450-1464 (2006).

[23] P. P. Lawson. R. E. Stewart, L. J. Angus, "Observations and numerical simulations of origin and development of very large snowflakes", J. Atmos. Sc. 55, pp. 3209-3229 (1998).

[24] S. Sheikh Muhammad, P. Kohldorfer, E. Leitgeb: "Channel Modeling for Terrestrial Free Space Optical Links", ICTON 2005.

[25] Achour M., "Simulating Atmospheric Free-Space Optical Propagation part I, Haze, Fogand Low Clouds, Rainfall Attenuation", Optical Wireless Communications,Proceedings of SPIE, 2002.

Analysis on Hand & Eye Gesture Detection Algorithms

Alok Kumar
Department of ECE
CT Group of Institution
Jalandhar, Punjab, INDIA
Kalok2705@gmail.com

Jagdeep Singh
Department of ECE
CT Group of Institution
Jalandhar, Punjab, INDIA
Sandhuereka@gmail.com

Vikrant Sharma
Department of ECE
CT Group of Institution
Jalandhar, Punjab, INDIA
vikrant.sharma@yahoo.com

Kunal Chaudhary
Department of ECE
CT Group of Institution
Jalandhar, Punjab, INDIA
Err.kunal@gmail.com

Abstract*—Gesture recognition is mainly curious on analyzing the functionality of human intellects. The main focus of hand gesture recognition is to implement the output on a system which able to recognize human gesture and utilize them to deliver information or device control. Hand Gesture provides a harmonizing modality to speech for expressing own ideas. Information communicates with a hand gesture relate in a conversation degree, discourse structure, spatial and temporal structure. The approaches could be divided into two parts Data Glove Based and Vision Based approaches. Secondly, eye detection is isolate into eye position and eye contour detection. Existing work in eye detection could be categorized into two key categorized: Traditional based passive approach and IR based approach. The purpose of this paper is to equate various human gesture recognition systems for interfacing technology straightly to human intellects without any corporeal medium in ambient surroundings.

***Keywords*—CLUT (Color look up table) , Histogram, thresholding and color segment**

I. INTRODUCTION

The existing human computer interaction (HCI) techniques, keyboards and mice are probably the best-known input devices. However, these devices constrain the dexterity and naturalness of interaction with computer-controlled applications. This limitation becomes more apparent when we employ these devices settings such as virtual reality applications that require a wide variety of input from the user. Thus, in recent years there has been a tremendous push in research toward novel devices and techniques for natural and friendly interaction with computers. In particular, the analysis of hand motion has proved attractive to many computer animation and virtual reality researchers since the human hands perform most everyday tasks, assist in communication with other individuals and can sometimes express our feelings [1].Hand gesture in one of the best method to expressing thought along with speech. Hand Gestures give degree, discourse structure, spatial and temporal structure to the conversation of any language[2]. The hand gesture technique is isolate into Data-glove and Vision based approaches.

The first implemented approach was data glove approach. This approach was based on sensor technique which gave parametric data on the basis of wrist moment and finger motion. This approach was not so famous among scholars due to its cost and gives complex experience to the novice [3]. In contrast due to more complexity of data glove approach a vision based approach is implemented. In this approach, a video camera is used to judge the hand expression on real time basis. But unfortunately this approach poses a new challenge that these systems need clear background and good light intensity [4].

An important face feature point is eye and eye detection and tracking is also a vital challenge for researcher. Eye detection is divided into two categories eye position and eye contour detection. In existing work the eye detection is classified into mainly two categories traditional image based and IR based approaches.

This paper is organized as follow: In Section II, hand gesture detection and resolution algorithms are defined. In section III, comprises of eye detection and tracking algorithms. Finally, a conclusion is given in Section IV.

II. HAND GESTURE DETECTION AND RESOLUTION ALGORITHMS

A. Hidden Markov Model

In this model, the gestures are obtained from a sequence of video pictures which are trailing by a skin color blobs. This Algorithm is concerned with dynamic aspects of gesture. The basic goal of this model is to recognize two classes of gesture: deictic and symbolic. The video picture is filtered by using CLUT (Color look up table) of skin color pixels in YUV color space and gather into blobs. Blobs are geometrical object relates with colrimetry (YUV) and location (x,y) of skin color pixel to determine identical area. This method define the gesture classes which indicates Deictic gesture directing movement towards left or right side of face space whereas symbolic gesture projected to execute command on left or right shoulder[5]

B. YUV color space and CAMSHIFT algorithm

This algorithm deals with recognition of hand gesture which is done in following ways

- Digital Camera is used to record hand gesture.

- All video frames are visualize carefully and then applied YUV color space (skin color based) for segmentation. This technique isolates the intensity indicates by symbol Y and chrominance components specify by U (saturation) and V (hew).
- Then CAMSHIFT algorithm is used to shift hand gestures. Meanwhile hand is the largest inter-connected region of the body so we have to segment it from the body [6].
- After this the position of hand centroid is calculated in each frame. The basic method for the same is, first calculating the zeroth and first moment and then with the help of calculation information could be extracted.
- At last the various centroid points are combined and form trajectory. This shows the path of hand movement and tracking procedure is done

C. Naive Bayes Classifier Algorithm

This is a fast and effective algorithm for static hand gesture recognition. This technique is categorizing different hand gestures based on geometric invariants which are extracted during segmentation. Unlike other techniques this isn't based on skin color, gestures are obtained from video frame with static background. The process of segmentation is done by elimination of background pixel according to their corresponding histogram of each image. Naive Bayes algorithm is evaluate the probability of each gesture and classified it by merging with weighted K-nearest neighbors algorithm. When this technique is tested with JAST Human Robot dialog system it gives more than 93% gestures accurately [7]

This method is based on three main steps:-

- Segment and label the object of concern and fetch geometric invariant from them.
- Then hand gestures are categorized using K-nearest neighbor with distance weighting algorithm (KNNDW) to deliver accurate data for locally weighted Naive Bayes classifier.
- The invariant of each region of interest take as the input of classifier while output is a type of gesture.
- Once gesture is categorized then finally locate the specific property of the gesture.

D. Vision based hand gesture recognition

In this algorithm, the movement of hand is recorded by video cameras. The input video is decomposed in set of feature. Gestures are nothing but a continuous sequence of hand poses which are traced, when hands are isolate from body part and background objects [8].

Vision based hand posture and gesture recognition approaches are isolate in two types.

- 3D hand model based approach
- Appearance based approach

a) 3D Hand model Based approach

This model based on 3D kinematic hand model with considerable DOF's. Three dimensional hand model based approach approximate the hand gesture parameters by taking comparison between input images and 2D appearance. This approach is realistic interaction with virtual environment. There are several disadvantage of this approach that have kept it from real world. This approach is very sensitive to noise in the imaging process and cannot fetch predictable self-occlusion of hand

b) Appearance based approach

In this approach image features are related with visual appearance of hand gesture. These parameters compare with extracted image features and display the resulted output. This algorithm has many real time advantages due easier 2D image feature. It is a simple and straightforward approach method that is frequently utilized to search skin color region in image. In spite of its popularity, it has some drawbacks like skin color pixels are very sensitive to light condition.

III. EYE DETECTION AND TRACKING ALGORITHMS

A. Template Based Technique

In this algorithm, a generic eye model constructed on eye shape is design at first. Template matching is based on searching an image by eyes. This method is detected image accurately but somehow it is time consuming. In order to increase the accuracy this technique first matches the whole face with eye template in order to pixel by pixel. If scholars are not aware of the eye size then we have to repeat the matching process with eye templates of different size which consume a lot of time. But with advancement feature based method is introduce, which first detects two rough region of the face and then matching will be performed. It is the best solution to improve the efficiency of template matching method. This technique focuses on two rough points which ultimately reducing the area of scanning and cutting down the times of matching. In contrast, the size of eye template can be calculated according to region size. Moreover it gives profitability from evaluation of eye size and matching is done only once. Both the algorithm combines the efficiency of feature based algorithm and accuracy of template based algorithm [9].

B. Appearance Based Method

This algorithm detects eyes on their photometric appearance. This technique usually needs a collection of huge amount of data which represent eyes on different entity under unlike illumination condition. This huge training data is used by classifier and for face detection [10]

C. ADA boost and SVM Classifier

ADA boost algorithm is used to select features instead of being a classifier. In this technique a subset of feature is selected at each step and each chosen feature is uncorrelated with preceding feature. The selected features in ADA boost method improve the speed of the operation. Beside this, SVM classifier is implemented on the features selected by the ADA boost algorithm. In SVM technique a 5-Fold Cross validate scheme is used. The finest parameters are extracted by grid search method [11]. The essence parameters of SVM are Sigma (RBF kernel parameter) and penalty C which is selected between 25-220 experimentally. In SVM algorithm facial image is placed into cascaded classifier via rectangular features and result is scaled and transform into a pattern map.

D. Color based detection algorithm

This algorithm is totally different from other techniques and fully based on eye map technique. It is one of the useful ways to detect eye by using color characteristics. In this technique two maps are made on the basis of their components and after that merge them to obtain final map. Scholars have to generate an extra phase on this final map to determine suitable eye pair. This extra phase consist of geometrical tests and flexible thresholding [12]. These two terms are help to selected or generated proper candidates respectively. This technique works in steps, first facial image is separated into two eye maps. First eye map indicates the chrominance component and second eye map reflects the luminance component in the eye.

IV. CONCLUSION

This review paper shows various algorithms for hand gesture detection and eye detection and tracking. Observation based on the hand gesture tracking and gesture demonstrates that using YUV color segment followed by CAMSHIFT technique will help to detect and track the centroid value of hand gesture very easily by calculating and comparing each point with preceding values and after that it will implement with Hidden Markov model algorithm for various applications. Whereas for analysis of eye detection and tracking it could be analyzed that a algorithm which would be a combination of template based and color based eye detection techniques. Color based eye detection play a vital role in texture recognition of facial regions and after that template based technique helps to find the perfect location of eyes using eye maps. Exclusively the accuracy of template based is 99% whereas the efficiency of color based algorithms are and 98.2%. Together both will implemented in various application which takes the research on new apex.

REFERENCES

[1] T. S. Chou, A. Gadd, and D. Knott, "Hand-eye: A vision based approach to data glove calibration," in Proc. Human Interface Technol.,pages. 2–7,2007.

[2] James M. Rehg, et. al."Visual Tracking of High DOF Articulated Structures: an Application to Human Hand Tracking", Third European Conf. on Computer Vision, pp 35-46, may 1994.

[3] Wu, Y., Lin, J., and Huang, T. S.,"Capturing natural hand articulation", In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pp. 426—432, 2001

[4] G. R. S. Murthy & R. S. Jadon,"A review of vision based hand gestures recognition", *Volume 2, No. 2, pp. 405-410, December 2009*

[5] Chen, F.-S.; Fu, C.-M.; Huang, C.-L. , "Hand gesture recognition using a real-time tracking method and hidden Markov models", Science Direct – Image and Vision Computing, Volume 21, Issue 8, 1 August 2003, pp.745-758

[6] Vadakkepat, P, Lim, P. ; De Silva, L.C. ; Liu Jing ; Li Li Ling, "Multimodal Approach to Human-Face Detection and Tracking", Industrial Electronics, IEEE Transactions on Issue Date: March 2008, Volume: 55 Issue:3, pp.1385 – 1393

[7] Pujan Ziaie, Thomas Müller, Mary Ellen Foster, Alois Knoll, "Using a Naïve Bayes Classifier based on K-Nearest Neighbors with Distance Weighting for Static Hand-Gesture Recognition in a Human-Robot Dialog System" Advances in Computer Science and Engineering Communications in Computer and Information Science, 2009, Volume 6, Part 1, Part 8, pp.308-315

[8] P Garg, N Aggarwal, S Sofat, "Vision Based Hand Gesture Recognition", World Academy of Science, Engineering and Technology, pp.1-6 (2009).

[9] K Peng, L Chen, S Ruan, "A robust algorithm for eye detection on gray intensity face without spectacles", Journal of Computer Science & Technology, Oct, 2005

[10] V Kith, M El-Sharkawy, "A feature and appearance based method for eye detection on gray intensity face images",Computer Engineering & Systems, ICCES 2008, International Conference, Issue Date: 25-27 Nov. 2008, pp.41 – 47

[11] X Tang, Z Ou, T Su, H Sun, P Zhao, "Robust Precise Eye Location by Adaboost and SVM Techniques", Advances in Neural Networks – ISNN 2005, Lecture Notes in Computer Science, 2005, Volume 3497/2005, pp.93-98.

[12] M Shafi, PWH Chung, "A Hybrid Method for Eyes Detection in Facial Images", Proceedings Of World Academy Of Science, Engineering And Technology, Volume 32, August 2008.

Automatic Number Plate Recognition System under Image Processing: A Review

Sarbjit Kaur
Dept. of CSE
CT Group of Institutions,
Jalandhar
sarbjitkhinda1991@gmail.com

Sukhvir Kaur
Dept. of CSE
CT Group of Institutions,
Jalandhar
sukhsain.17@gmail.com

***Abstract*—Automatic Number Plate Recognition (ANPR) is an important area of research due to its importance in wide range of commercial applications. It is an image-processing technology that extracts the number plate information from vehicle image or from sequence of images hence identifies vehicles by their number plates. ANPR work is generally framed into 4 stages. First stage is to acquire the image and performing some pre-processing on it; Second stage is the detection and extraction of number plate area from the whole body of vehicle image; Third stage is segmentation or isolation of characters from the extracted number plate and fourth stage is to recognize the segmented characters and to display the output result. Only the number plate area is extracted from the entire input vehicle image after detection and processed further in character segmentation. From the extracted number plate each character is isolated by segmentation in character segmentation phase. After this the isolated (segmented) characters are recognized in the character recognition phase. In this paper we represent review of various methods used in each phase of automatic number plate recognition system and difficulties occur during that phase.**

Keywords — Automatic number plate recognition, histogram equalization, number plate extraction, character segmentation, morphological operations, robert edge detection, sobel edge detection, character recognition, template matching, character normalization, feature extraction.

I. INTRODUCTION

Automatic number plate recognition (ANPR) system place important role in Intelligent Transportation System (ITS). In transportation vehicles play very important role. In the recent years the use of vehicles has been increasing due to population growth. So the control of vehicles has been becoming a big problem. Automatic number plate recognition system is used for the effective control of these vehicles. ANPR is also known as license plate recognition, number plate tracking, car plate recognition, automatic license plate reader, automatic vehicle identification, vehicle number plate recognition etc.

Mostly in developed and developing country the attributes of the number plates are maintained strictly. The background color of plate, character color, size (aspect ratio) of number plate, font size, font style, spacing between the subsequent characters, script etc. are attributes that are maintained strictly. For the standard plates aspect ratio is very important factor and in all developed and mostly in all the developed country the aspect ratio of vehicles plate is same means all the number plates are of same size where aspect ratio of a region is defined as ratio of length to width of that region and it is calculated as

$$\text{Aspect ratio} = \text{Length}/\text{Width} \qquad (1)$$

In India, two types of number plates :- 1) Plates for private vehicles consist of white background with black letter as shown in Fig.1. 2) Plates for commercial vehicles (taxi, bus, trucks) consist of yellow background with black letter as shown in Fig.2. In India the vehicle registration scheme consists of two-letter state code followed by two-digit district code followed by series code. This is followed by four digit code that is the actual registration number called license plate number and it is unique for each vehicle.

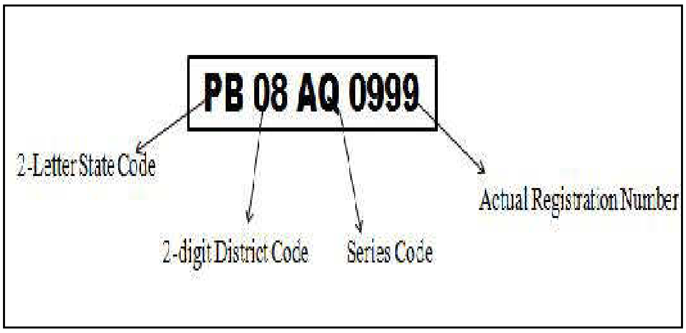


Fig.1. Standard Format for Indian Number plate for Private vehicles

Fig.2. Standard Format for Indian Number plate for Commercial vehicles

The ANPR system consist of four phases: - First phase is to acquire the input image and performing some pre-processing on it to make good quality input image. In the second phase the detection and extraction of exact number plate area is done from vehicle background image. In the third phase the segmentation of characters on number plate is done and in last character recognition phase the character are recognized by matching with the template database and output result are displayed. ANPR system as shown in fig.3 consists of four phases:-

1) Image Acquisition and Pre-processing phase

2) Number Plate Extraction phase

3) Character Segmentation phase

4) Character Recognition phase

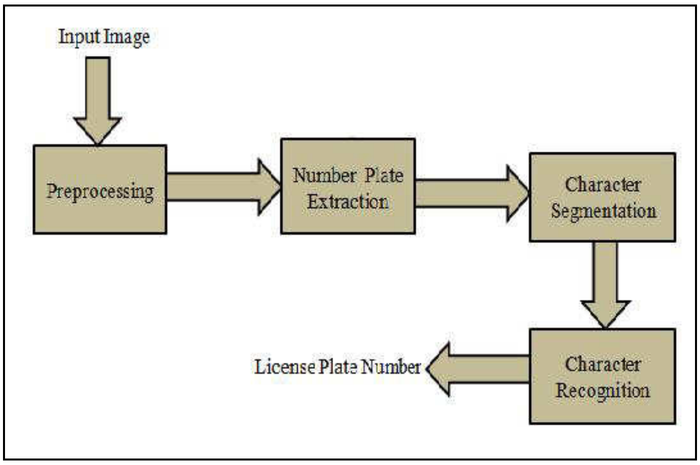


Fig.3. ANPR system

Firstly the input image is acquired and pre-processing is done to improve the contrast of image, to remove noise in the image and to enhance the processing speed. To remove the noise various types of filters are used and to improve the contrast of input image, images are enhanced by various techniques like contrast stretching, histogram equalization etc. After pre-processing the exact plate area is detected and then extracted in number plate extraction phase. After this the character on the number plate area are segmented and then segmented character are recognized. The extraction of number plate is difficult and directly influences the accuracy of ANPR because all further steps depend on accurate extraction of number plate. The difficulties in detection and extraction of number plate occur due to the following main reasons:-

1. Number plates exist on different location for different vehicles
2. Number plates occupy a small portion of whole input vehicle image.
3. Scene complexity (background objects, vehicle occlusion, multiple objects etc.) affect the efficiency of extraction of number plate area.
4. Lighting results into contrast problems
5. Weather conditions (fog, snow, rain) responsible for introducing image "noise".
6. Distortion occurs due to wrong plate or camera position
7. Some number plates contain various screws and frames in addition to unwanted characters.
8. The difference in size of plate, background graphics, single versus double-row plates, dirty plates, color of characters, font size, font style, font script, number of lines in the plate, background color etc. increase difficulty during detection of actual number plate area.
9. Vehicle motion, blurred input image, low or uneven illumination, low resolution of input image, distorted characters, reflection, shadows etc. affect the efficiency of extraction of number plate area.

ANPR system used in number of applications as:-

a) Parking - The ANPR system is used for automatic ticketing of vehicles at parking area. By extracting the number plate using ANPR, the vehicle plate number is used to automatically enter the vehicle and hence calculate the parking fee during exit.

b) Traffic law enforcement- Tracking vehicles during signal violation.

c) Access control- The gate automatically opens for authorized members. Thus replacing the security guard. The history of events is stored in the database.

d) Border control- Due to high security in border area, vehicle number plate is registered in entry or exits.

e) Section control- To measure average vehicle speed over long distances.

f) Traffic monitoring, automatic toll collection at toll plaza, traffic control, journey time measurement, stolen vehicle detection, marketing research etc.

This paper is organized in VII sections. Section II explains the image acquisition and preprocessing phase. Section III explains the number plate extraction methods, Section IV explain various character recognition methods and character recognition methods are discussed in section V. Conclusions and future work is given in Section VI and References are given in section VII.

II. IMAGE ACQUISITION & PREPROCESSING

Image acquisition means how to acquire the input image. In ANPR system digital camera of high resolution is used to acquire the input image. Images are taken in different background and illumination conditions. In pre-processing RGB image is firstly converted into gray level image and then into binary image. If the image consists of noise then filters are used to remove the noise. The basic purpose of pre-processing is to enhance the processing speed, to remove noise from input image, to improve the contrast of the image so that to increase and improve the visibility and quality of input image. For example, fig.4 shows input image, fig.5 shows RGB to gray-scale conversion of input image and fig.6 shows filtered gray scale image.

Fig.4. Input image

Fig.5. Gray Scale image

Fig.6. Filtered Image

III. NUMBER PLATE EXTRACTION

The captured input image has number plate covered by vehicle body, so by this step only number plate area is detected and extracted from whole body of vehicle. The number plate extraction phase influence the accuracy of ANPR system because all further step depend on the accurate extraction of number plate area. The input to this stage is vehicle image and output is a portion of image containing the exact number plate. Number plate can be distinguished by its features. Instead of processing every pixel, the system processes only the pixels that have these features. The features are derived from number plate format and the characters constituting the plate. Number plate color, rectangular shape of number plate boundary, the color change between the plate background and characters on it etc. can be used for detection and extraction of number plate area. The extraction of Indian number plate is difficult as compared to the foreign number plate because in India there is no standard followed for the aspect ratio of plate. This factor makes the detection and extraction of number plate very difficult. The various methods used for number plate extraction are as follows:-

A. Number Plate Extraction using Boundary/Edge information

Normally the number plate has a rectangular shape with certain aspect ratio, so number plate can be extracted by finding all possible rectangles from the input vehicle image. Various edge detection methods commonly used to find these rectangles. Sobel, Prewitt, Robert, Canny etc are edge detection filter that used to find edges. For boundary based extraction Hough transform (HT) is used. This method detects parallel lines in the image to detect number plate.

B. Number Plate Extraction using Texture Features

The difference between the character color and number plate background color known as texture is used to extract the number plate region. Vector quantization (VQ) is used to find the content of image regions. Gabor filter is commonly used for texture analysis.

C. Number Plate Extraction using Character Features

The number plate extraction based on character feature examines the image for the presence of characters in the input vehicle image. When the characters are found that region is extracted as number plate region.

D. Number Plate Extraction using Color Features

Some countries have specific colors for character color and background color of vehicle number plates; so the extraction of number plates can be done by locating these colors from the vehicle image. The idea is that the color combination of background color of number plate and characters is unique and this combination occurs almost only in number plate region so by finding this type of combination we can extract the number plate using this color combination feature. Genetic algorithm (GA) is good method for searching the number plate color.

E. Number Plate Extraction using Global image Information

Connected component analysis is an important technique. It scans the binary image and assign label to its pixels based on pixel connectivity.

F. Number Plate Extraction by combining two or more Features

For more accurate detection of number plate area, some method use two or more feature of number plate. The extraction method in this case is called hybrid methods. For example, color feature and texture feature is combined for extraction and edge detection operators are used with morphology operations for accurate detection of number plate area.

IV. CHARACTER SEGMENTATION

This step acts as bridge between the number plate extraction and character recognition phase. In this phase the characters on number plate area are segmented one by one. There are many factors such as image noise, space mark, plate frame, plate rotation and illumination variance etc. that make the character segmentation task difficult. The various methods of character segmentation are:-

A. Segmentation using Pixel Connectivity

In this method segmentation is performed by labeling the connected pixels in the binary number plate region. This method fails to extract characters when the characters are joined or broken.

B. Segmentation using Projection Profiles

In this method project algorithms are used for segmentation. Two types of projection algorithms are used. Vertical projection algorithm project the binary number plate vertically to determine the starting and end position of characters and then horizontal projection algorithm is used to extract each character alone.

C. Segmentation using Character Contours

Contour algorithm is used for character segmentation. Active contour model is established for segmentation.

D. Segmentation using Combining two or more Features

For effectively segmentation of characters two or more features are used. By combining two or more features we can get better result.

V. CHARACTER RECOGNITION

This is the last phase of ANPR system. The inputs to this phase are segmented characters. In this phase the segmented characters are recognized and output of this phase is license plate number. Character recognition in ANPR system has some difficulties. Sometimes the segmented or extracted characters do not have the same size. The better way to overcome this problem is to resize the characters into one size before recognition starts. The character recognition mainly done by following methods as: -

A. Character Recognition using Raw Data

Template matching is simple method used for recognition. The similarity between the template and characters is measured and the template that is most similar to the character is recognized as target. Template matching is done after resizing the segmented characters into the same size. Template matching consists of three steps:

1) Loading template and Character normalization: - This operation loads the template of character. Firstly we make template by taking A to Z alphabet and 0 to 9 number images. Read all image and store them in database. It becomes 36 character templates. Sometimes the segmented or extracted characters do not have the same size. The better way to overcome this problem that all the characters are normalized to predefined size.

2) Matching segmented characters with template characters: - The matching of segmented character with template character is done by many similarity measuring techniques such as correlation, Hamming distance, Bayes decision technique, Hausdorff distance etc. For example when the matching is done by correlation, the value of correlation is calculated by comparing the normalized segmented character image with each template character image and selecting the most relevant image and writes that character into text file.

B. Feature based character Recognition

Feature Extraction mainly done for two purposes. One is to extract the properties that can uniquely identify the character. Second is to extract properties that can differentiate between characters.

VI. CONCLUSION & FUTURE WORK

The Automatic Number Plate Recognition (ANPR) system consists of the following four stages: image acquisition and Pre-Processing, number plate extraction and number plate segmentation and number plate recognition phases. Number plate recognition systems are used for the purpose of effective control of vehicles. ANPR has number of applications as parking, access control, tolling, border control, detection of stolen cars, traffic control, traffic monitoring, journey time measurement etc. Extraction of number plate from background vehicle image is a difficult task. Vehicle motion, blurry image, uneven or low illumination, low resolution of the image, distorted characters, dirty plate, shadows or reflection etc. could influence the efficiency of the extraction. The complexity of smart number plate recognition work varies throughout the world .In addition to these problems; the extraction of Indian number plate is difficult as compared to the foreign number plate as there is no standard followed for the aspect ratio of Indian number plate. For the standard number plate ANPR system easily read and recognizes the exact licence plate number. In India this task becomes difficult due to variation in plate model and their size. Character recognition is also very difficult for Indian number plate so some flexible algorithm required for solving to remove these difficulties.

The future research of ANPR should concentrate on video-based ANPR using temporal information, multi-style plate recognition, high definition plate image processing, multi-plates processing, and ambiguous-character recognition and so on. There are many other open issues for the future research as follow:-

1) To deal with the illumination and low contrast problems, good pre-processing methods (image enhancement) should be used to make better quality input image.

2) For video-based ANPR we need to first extract the frames that have the passing cars. It needs either motion detection or frame differencing. Extracting the correct frame with a clear vehicle number plate image is another challenge especially when the vehicle speed is very fast.

3) Future research should concentrate on improving the recognition rate on ambiguous characters such as (O-0), (I-1), (A-4), (C-G) and broken characters.

4) Multi-style plate recognition, high definition plate image processing and multi-plates processing.

REFERENCES

[1] Christos-Nikolaos E. Anagnostopoulos, "License Plate Recognition: A Brief Tutorial", Intelligent Transportation Systems Magazine IEEE, Vol.6, Issue.1, pp.59 – 67, 2014.

[2] Shan Du,Mahmoud Ibrahim,Mohamed Shehata and Wael Badawy, "Automatic License Plate Recognition (ALPR):A State-of-the-Art Review", IEEE Transactions on Circuits & Systems for Video Technology, Vol. 23, Issue.2, pp.311-325, IEEE 2013.

[3] Sahil Shaikh, Bornika Lahiri, Gopi Bhatt and Nirav Raja,"A novel approach for Automatic Number Plate Recognition", International Conference on Intelligent Systems and Signal Processing (ISSP), pp.275 – 380, IEEE 2013.

[4] Norizam Sulaiman,"Development of Automatic Vehicle Plate Detection System",3rd International Conference on System Engineering and Technology, pp.130-135, IEEE 2013.

[5] Reza Azad and Hamid Reza Shayegh," New Method for Optimization of License Plate Recognition system with Use of Edge Detection and Connected Component",3rd International Conference on Computer and Knowledge Engineering (ICCKE 2013), pp. 21 – 25, IEEE 2013.

[6] Ronak P Patel, Narendra M Patel and Keyur Brahmbhatt,"Automatic Licenses Plate Recognition", International Journal of Computer Science and Mobile Computing (IJCSMC) , Vol. 2, Issue. 4, pp.285 – 294, 2013.

[7] Najeem Owamoyo, A.Alaba Fadele and Abimbola Abudu, "Number Plate Recognition for Nigerian Vehicles", Academic Research International Journal (ARIJ), Vol.4, Issue.3, pp.48-55, 2013.

[8] Sourav Roy, Amitava Choudhury and Joydeep Mukherjee, "An Approach towards Detection of Indian Number Plate from Vehicle", International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol.2, Issue.4, pp.241 – 244, 2013.

[9] Divya gilly and Dr. Kumudha Raimond, "License Plate Recognition- A Template Matching Method", International Journal of Engineering Research and Applications (IJERA), Vol. 3, Issue. 2, pp.1240–1245, 2013.

[10] Isack Bulugu,"Algorithm for License Plate Localization and Recognition for Tanzania Car Plate Numbers", International Journal of Science and Research (IJSR), Vol. 2, Issue.5, pp.12–16, 2013.

[11] Rupali Kate, "Number Plate Recognition Using Segmentation", International Journal of Engineering Research & Technology (IJERT), Vol.1, Issue.9, pp.1-5, 2012.

[12] P.Mohan Kumar, P.Kumaresan and Dr.S.A.K.Jilani," The Real Time Vehicle License Plate Identification System ", International Journal of Engineering Research and Development, Vol. 2, Issue. 4, pp.35-39, 2012.

[13] Hadi Sharifi Kolour, "An Evaluation of License Plate Recognition Algorithms", International Journal of Digital Information and Wireless Communications, ISSN2225-658X, pp.247-253, 2011.

[14] Kumar Parasuraman and P.Vasantha Kumar, "An Efficient Method for Indian Vehicle License Plate Extraction and Character Segmentation",A IEEE International Conference on Computational Intelligence and Computing Research, 2010.

[15] S. Hamidreza Kasaei, S. Mohammadreza Kasaei and S. Alireza Kasaei,"New Morphology-Based Method for Robust Iranian Car Plate Detection and Recognition", International Journal of Computer Theory and Engineering, Vol. 2, No. 2, pp.264-268, 2010.

Review on Image Fusion using Biorthogonal Wavelet Transform

Savroop Kaur
E.C.E Dept.
G.I.M.E.T, Amritsar
Ksavroop@gmail.com

Hartej Singh Dadhwal
E.C.E Dept.
G.I.M.E.T, Amritsar
dadhwalz@yahoo.com

***Abstract*—The process of image fusion combines the relevant information from two or more images of the same view to produce the image which is more informative. It is an important component of many applications such as inspection, night vision, medical diagnosis and multispectral targeting. Different methods of image fusion have been introduced in recent past. They can be either in spatial domain or in wavelet domain. In this paper a comparison between different techniques is done and results are in favor of biorthogonal wavelet transform which preserves the edge information and reduces the distortion which the spatial domain techniques can cause.**

Keywords—Image fusion, BWT, multifocus, multiresolution, registration.

I. INTRODUCTION

One of the most challenging applications in computer vision is to combine the relevant information from two or more images of the same view having different focus points. Due to different focus points the image can deliver different information even for the same view [1,2].

The imaging systems can be divided into two categories, the visible light and the non-visible light systems. The non-visible imaging systems have longer wavelength. They can focus on large area. On the other hand, visible light has shorter wavelength and so it can focus on small areas only. The region of interest can be shown a very clear image even, but in the entire imaging area if more than one area and the imaging depth of different objects need to observe, the visible imaging system is on the powerless condition. Because of the different type of sensors used in image capturing devices, and their principle of sensing, and also due to the limited depth of focus of optical lenses used in camera, it is possible to get several images of the same scene providing different information. Therefore, it is important to retrieve a new image with improved information. It becomes an important area of research. Image fusion can be used in many applications such as medical images, remote sensing, forensic science, surveillance etc [3].

The images used in image fusion should be registered. Image registration is a process of transforming different sets of data into one co ordinate system. The registration of image leads to preprocessing of Image Fusion. Two images taken in different angles of scene sometimes cause distortion. Most of objects are the same but the shape changes a little. For fusing the two images the pixels should be correlated. This is done by registration. Two images having same scene can register together using software to connect several control points. After registration, resampling is done to adjust each image that about to fuse to same dimension. Now, each image will be of the same size. Several interpolation approaches can be used to resample the image. This is because most of the approaches used all pixel-by-pixel fused. Images with the same size will be easy for fusing process. After resampling, fusion algorithm is applied. Sometimes the image is to be transferred into different domain, and then inverse transfer is necessary [4].

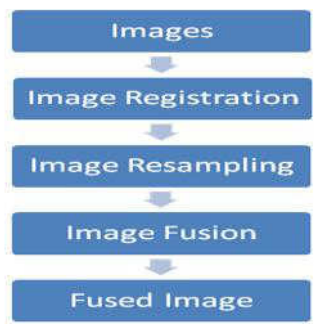


Fig.1. Preprocessing of Image Fusion

Various spatial domain and transform domain methods have been proposed in literature. Some popular spatial domain methods are Arithmetic Averaging, Principal Component Analysis (PCA), Sharpness criteria and Intensity Hue Saturation based on fusion scheme [5]. However these methods are complex and time consuming which are hard to be performed on real-time applications. These methods produce edge distortions in the fused image.

The fused image produced should contain the relevant information and also the artifacts should not be present in the fused image. Fusion must not include any noise or unexpected feature. Image fusion can be performed at three levels that is, pixel level, region level and decision level. The pixel level fusion methods are easy to implement and provide original information in the fused image as it deals with information associated with each pixel. In wavelet domain several methods are available to implement image fusion and these are based on multi scale decomposition of image. Here fusion rule is used i.e. first decomposition of source image is done and then selection of coefficients from the decomposed image. Decomposition of image produces coefficients in transform domain and fusion rule merges these coefficients without losing original information and without introducing any artifacts in the fused image.

II. WAVELET BASED IMAGE FUSION

Wavelets are finite duration oscillatory functions with zero average value. The irregularity and good localization properties make them better basis for analysis of signals with discontinuities. Two functions, scaling function f(t) and the translation parameter ψ (t) can describe wavelets[4]. The wavelet family can be defined as:

$$\psi_{j,m}(t) = a_0^{-j/2}\psi\left(a_0^{-j}t - mb_0\right)j, m\varepsilon Z \qquad (1)$$

The wavelet domain image fusion method used traditionally was based on Mallet algorithm is complex as it uses convolution to process large numbers of image data. Therefore it needs more memory space for the operations like read or write, which is costly for real-time imaging applications. In addition orthogonal filter of wavelet transform does not have the characteristics of linear phase, therefore phase distortion will lead to the distortion of the image edges and hence loss of the important image content. These limitations of orthogonal transform can be overcome by the use of biorthogonal wavelet transform.

III. BIORTHOGNAL WAVELET TRANSFORM

The biorthogonal wavelet transform is designed to provide symmetry property and the exact reconstruction by using two wavelet filters instead of one. The orthogonal filters cannot provide linear phase except for Haar filters. The general biorthogonal wavelet fusion method is shown in the diagram below:

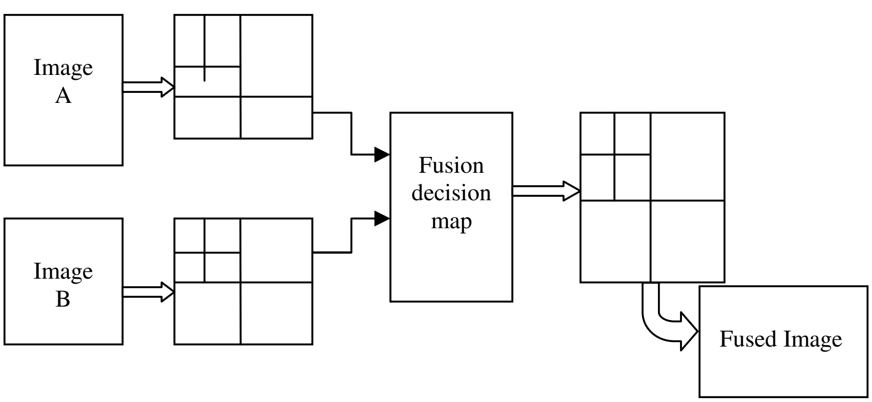


Fig.2. General Biorthogonal Wavelet based Image Fusion Scheme

Biorthogonal Wavelet transform is used for image denoising as it has the property of linear phase. The orthogonal filter of wavelet transform does not have the characteristics of linear phase, therefore the phase distortion will lead to the distortion of the image edge. This problem is reduced by the use of biorthogonal wavelet, as it contains spline wavelets. This will help in perfect reconstruction of the image by using Finite Impulse Response filters. This property is not present in orthogonal filters[6,7].

IV. PERFORMANCE MEASURES

In image fusion application, the aim of fusion is to process the significant parts of source images, like edges and regions with high contrast. This can be performed in quantitative measures such as[8,9]

(i) Standard Deviation ( ): It is the measure of the contrast of the fused image and is calculated as

$$\sigma = \sqrt{\sum_{i=0}^{L-1}(i - \bar{i})\, h_F(i)} \qquad (2)$$

(ii) Information Entropy (Q): It is the amount of information contained in the fused image and is calculated as

$$Q = \sum_{i=0}^{L-1} P_i \log_2 P_i \qquad (3)$$

(iii) Fusion Factor(FF): It is the sum of mutual information of source images and fused image.

$$FF = M_{AF} + M_{BF} \qquad (4)$$

M_{AF} and M_{BF} are the mutual information present between the source image and fused image.

V. EXPERIMENTAL RESULT

The image fusion is to be performed on different images which are multimodal and multifocus images which are based on BWT. The BWT based fusion is compared with spatial domain based image fusion methods like Principal Component Analysis (PCA), Sharpness criteria, linear fusion. White Gaussian noise with variance 0.01 to both input images.

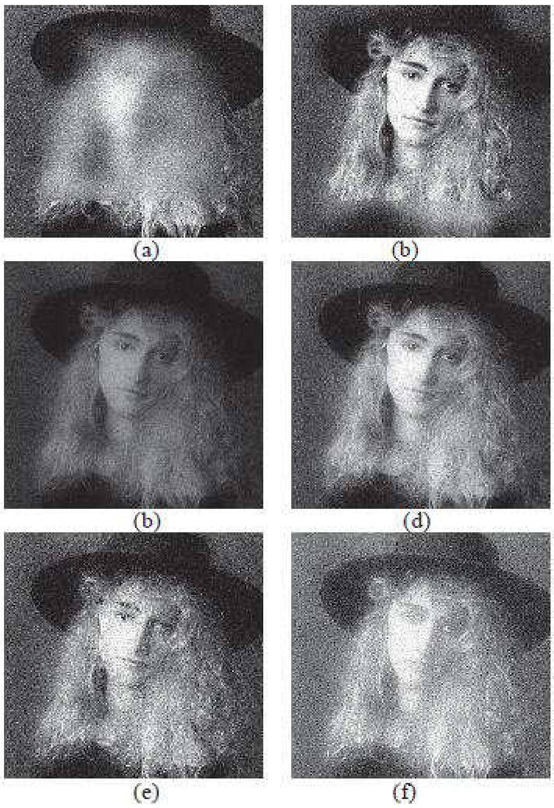


Fig.3.Illustrations of fusion results for multifocus images in presence of zero mean Gaussian noise with variance 0.01. (a) hoed A, (b) hoed B, (c) linear fused image, (d) PCA fused image, (e) Sharpness fused image, (f) The BWT

The qualitative performance does not provide all the information so the quantitative measures are needed

Table I. Comparative Quantitative performance measures

Fused Image	Method used	Fusion metric used		
		Fusion Factor	Entropy	Standard deviation
Hoed	Linear fusion	3.2487	7.1177	38.4703
	PCA fusion	4.3701	7.5261	55.6481
	Sharp fusion	4.1111	7.6594	60.4910
	BWT	4.4001	7.7546	61.7299

VI. CONCLUSION

This paper puts forward an image fusion algorithm based on Wavelet Transform i.e. Biorthogonal Wavelet Transform. It includes multiresolution analysis ability in Wavelet Transform. Image fusion is a process of combining important information from two or more images and generating a new image which is more informative. Biorthogonal wavelet transform provide a framework in which an image is decomposed, with each level corresponding to a coarser resolution band. The wavelet sharpened images have a very good spectral quality. It can retain information of individual images like edges, lines, curves, boundaries in the fused image in better way.

REFERENCES

[1] R. S. Blum and Z. Liu (Eds.), “Multi-sensor image fusion and its applications” (special series on Signal Processing and Communications), Taylor & Francis, CRC Press, July 2005.

[2] Yi Chai, Huafeng Li and Zhaofei Li, “Multifocus image fusion scheme using focused region detection and multiresolution,”,2011.

[3] Ren, Haozheng, Yihua Lan, and Yong Zhang. "Research of multi-focus image fusion based on M-band multi-wavelet transformation." In Advanced Computational Intelligence (IWACI), 2011 Fourth International Workshop on, pp. 395-398. IEEE, 2011.

[4] R.J.Sapkal, S.M.Kulkarni, “Image fusion based on Wavelet Transform for Medical Application.

[5] V.P.S.Nadiu and J.R.Raol, “Pixel-level image fusion using wavelets and principal component analysis.” 2008.

[6] W. Sweldens, “The Lifting Scheme: A Construction of second generation wavelets”, SIAM J. Math. Anal., 1997.

[7] W. Sweldens, “The Lifting Scheme: A Custom Design construction of Biorthogonal”, Wavelets Appl. Comput. Harmon. Anal., Vol. 3, 1996.

[8] S. Li, B. Yang, and J. Hu, “Performance comparison of different multi-resolution transforms for image fusion”, Information Fusion, Vol. 12, pp. 74-84, 2011.

[9] M. Deshmukh and U. Bhosale, “Image fusion and image quality assessment of fused images”, Int J Image Process, Vol. 4, pp. 484–508, 2010.

Comparative Analysis of Image Filtering Algorithms

Suraj Pal
Department of Computer Science
Guru Nanak Dev University
Amritsar, India
ersurajpal8@gmail.com

Amit Chhabra
Department of Computer Science
Guru Nanak Dev University
Amritsar, India
amit.cse@gndu.ac.in

***Abstract*— This paper has focused on the different image filtering techniques. Image filtering is one of the most important techniques for increasing the quality of the image as it act as the pre treatment to the images. All the other image processing techniques like image segmentation, image enhancement etc. are performed after image filtering. Image filtering is used to remove the different noisy pixels present in the image. To increase the picture quality image filtering can specifically remove the different kind of unwanted noises present in the given image by keeping the quality of the image. The main objective of this paper is to compare the different existing image filtering techniques using different parameters like noise density, peak to signal ratio, quality of image etc.**

***Keywords*— salt and pepper noise, Gaussian noise, poisson noise, speckle noise.**

I. Introduction

Digital image is an important source to obtain information in many scientific fields. In image processing whole image is divided into small unit called pixel, and then various operation have been carried out.When an image is acquired by a camera or other imaging system, the image get corrupted by random variations in intensity values, called noise which will reduce the image quality. Various reasons could be due to imperfection in image collection system, transmission medium and imaging system.

Some common types of noise are salt and pepper noise, impulse noise, and Gaussian noise. The presence of single dark pixels in bright regions, or single bright pixels in dark regions, is called salt and pepper noise. The analogy to real life is obvious. Often, salt and pepper noise is the natural result of creating a binary image via thresholding. Salt corresponds to pixels in a dark region that somehow passed the threshold for bright, and pepper corresponds to pixels in a bright region that were below threshold. [8]

Noise when get added to image destroy the details of it.Noise elimination is a key step in image processing, since much of the image post-treatment e.g. edge detection, pattern recognition, and image segmentation is depended effect and quality of the noise elimination. Image visibility restoration, these days, has become an active research area in the domain of Image processing. Despite several researches has undertaken in number of different fields, Image visibility restoration stands out uniquely with its own recognition.

Today's world is globe of Internet wherein information is needed to be exchanged across this globe, within fraction of second. This information may be composed of text, videos, or images. Images while transmissions over the communication media, are get corrupted due to insertion of noise. In order to recover the original image, the noise should be removed resulting as the concept of Image filtering. Following picture estates the purpose of image filtering.

Fig 1. (a) Before Filtering (b) After filtering

II. Various Types Of Noises

Although there are different kind of noises present in the image but each of them belong to a particular type of noise described below:-

A. Salt and Pepper Noise

Various terms that are used for this type of noise are random noise, impulse noise or independent noise. This particular type if noise appears in the form of black and white dots in the image because of which it is named as salt and pepper noise. This type of noise is caused due to

sudden changes in the image signal .The noisy pixels of these types could be easily distinguished from the normal pixels because the intensity of the noisy pixel is totally different from the surrounding ones and has no relation with them at all.

B. Gaussian Noise

Unlike the salt and pepper noise in Gaussian noise the noisy pixels are not present in the small portion of the image in fact they are evenly distributed all over the image. The name is given on the behalf of the person who gave a value to the corrupted pixel as random Gaussian distributed noise value and due to the even distribution each pixel in the image is the sum of the true pixel value and random Gaussian distributed value. Cause of this type of noise is the random fluctuations in the image signal.

C. Poisson noise

It is also called as photon shot noise. Poisson noise is a type of electronic noise present in the lighter parts of the image. In the image suffering from this type of noise each pixel has different noise value. This type of noise is caused when the numbers of photons carry energy which is small enough to give rise to the detectable statistical fluctuations in the measurement.

D. Speckle noise

It is a kind of granular noise that degrades the quality of the image and occurs in the Synthetic Aperture Radar (SAR) images. This noise is originated because of coherent processing of back scattered signals from multiple distributed points. Speckle noise results from random fluctuations in the return signal from an object that is no bigger than a single image-processing element. Speckle noise in SAR is generally more serious, causing difficulties for image interpretation.

III. RELATED WORK

A. First Order Neighbourhood Decision Based Median Filter

Priyadarshi Kanungo et al. [2012] [9] proposed a First Order Neighbourhood Decision Based Median Filter for the removal of noisy pixels in order to restore the grayscale images. The proposed algorithm detected the noisy pixels and considered the First Order Neighbourhood pixels of that detected pixel for further decision. The author has proved that the proposed algorithm is better than the standard median filter algorithm and modified decision based unsymmetric trimmed median filter in terms of peak signal to noise ratio and signal operation.

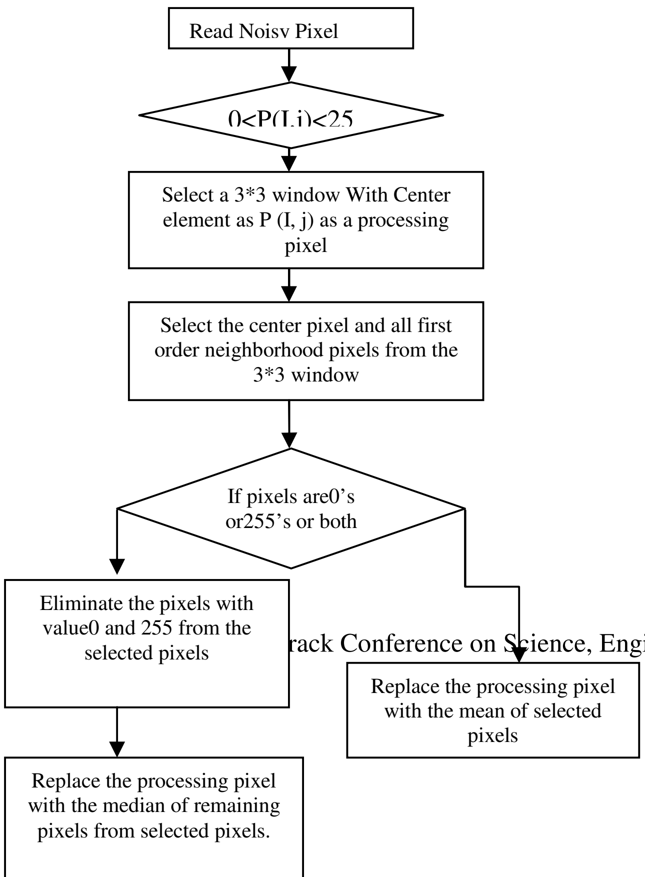


Fig.2. Flow diagram of the algorithm

B. High Performance Decision Based Median Filter for Salt and Pepper Noise Removal in Images

Golam Moktader Daiyan et al. [2012] [2] proposed a high performance decision based median filter for removal of salt and pepper in image.

Step1.Detection of noise is done by expanding the mask until 7*7 to maintain local information extraction.

Step2. If the noise free median value is not available at 7*7 processing window, the last processed pixel take into consideration if it is noise free.

Step3. If the last processed pixel is noisy, the algorithm select a window size with 15*15 dimension and calculate the number of 0's and 255's in the processing window.

Step4. Then replace the processing pixel with 0 or 255 which is more in number in the selected window.

C. Unsymmetric trimmed median filter

S.Esakkirajan et al [2011] [9] proposed this algorithm which detects the noisy pixel as the processing pixel. The steps of are as follows:

Step1. Select 2-D window of size 3*3. Assume that the pixel being processed is $P_{ij.}$

Step2. If0< P_{ij} <255, then P_{ij} is an uncorrupted pixel and its value is left unchanged.

Step3. If P_{ij} =0 or P_{ij} =255, then P_{ij} is a corrupted pixel. then two cases are possible as given:Case1:If the selected window contains all the elements as 0's and 255's, then replace P_{ij} with the mean of element of the window.Case2:If the selected window contains not all elements as 0's and 255's ,then eliminate 0s and 255's and find the median value of the remaining elements. Replace P_{ij} with the median value.

. D. Standard median filter

PENG Lei et al [2001] [7] proposed a median filter which is normally used to reduce noise in an image, somewhat like the mean filter.

Methodology

Step1: Considers each pixel in the image and look at its nearby neighbours.

Step2: Decide whether or not it is representative of its surroundings.

Step3: Sort all the pixel values from the surrounding neighborhood into numerical order

Step4: Calculate the median of those values.

Step5: Then replace the pixel being considered with the middle pixel value.

E. Adaptive Median Filtering

H.Moro et al [2001] [7] proposed the adaptive median filtering which has been applied widely as an advanced method compared with standard median filtering. The Adaptive Median Filter performs spatial processing to determine which pixels in an image have been affected by impulse noise.

Purpose

1). Remove impulse noise

2). Smoothing of other noise

3). Reduce distortion, like excessive thinning or thickening of object boundaries

F. A new kind of weighted median filtering algorithm used for image Processing

DENG Xiuqin et al. [2008] [3] proposed new adaptive weighted median filtering algorithm which aimed at the excellence and shortcoming of the traditional median filtering algorithm. The author has shown the advantage of large window filtering and small window filtering.

Methodology:-

Step1: Noise detection over the image

Step2: Adaptive determination of the filtering window size

Step3: Noise pixels filtering

G. Mean Filter

S. Rakshit et al [2006] [4] have proposed a novel method for mean filtering that has been tested on a variety of greyscale images and neighbourhood sizes with promising results

Methodology:-

Step1: Consider an L level image $F\ (P \times Q)$, whose pixel greylevels are stored in a 2-D array, say, *data* [P] [Q] such that *data* [0] [0]

Step2: *data* [P −1] [Q−1] contains the first and last pixels, respectively

ALGORITHM	Peak Signal To Noise Ratio (PSNR)	TYPE OF NOISE REMOVED	ADVANTAGES	LIMITATION
1.First Order Neighborhood Decision Based Median Filter	21.59 when noise density upto 50%	Impulse noise	Restoration of highly corrupted images.	Unable to maintain the original details of the image.
2.Decision Based Median filter	22.87 when noise density is 10%	Salt and pepper noise	Less computation time.	Efficient only when the noise density is very low.
3.Unsymmetric Trimmed Median Filter	Good but when noise density is below70%	Salt and pepper noise	Preserve the fine details of the image	Lesser efficient during high noise density and peak signal to noise ratio is low.
4.Standard Median Filter	20.37 when noise density is 10%	Impulse noise	This method is simple and efficient in terms of noise removal capability.	Blurs the image detail such as points and short thin lines.
5.Adaptive Median filter	25.66 when noise density is 20%	Non-impulsive noise and salt and pepper noise.	Maintains the details such as points and short thin lines	Could perform only when the noise density is below 20%.
6.Weighted median filtering	23.3625 at 60% noise density	Impulse noise	Detail preserving capabilities due to self adjusting size of the filtering window.	Could perform good only when the noise density is upto 60%
7.Mean Filter	19.95 upto 10% noise density	Gaussian noise	Relative time requirement reduces with increase in image size.	Very complex technique.
8.Prescanned minmax centre weighted filter	16.431 when noise rate is 0.1	Impulse noise	This algorithm has solved signal distortion problem.	Preserves details only when noise density is quite low.
9.Fast Filtering Algorithm	20.23 upto 10% noise density	Salt and pepper noise	Few modifications on the original image structures	Obtaining four different sub images is a difficult task.
10.Feed Forward Algorithm	20.56 upto 10% impulse noise	Uniform noise and impulse noise	Good performance in the presence of the long-tailed impulse noise and short tailed uniform noise.	Whole procedure is time consuming.

Table I.Comparison Table of Algorithms

Step3: Then image is *padded* over a rectangular neighbourhood window $m \times n$ (m & n are odd positive integers

Step4: The average greylevel of a pixel *data*[*row*][*col*] over an $m \times n$ neighbourhood is then calculated.

H. Prescanned minmax centre weighted filter

J.H.Wang et al [1999] [8] proposed the prescanned minmax centre-weighted filter which is capable of restoring the images severely corrupted by impulsive noises.

Methodology;

This filter works in two phases

Phase1: in the first phase prescanning operation is applied to the input image such that maximum and minimum of each ranked set of the running window are grouped and filtered.

Phase2: All the remaining samples are filtered in this particular phase.

IV.CONCLUSION

The technique of image filtering plays an important role in the world of digital image processing. Different techniques for the process of image filtering have been discussed in this paper along with their comparison. On comparing them it has been found that by keeping into account the quality of image and peak signal to noise ratio decision based median filter is the best method to remove the noises from the image at high noise density. However even this particular method is not able to maintain the actuality of the image because it also reduces the quality of the image up to some extent. Therefore a more optimal method could be designed for image filtering which along with the noise removal also maintains the originality of the image without reducing its quality because this is the most important parameter for the digital image

.

REFERENCES

[1] Barney, Blaise. "Introduction to parallel computing." Lawrence Livermore National Laboratory 6, o. 13 (2010): 10.

[2] Daiyan, Golam Moktader, M. A. Mottalib, and Muhammad Mizanur Rahman. "High Performance Decision Based Median Filter for Salt and Pepper Noise Removal in Images."

[3] Deng, Xiuqin, Yong Xiong, and Hong Peng. "A new kind of weighted median filtering algorithm used for image

[4] Processing." In Information Science and Engineering, 2008. ISISE'08. International Symposium on, vol. 2, pp. 738-743. IEEE, 2008.

[5] R.C. Gonzalez, R.E.Woods, Digital Image Processing, second ed., Pearson Education, 2002.

[6] Kaur, Preeti, and Anil Kumar. "Utilization of Multicore Processors Using Parallel Image Enhancement Algorithms."

[7] Li, Li, and Li Xiuhua. "A new adaptive median filtering based on multidirectional criterion." In MultiMedia and Information Technology, 2008. MMIT'08. International Conference on, pp. 342-344. IEEE, 2008.

[8] H.Moro, T.Watanabe, A.Taguchi and N. Hamada, "On the adaptive algorithm and its convergence rate improvement of 2-Dlattice filter", 1988 IEEE International Symposium on Circuits and Systems, Proceeding vol. 1 of 3, pp. 430-434

[9] WANG, J.H., and LIN, L.D.:'Improved median filtering using minmax algorithm for image processing ', Electron. Lett, 1997, 33, (16), pp.1362-1363.

[10] Samantaray, Aswini Kumar, and Priyadarshi Kanungo. "First order neighborhood decision based median, 2012. Filter." In Information and Communication Technologies (WICT), 2012 World Congress on, pp. 785-789. IEEE

[11] Youlian, Zhu, Huang Cheng, and Zhai Lifang. "A Median Image Filtering Algorithm Based on Statistical Histogram." In Measuring Technology and Mechatronics Automation (ICMTMA), 2013 Fifth International Conference on, pp. 17-20.

[12] L. B. Jackson, Digital Filters and Signal Processing. Kluwer Academic Publishers, 1989.

[13] Pushpavalli, R. & Sivaradje, G. (2013). Feed forward back propagation algorithm for eliminating uniform noise and impulse noise. European Journal of Natural and AppliedSciences,1(1),10-27.

Review of Fog Removal Techniques from Color Images

Rajbir Kaur
Dept. of Electronics and Communication Engineering
Amritsar College of Engineering and Technology
Amritsar, Punjab, India.
rajbirkaurr@yahoo.com

Vijay Kumar Banga
Dept. of Electronics and Communication Engineering
Amritsar College of Engineering and Technology
Amritsar, Punjab, India.
vijaykumar.banga@gmail.com

***Abstract*--This paper presents a review of well-known fog removal techniques. Fog formation is because of attenuation and airlight. Attenuation reduces the contrast and airlight increases the whiteness in the scene. Atmospheric conditions created by suspended particles, such as fog and haze, severely degrade image quality. Fog removal from a single image of a weather-degraded scene found to be a difficult task, because the fog is dependent on the unknown depth information. The overall objective of this paper is to explore the various techniques for efficiently removing the fog from digital images. This paper ends up with the short comings of the existing techniques.**

Keywords—Fog Removal, Bilateral Filter, CLAHE, Image Contrast.

I. INTRODUCTION

Fog is a collection of liquid water droplets or ice crystals suspended in the air at or near the Earth's surface. The term fog is typically distinguished from the more generic term cloud in that fog is low-lying, and the moisture in the fog is often generated locally (such as from a nearby body of water, like a lake or the ocean, or from nearby moist ground or marshes)[1].

Reduced visibility degrades the quality of the image and the performance of the computer vision algorithms such as observation, object detection, tracking and segmentation. Reduced visibility in bad climate such as fog, smog and cloud are caused due to the water droplets present in the air. These droplets are very small and gradually float in the air. Due to these droplets in the air when the images are captured in such condition the light gets spread before reaching the camera which reduces the quality of the image.

A light beam travels from a scene point through the atmosphere, gets attenuated due to the distributed by the atmospheric particles, this phenomenon is called attenuation which reduces the contrast in the scene. Light that is coming from the source is scattered by fog and part of it also travels towards the camera [2]. Air-light increases with the distance from the object.

When there is fog, smog, rain and other bad weather conditions, a number of deterioration will occur in the image obtained by the camera, which reduces the purpose of the outdoor monitoring system and causes difficulty for the extraction of image features [3]. But the image defogging tools can effectively remove the weather effects of the image, it is important in the restoring of the image contrast. The natural images to fog algorithm has become the hotspot in the computer vision system and image processing field.

For obtaining the fog free images contrasts are restored from a single input image by maximizing the contrasts of the direct transmission while assuming a smooth layer of air-light. The fog free image must have better contrast compared to the input foggy image and fog removal is processed by maximizing the local contrast of the restored image. The most of patches in the fog free images have some pixels which have low intensity at least at one color channel. [2]

The current fog removal method can be divided into two categories: image enhancement and image restoration. Image enhancement excludes the reasons of fog corrupting the image quality. This method is applicable to a broader scope; it can improve the contrast of haze image, but can also lead to the loss of information in the image. On the other hand, image restoration first studied the physical features of the image and the model for the degradation of the fog is applied to it. In computer vision, the optical model, this is widely used to approximate the image formation in bad weather

$$I(x) = j(x)t(x) + (1 - t(x))A \quad (1)$$

Various Fog Removal techniques are discussed in this paper such as:

- Single Image Defogging
- Contrast Limited Adaptive Histogram Equalization
- Image Haze Removal of Wiener Filtering
- Dark channel prior
- Single Image Fog Removal Using Bilateral Filter

II. SINGLE IMAGE DEFOGGING

This method uses a single image defogging algorithm which deals with the global or local rectifications. First of all the estimation of atmospheric light and transmission path is carried out and then the scene radiance is recovered. After which the global and local rectification is done to get the fog free image.[3]

Estimation of transmission map is done according to the following equation. We assume that transmission t(x) at local patch (x) is centered at x is constant

$$\min_{y\epsilon\Omega(x)}\left(I^c(y)\right) = t(x)\min_{y\epsilon\Omega(x)}\left(J^c(y)\right) + \left(1 - t(x)\right)A^c \quad (2)$$

Where t(x) is estimation of transmission map

And c ϵ\{r,g,b\}

As A^c is always positive so dark channel normalization of haze image can be achieved at[4]

$$J^{dark}(x) = \min_c\left(\frac{\min_{y\epsilon\Omega(x)}\left(I^c(y)\right)}{A^c}\right) \quad (3)$$

And the transmission estimation t(x) can be given as

$$t(x) = 1 - \min\left(_c\frac{\min_{y\epsilon\Omega(x)}\left(I^c(y)\right)}{A^c}\right) \quad (4)$$

Once the atmospheric light and transmission map t(x) is estimated[5],the scene radiance is recovered according to the following equation

$$J(x) = I(x) - \frac{A}{\max\left((t(x),t_\circ)\right)} \quad (5)$$

In conditions of bad weather, mostly in daylight where the sky is usually dark, we can overlook the presence of the sunlight, and assume that the atmospheric light is globally constant [3]. To improve the reality, the brightest pixels in dark channel image are stored and then among them, the pixels with maximum power in the input image are selected as the atmospheric light.

Now the global rectification is done, in which first of all limitation of dark channel prior is done. The fog removal method based on dark channel prior will improve diffusion of image as well as improve definition. If the scene objects are big enough and similar to the atmospheric light, the recovered scene vivacity on them will get higher diffusion. Then local rectification is done in which image matting and iterative image optimization is done. Image mating is separating foreground object from the background involving both full and partial coverage of pixels.

III. CLAHE

CLAHE known as Contrast Limited Adaptive Histogram Equalization is used for defogging color images [6].The color images taken by camera in foggy conditions are transformed from RGB color space to HSI color space. The motive of converting is because the HIS represents colors is same how the human eye senses colors. Then, the intensity component of the image is processed by CLAHE. The hue and saturation are unaffected. Finally, the image processed in HSI color space is converted back to RGB color space

A. RGB Color Space

The RGB color space is color space based on the RGB color model. It is defined by the three colors red(R), green (G) and blue (B) additive primaries, and can produce any chromaticity that is the triangle defined by those primary colors.

B. HSI Color Space

The HSI color space is very important model for image processing applications because it understands colors similarly how the human eye represents colors. Thiscolor model presents every color with three components: hue (H), saturation (S) and intensity (I).

C. Color Conversions

The color conversions from the RGB color space toe HSI color space is done with the help of following equations. The HSI (hue, saturation and intensity) is represented as:

$$I = \frac{R+G+B}{3}$$

$$H = \frac{1}{360}\left[90 - \frac{\arctan(2R-G-B)}{\sqrt{3(G-B)}} + \{0, G > B; 180\}\right] \quad (6)$$

$$S = I - \frac{\min(R,G,B)}{I}$$

And for the conversion purpose RGB is given as[6]

$$R = \frac{1}{\sqrt{3}}\left[1 + \frac{s\cos(H)}{(\cos 60 - H)}\right] \quad (7)$$

$$B = \frac{I}{\sqrt{3}(1-S)}$$

$$G = \sqrt{3}I - R - B$$

Once the color conversion is done then the histogram equalization is done.

D. Contrast Limited Adaptive Histogram Equalization

The CLAHE method applies histogram equalization to a background region. Each pixel of original image is centered at the background region.[7] The original histogram is briefed and the briefed pixels are redistributed to each gray level.

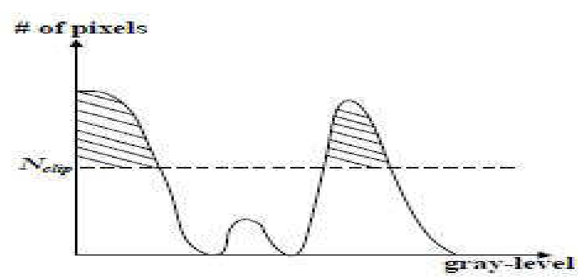


Figure 1 (a) Original Histogram

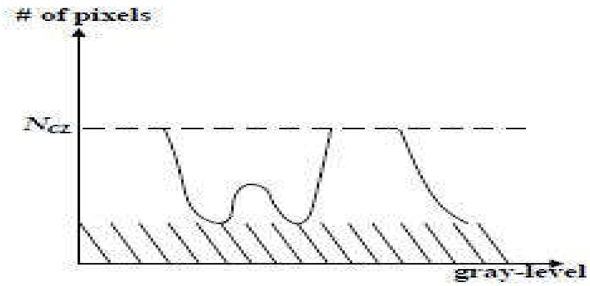


Figure 1 (b) Clipped Histogram

Contrast-limited histogram equalization techniques (CLAHE) Sharp field edges can be maintained by selective enhancement within theboundaries. Selective enhancement is done by detecting the field edge in a foggy image and then onlythose regions of the image are processed that lie inside the field edge. Noise can be reduced whereas the high spatial frequency content of the image is kept constant by applying a combination of CLAHE, median filtration and edge sharpening [7].

Figure 2(a) Input Foggy Image

Figure 2(b) Restored Image by CLAHE

IV. IMAGE HAZE REMOVAL OF WIENER FILTERING

The wiener filtering method can make image more clearer using filtering and reducing the noise in the image. And it can make a better quality of the image contrast after dark colors processing image without losing the details of the image [8]. This method is a mending method, which is a combination of the numerical character of the function and noise.

In this method first of all the input foggy image is taken and then dark channel method is used to estimate the airlight of sky brightness. Then the median filter function is estimated and wiener filtering is applied to the image to obtain a clear fog free image.

The atmospheric scattering model used in these methods is given as [9]

$$I(x) = J(x)t(x) + (1 - t(x))A \qquad (8)$$

$I(x)$ is the color value in the x position of foggy image

$J(x)$ is the color value in the x position of defogging image

A is sky brightness and is a constant vector

Then the dark channel of the foggy image is estimated.

$$J^{dark}(x) = \min_c \left(\min_{y \in \Omega(x)} J^c(y) \right) = 0 \qquad (9)$$

As A is a positive value so dividing both sides by A and taking the minimum values of all the three color channels we get smoothness of constraint about t

$$t = 1 - \min_c \left(\frac{\min_{y \in \Omega(x)} I^c(y)}{A} \right) \qquad (10)$$

The image seems to be unreal so some amount of fog is retained that covers the distant scenery. Now the median filter is used to estimate the media function of the image. Median filter is a nonlinear filter which is used to recover the signal. First of all the grayscale pixels in the window are rearranged then the new pixel grayscale value is placed in the center of the original window and it became newly-ranked middle grey value [10].

Each gray image $I_1(x)$of every pixel in the fogimage $I(x)$ in the space based on dark colors, that is:

$$I_1(x) = \min \quad I^c(x) \qquad (11)$$

The wiener filtering method can make image smoother using Average filtering, dropping the noise in the image. And it can make a better change of the image contrast after dark colors processing image without losing the details of the image edge.

Figure 3(a) Input foggy image

Figure 3 (b) Restored Image by Wiener Filtering

V. DARK CHANNEL PRIOR

Dark channel prior is information of outdoor image haze removal. It is based on the fact that the local patches whose intensity is very low in at least one color channel in the foggy images. This information is used by dark channel prior method to improve the quality of image. The pixels whose intensity is low the thickness of the fog can be estimated and the high quality fog free image can be recovered. High quality depth map of the image can also be obtained in this method [11].

Fog as light spreading effect in amount to the depth of the scene adds Atmospheric light effect, hence minimum intensity of the dark channel of the foggy image is not zero.

The channel which has theminimum values of RGB is named as Dark Channel (DC). In foggy images, dark channel provides an estimation of the fog element present in the image. The dark channel phenomenon is unable to restore the images when scene element are very similar to Atmospheric light, as an example a white wall over a large local region. Dark Channel provides a good solution for most of the outdoor haze images.

First of all the atmospheric scattering model is obtained for the foggy image which is:

$$I(x) = J(x)t(x) + A(1 - t(x))$$

Where x is location of the pixel, Iis the observed intensity, J is the scene radiance, A is atmospheric global light and t is transmission medium.

Now the dark channel prior for the foggy outdoor image is estimated. For this at least one channel should have the low intensity at some pixels.[12] For the image J it can be given as:

$$J^{dark}(x) = \min_{c\epsilon\{r,g,b\}}(\min_{y\epsilon\Omega(x)}(J^c(y)))$$

$J^c$is color channel of J, (x) is local patch centered at x.

The channel which has theminimum values of RGB is named as Dark Channel (DC). In foggy images, dark channel provides an estimation of the fog element present in the image.

Figure 4 (a) input image (b) estimated transmission map

Figure 4(c) refined transmission map (d) output haze free image

VI. SINGLE IMAGE FOG REMOVAL USING BILATERAL FILTER

The efficiency of bilateral filter for estimating the image depth map is the base for the implementation of this algorithm. This requires pre and post processing steps. Histogram equalization is used as a preprocessing. This preprocessing increases the contrast of the image earlier the fog removal and results in the better judgment of air light map.

Histogram stretching is used as a post processing, which increases contrast of the fog removed image which is more often a low contrast image. Transfer function of the stretching is adjusted according to the image content. This algorithm does not require any user involvement and is applicable for color as well as gray level images [2].

The air light map is represented as

$$A(x, y) = I_{\infty}(1 - e^{-kd(x,y)})$$

Where I_{∞} is global atmospheric constant and is also known as sky intensity.

In this method first of all preprocessing is done in which histogram equalization is carried out [14].Then the preprocessing on the foggy image results in the better estimation of the air light map. Then the air light map is refined using the bilateral filter which restores the foggy image. The restored image has low contrast so the post processing is performed in which histogram stretching is carried out which results in the better image quality by improving the contrast of the restored image.It has a wide application in tracking and navigation, consumer electronics and entertainment industries.[2]

Figure 5 (a) has shown the input image for experimental purpose. It has been clearly shown that the image contains a foggy image so the image is a foggy image, so the main objective is to remove it.

Figure 5 (a) Foggy image

Figure 5 (b) has shown the result of the existing dark channel prior algorithm. It has been clearly shown that the fog has been removed efficiently. But still some improvement is required to enhance the result further.

Figure 5 (b)Single image fog removed by bilateral filter

VII. CONCLUSION

In this paper we review the existing techniques of image restoration of foggy images. We discussed the techniques to detect the airlightmap in the foggy images and then restore then by using various techniques. Fog removal algorithms has become more useful for many vision applications. It is found that most of the existing researchers have neglected many issues; i.e. no technique is better for different kind of circumstances.

The existing methods have neglected the use of gamma correction and histogram stretching to reduce the noise problem which will be presented in the output image of the existing fog removal algorithms. To reduce the problems of existing literature a new integrated algorithm will be proposed in near future that will modify the dark channel prior with CLAHE to enhance the results further.

REFERENCES

[1] Wei, Sun, and Han Long. "A New Fast Single-Image Defog Algorithm." Intelligent System Design and Engineering Applications (ISDEA), 2013 Third International Conference on. IEEE, 2013.

[2] Tripathi, A. K., and S. Mukhopadhyay. "Single image fog removal using bilateral filter." Signal Processing, Computing and Control (ISPCC), 2012IEEE International Conference on. IEEE, 2012.

[3] Chen, Mengyang, et al. "Single image defogging." Network Infrastructure and Digital Content, 2009. IC-NIDC 2009. IEEE International Conference on. IEEE, 2009.

[4] He, Kaiming, Jian Sun, and Xiaoou Tang. "Single image haze removal using dark channel prior." Pattern Analysis and Machine Intelligence, IEEE Transactions on 33.12 (2011): 2341-2353.

[5] Shwartz, Sarit, EinavNamer, and Yoav Y. Schechner. "Blind haze separation."Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. Vol. 2. IEEE, 2006.

[6] Xu, Zhiyuan, Xiaoming Liu, and Xiaonan Chen. "Fog removal from video sequences using contrast limited adaptivehistogramequalization."Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference on. IEEE, 2009.

[7] Oakley, John P., and Brenda L. Satherley. "Improving image quality in poor visibility conditions using a physical model for contrast degradation." Image Processing, IEEE Transactions on 7.2 (1998): 167-179.

[8] Xu, Zhiyuan, Xiaoming Liu, and Na Ji. "Fog removal from color images using contrast limited adaptive histogram equalization." Image and Signal Processing, 2009. CISP'09. 2nd International Congress on. IEEE, 2009.

[9] Guo, Fan, et al. "Automatic Image Haze Removal Based on Luminance Component." Wireless Communications Networking and Mobile Computing (WiCOM), 2010 6th International Conference on. IEEE, 2010.

[10] Bao-chang, Xu, Zhang Ding-yuan, and LuoXiong-lin. "A new neural network soft sensing method based on the multi-resolution data fusion." Intelligent Control and Automation (WCICA), 2010 8th World Congress on. IEEE, 2010.

[11] Shuai, Yanjuan, Rui Liu, and Wenzhang He. "Image haze removal of wiener filtering based on dark channel prior." Computational Intelligence and Security (CIS), 2012 Eighth International Conference on. IEEE, 2012.

[12] Xu, Haoran, et al. "Fast image dehazing using improved dark channel prior."Information Science and Technology (ICIST), 2012 International Conference on. IEEE, 2012.

[13] Xu, Haoran, et al. "Fast image dehazing using improved dark channel prior."Information Science and Technology (ICIST), 2012 International Conference on. IEEE, 2012.

[14] Tarel, J-P., et al. "Improved visibility of road scene images under heterogeneous fog." Intelligent Vehicles Symposium (IV), 2010 IEEE. IEEE, 2010.

Analysis of Various Algorithms for Automatic Speech Recognition: A Case Study

Sakshi Dua
PTU
Jalandhar, Punjab, India
sakshi_nancy@yahoo.in

Ashu Gupta
ApeeJay Institute of Management Technical Campus
Jalandhar, Punjab, India
guptashu1@rediffmail.com

Abstract—**Research on Automatic Speech Recognition is going on and on. No language is left out upon which work for ASR has not been done, but which algorithm can be proven as the 'best approach' for enhanced accuracy and Speedy recognition is still under the researchers' microscope. So many approaches are there for recognizing the speech spoken by speaker but the choice for highly accurate algorithm is still a keen question because under different circumstances different approaches or algorithms can work in a different manner. So this is one of the crucial stages where researcher has to decide which algorithm is to be opted for the process of recognition yet it must be optimal too. The primary objective of this paper is to study and compare the various algorithms that are in trend for Automatic Speech Recognition.**

Keywords- Accuracy, Algorithms, Recognition, Approaches.

I. INTRODUCTION

Speech is a natural way of communication, human beings communicates via speech, whereas, other species communicate with each other too by their natural way of expressing the feelings. Speech comprises of sender, transmission channel, and a finally a receiver [14]. Speech is an acoustic model that incorporates so many factors, such as syllables, phonemes, words and sentences. From engineering aspect, in speech, there is a sender who encodes and transmits the acoustic message it may be via air, whereas, at other end, receiver captures such signals and decode the message. Speech is transformation of speakers' emotions and feelings [11]. Automatic Speech Recognition is a concept that is associated with speech under which that translates the human speech into text, which further can be processed or displayed for human understanding [7]. For this, various Recognizers can be used that typically depends upon the application for which ASR is to be developed [2,3]. Various algorithms or various techniques are there which can be used in ASR. Gaussian Mixtures, Dynamic Time Warping (Dtw), Hidden Markov Model (Hmm) , Neural Network, Arificial Intelligence, Vector Quantization (Vq), Connectionist Approaches, (Artificial Neural Networks), Support Vector Machine (Svm), Time Slice Approach And Syllable Based Approach. These are the different techniques which are providing different accuracy rates, depending upon their applications and type of ASR.

II. NEED OF STUDY

The main concentration in this paper is on studying the various existing algorithms or approaches for the process of automatic speech recognition and depending upon their applications exploring the results of each algorithm. Before proceeding to get involve into working of ASR it is necessary to have good knowledge of numerous algorithms in prior.

III. PROBLEM STATEMENT

Study and compare the various algorithms existing for ASR and evaluation of their results, on the basis of their performance.

IV. OBJECTIVES

1) To study the entire process of ASR
2) To study the various algorithms for ASR
3) To evaluate the results of ASR by using some particular algorithms.

V. SPEECH RECOGNITION PROCESS

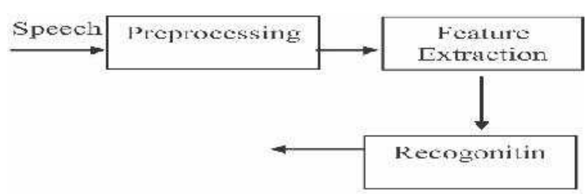


Fig.1. An overview of Speech Recognition Process [15]

A. *Speech signal:*

Speech signal is an analog signal, whose main characteristic is it varies time to time and to digitally process such signal, it becomes necessary to pass the signals through the process of Sampling. In this process, continuous time signals are sampled or converted to discrete time signals which are further converted to discrete time and finally to digital signals. Just because this conversion is quite slow process, mean while speech is been divided into some segments or frames where each of the frame posses some certain features and they can be extracted via some known techniques [6] [8].

B. *Feature extraction:*

Feature extraction is the process of spectral analysis where main objective is to extract the features from the frames or segments generated from speech signals. Under this phase of ASR, the thing that happens is to find out the parameters of signal waveform which are called as Features [13]. In other words, it can be concluded as, the main goal of feature extraction in ASR is to calculate the sequence of feature vectors that provide a clear picture about input or speech signal. It is carried out in these steps- First step is the analysis of speed. Second step involves compilation of static and dynamic feature of an extended feature vector. In third step, the compact form of these vectors is forwarded to Recognizer. Various methods for Feature Extraction in speech Recognition [1]-

- Principal Component Analysis (PCA)
- Linear Discriminant Analysis (LDA
- Independent Component Analysis (ICA)
- Linear Predictive coding
- Cepstral Analysis
- Mel-frequency cepstrum (MFFCs)
- Cepstral mean subtraction

C. *Recognition* :

There are three approaches [1] that are used for the purpose of speech recognition. They can be categorized as-

- Acoustic phonetic approach
- Pattern recognition approach
- Artificial intelligence approach

VI. VARIOUS ALGORITHMS FOR ASR

A. *HMM* :

This is a stochastic process in which the as a problem is modeled as a "doubly stochastic process", according to which the observed data is taken as result of being passed from "true" or "hidden" process via some second process. But the issue that arises with this problem is that number of Markov chains are not known here. Moreover, number of states and their associated probabilistic functions are unknown. Due to fact, all these properties are kept to be hidden; the name of the model is Hidden Markov Model [10]. Hidden Markov models are especially known for their application in temporal pattern recognition such as speech, handwriting, gesture recognition, part-of-speech tagging, partial discharges and bioinformatics.

Structure-

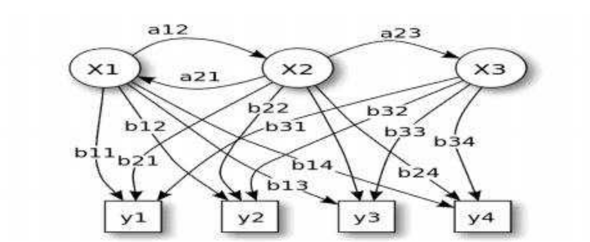


Fig. 2 . Probabilistic parameters of a hidden Markov model, example(http://en.wikipedia.org/wiki/Hidden_Markov_model)

x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Above diagram shows general architecture of HMM. Each of the oval shape is representing a random variable that can have any of a number of values. The random variable $x(t)$ is the hidden state at time t, $x(t) \in \{ x_1, x_2, x_3 \}$. The random variable $y(t)$ is the observation at time t, $y(t) \in \{ y_1, y_2, y_3, y_4 \}$. The arrows in the diagram denote conditional dependencies, often known as trellis diagram.

W. Ghai and N. Singh [19] have worked for Punjabi based Automatic Speech Recognition, in their paper they have described a Recognizer that recognizes isolated and connected speech using HMM toolkit HTK 3.4.1 speech engine. For isolated word recognition the system has shown 92.05% accuracy where as for connected speech system has gained 97.14% accuracy.

B. *NEURAL NETWORK BASED APPROACH:*

This is one of the modest approaches that have been incorporated into the field of speech recognition. All the working of this approach copies the working of a neuron in human brain and has occupied place in computer science, electrical engineering, mathematics, physics, psychology, and linguistics as well. Using this approach the model can attain or acquire some of the following properties, like-

1) Networks under this approach often provide u uniform computational paradigm that can integrate the constraints from various inputs in an easy manner.
2) Speech signals are often noisy, but Networks are tolerant of all kind of damaged and noisy data.
3) Networks are taught in such a way to form association between input and output patterns. By this Trainability feature, network becomes capable of performing classification itself.
4) These kinds of models are able to process the data or speech at high speed because Networks are parallel in nature and parallelism is beneficial for massive parallel computers [17]

Numerous Networks are available under Neural Networks, with different workings with different architectures and different applications as well but all the networks share common processing elements like neurons or nodes along with weighted inputs, upon which different functions are applied to yield desired output. Some extra bias values are also applied to set of values when passed to hidden layer. Network also comprises of training set, a set that comprises of inputs and output units.

Joe Tebelskis in his Thesis Report evaluated the MS-TDNN on the Conference Registration database and Resource Management database. These two sets of experiments were performed under different experimental conditions, with databases of different sizes. The device resulted with Conference Registration database, word accuracy was 72% after frame level training and 81% after word level training and on the Resource Management database.word accuracy was 89.2% after frame level training, and 90.5% after word level training. He has succeeded in showing

MODEL/ APPROACH	RECOGNITION ACCURACY
HMM	ISOLATED-92.05% CONNECTED- 97.14%
NEURAL NETWORK	CONFERENCE REGISTRATION DATABASE - 72% (Frame level) 81%(Word level) RESOURCE MANAGEMENT DATABASE- 89.2%(Frame level) 90.5% (Word level)
SYLLABLE-BASED	TAMIL-48.7% TELUGU- 45.36%
MODGDF and MFCC	CONTINUOUS- 86-87%.

Table I. Resulting Table of different Approaches

that Neural Network can be used for recognizing the speech if it is used carefully and wisely [17]

C. ASR WITH SYLLABLE BASED APPROACH

This is one of the approaches that can work well for continuous speech, where speech is first segmented into syllable like units automatically, and then similar kind of syllables are then grouped together by using some appropriate techniques.

A SYLLABLE is a well organization of speech sounds, arranged in particular sequence. For example if we talk about Punjabi language, there are seven types of syllables. They are- V, VC, CV, VCC, CVC, CCVC and CVCC; where V and C represent vowel and consonant respectively [16]. SEGMENTATION is another term that is associated with this approach, where speech signal is decomposed into set of various phonetic units (phoneme or a syllable). For the execution of this phase, there exist so many algorithms. Researchers have tried variety of algorithms to attain good performance or high accuracy. *Prasad (2002*) has demonstrated that Manual segmentation at syllable-like units followed by isolated style recognition of continuous. Speech results in 65% recognition of syllable-like units.

G Lakshmi Sarada, A Lakshmi, Hema A Murthy and T Nagarajan [4] in their paper, have worked for ASR by using syllable like units methods, in which for the execution of segmentation phase, they opted sub-band and group delay based segmentation algorithm. Here, the thing that was done at very first was the removal of long silences. The modified speech is then forwarded via three filters, they are- all-pass, low-pass and band-pass. With this segmentation, accuracy that was achieved is 80%.

LABELLING is another phase of the approach, where acoustic models for the speech segments are developed by speech recognizer. Grouping is done here for the similar sounds to yield unique models for particular sound unit.

Using syllables based method *G Lakshmi Sarada, A Lakshmi, Hema A Murthy and T Nagarajan [4]* have gained 48.7% accuracy for tamil language and 45.36% accuracy for Telugu language.

D. HYBRID TECHNIQUE- MODIFIED GROUP DELAY FUNCTION AND MFCC

The Group Delay Function [12] defined as the negative derivative of phase, can be effectively used to extract various system parameters when the signal under consideration is a minimum phase signal.

$$X(z) = \sum x(n) z^{-n} \qquad (1)$$

Interpretation- here Xi(z) can either be a first order polynomial or second order polynomial.

Modified Group Delay Function reduces the effects of zeroes close to the unit circle, being used [5]. This spectrum is transformed to sequence of Cepstral coefficients. The thing that happens here is, initial phoneme recognizer is built by using features that deduced from those ceptras, after that results are compared with the features that are deduced from traditional mel frequency Cepstral coefficients, i.e, MFCC and with this it resulted with 39.2% more enhanced performance and good for continuous speech [5].

MFCC (Mel Frequency cepstral coefficients) are the coefficients which are derived from the Cepstral representation for any audio. Such frequency is quite better for audio compression. MFCC is basically deduced by-

1) Calculating the Fourier Transformation,
2) The powers of spectrum which are achieved by above, are mapped to the mel scale,
3) Take the log value for powers in each frequency,
4) Next is the need of taking discrete cosine transformation of mel log powers,

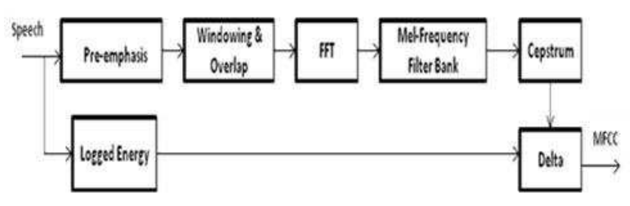


Fig. 3.Working of MFCC [18]

Rajesh M. Hegde, Hema A. Murthy, Gadde V. Ramana Raoin their paper by using this hybrid kind of approach they have concluded with the fact that such combination of Modified group delay function with MFCC improves 18% overall performance and accuracy 86-87%.

VII. RESULTS & DISCUSSION

From the above table, it can be easily calculated that though HMM is a conventional technique in field of speech recognition, still its yielding good results, in terms of accuracy it is showing quite influencing performance. At different parameters, Syllable based and MODGDF are yielding average performance where as NN and HMM are showing high accuracy rates out of which HMM is at high position. Blend of HMM with NN and many more techniques can yield much better performance.

CONCLUSION

The study has been carried out by keeping all the mentioned objectives clearly in front of the vision, among so many algorithms; four have been studied and resulted with the conclusion, though so many heuristic techniques have been arisen for the Recognition of speech still the conventional approach HMM is yielding maximum accuracy, depending upon application depending upon optimal use and application of technique. Combination of HMM with so many algorithms that have mentioned above, can introduce such good results in terms of accuracy which is the major issue when talk about recognition systems.

SCOPE FOR FURTHER STUDY

Among so many existing algorithms, here we have considered only four algorithms under the area of comparison. In future, others can be compared with each other so to introduce and determine the optimal algorithm for recognition of speech.

REFERENCES

[1] Anusuya M. A., katti S. K., "Speech Recognition by Machine: A review", International Journal of Computer Science and Information Security(IJCSIS), Vol. 6, No. 3, 2009).

[2] Davis S. and Mermelstein P., "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences", IEEE Trans. on Acoustics, Speech and Signal Processing, vol. ASSP-28, no. 4, pp. 357–366, 1980.

[3] Davis, K. Biddulph, R. and Balashek, S., "Automatic Recognition of Spoken Digits" Journal of the Acoustical Society of America, vol.24, pp. 637-642, 1952.

[4] G Lakshmi Sarada, A Lakshmi, Hema A Murthy and T Nagarajan, "Automatic transcription of continuous speech into syllable like units for Indian languages", S¯adhan¯a Vol. 34, Part 2, April 2009, pp. 221–233. © Printed in India 16 Prasad 2002.

[5] Hema A Murthy, Venkata Ramana Rao Gadde, "The Modified Group Delay Function and Its Application to Phoneme Recogntion".

[6] Joseph W.Picone.Sep. 1993. "Signal Modeling Technique in Speech Recognition". IEEE Proc.

[7] Mourtaga ehab, sharieh ahmad, Abdallah Mousa, "Speaker Independent Quranic Recognizer Based on Maximum Likelihood Linear Regression", Proceedings Of World Academy Of Science, Engineering And Technology Volume 20 April 2007 Issn 1307-6884.

[8] M.Karanjanadecha and Stephen A.Zahorian.1999, "Signal Modeling for Isolated Word Recognition", ICASS SP.

[9] Netti Chalapathy, rajput Nitendra, Verma Ashish, "A Large Vocabulary Continuous Spech Recognition System For Hindi".

[10] Nilsson Mikael, Ejnarsson Marcus, "Speech Recognition using Hidden Markov Model performance evaluation in noisy environment", Thesis Report, Degree of Masterof Science in electrical Engineering, Department of Telecommunications and Signal processing, Blekinge Institute of Technology, March 2002.

[11] Pylkkonen Janne, "Towards Efficient and robust Automatic Speech Recogniion: Decoding Techniques and Discriminaive Training", Aalto University Publication Series, DOCTORAL DISSERTATIONS 44/2013.

[12] Rajesh M. Hegde, Hema A. Murthy, Gadde V. Ramana Rao, "Continuous Speech Recognition using Joint Features derived from The Modified Group Delay Function and MFCC".

[13] R. K. Aggarwal and Mayank Dave , "Implementing a Speech Recognition System Interface for Indian Languages", IJCNLP-08 Workshop on NLP for Less Privileged, IIIT, Hyderabad, 11 January 2008.

[14] Shannon C. E., "A Mathematical Theory of Communication", Reprinted with corrections from The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, July, October, 1948.

[15] Singh Bhupinder, Singh Parminder, "Voice Based user machine Interface for Punjabi using Hidden Markov Model", IJCST Vol. 2, Issue 3, September 2011, I S S N : 2 2 2 9 - 4 3 3 3 (P r i n t) | I S S N : 0 9 7 6 - 8 4 9 1 (On l i n e).

[16] Singh Prem, "Sidhantik Bhasha Vigeyan", Patiala: Madan Publications, pp. 391.

[17] Tebelskis Joe, "Speech recognition using Neural Networks", PhD Thesis Report, School of Computer Science, Carnegie Mellon University, May 1995.

[18] Wei HAN, Cheong-Fat CHAN, Chiu-Sing CHOY and Kong- Pang PUN, "An Efficient MFCC extraction Method in Speech Recognition", 0-7803-9390-2/06/$20.00 ©2006 IEEE, ISCAS 2006.

[19] W. Ghai and N. Singh, "Phone Based acoustic modeling for automatic speech recognition for Punjabi language", Journal of Speech Sciences 1(3):69-83.2013.

Review on Audio Steganography Techniques

Shelza
Department of Computer Science and Engineering
Guru Nanak Dev University
Amritsar, India
shelzamahajan13@yahoo.in

Palak Sharma
Department of Computer Science and Engineering
Guru Nanak Dev University
Amritsar, India
palaksharma_1991@yahoo.com

***Abstract*— The main objective of this paper is to explore audio steganography and its applications in secure, real-time voice communication. Steganography is the process of hiding a secret message within a larger one such that the presence of the secret message cannot be detected. To explore the benefits and limitations of the audio steganography techniques some well known journals are selected for research papers. The comparison among audio steganography with cryptography is also drawn and also discusses the working of least significant bit (LSB) steganography.**

Keywords: Audio steganography, Cryptography, Data hiding, secure transmission.

I. Introduction

Data hiding (or information hiding) is a part of information security and a secret communication technology. The word steganography or data hiding is of Greek origin and means "concealed writing" from the Greek words "stegano" meaning "covered or protected", and "graphy" meaning "writing". Steganography [1] is the art and science of concealed hiding messages (or data) within data in such a way that no one apart from sender and intended recipient, suspects the existence of message, a form of security through obscurity i.e. Steganography is changing the image in such a way that only sender And intended recipient is able to detect the message sent through it. There are four types of steganography:

i. Steganography in Images: Image steganography [2] is very effective and can serve a variety of purposes included authentication, concealing of messages etc. Least significant bit method is used in which hidden messages will change the last bit of byte in an image. So by doing this, there will be relatively no change within the carrier image.
ii. Steganography in Audio**:** In audio Steganography [2] system, secret messages are embedded into digitized audio signal which results into altering binary sequence of corresponding audio files.
iii. Steganography in Video**:** Steganography in Videos [2] basically deals with hiding of information in each frame of video. This type reveals more information instead of hiding.
iv. Steganography in Text: Encoding secret messages in text [2] can be a very challenging task. This is because text files contain redundant data to replace with a secret message. Another cons is the ease of which text based Steganography can be altered by an unwanted parties by just changing the text itself or reformatting the text to some other form (from .TXT to .PDF, etc.).

A. *Comparison Of Secret Communication Techniques*

TABLE I. Comparison of secret communication techniques [3]

Secret Communication Techniques	**Confidentiality**	**Integrity**	**Unremovability**
Encryption	Yes	No	Yes
Digital Signatures	No	Yes	No
Steganography	Yes /No	Yes /No	Yes

B. *Difference between Steganography and Cryptography*

TABLE II. Difference between steganography and cryptography [4]

Steganography	**Cryptography**
Steganography deals with composing hidden messages so that only the sender and the receiver know that the existence of message	Cryptography is the study of hiding information and is used when communicating over an untrusted medium such as internet, where information needs to be protected from other third parties.
In Steganography, only the sender and the receiver know the existence of the message.	In cryptography the existence of the encrypted message is visible to the world.
Steganography removes the unwanted attention coming to the hidden message	Do not remove unwanted attention

Steganography is concerned with hiding the existence or presence of a message and protect both messages and communicating parties.	Cryptography is concerned with protecting the content of information or message but is not concerned with hiding its existence
Steganography does not alter the structure of secret message	Cryptography alter the structure of secret message
Unknown Message Passing	Known Message Passing
Once detected message is known	Strong current algorithms are currently resistant to attack, larger expensive computing power is required for cracking

II. AUDIO STEGANOGRAPHY

Embedding secret messages into digital sound is known as Audio Steganography [5]. It is usually a more difficult process than embedding messages in other media. Audio Steganography methods can embed messages in WAV, AU, and even MP3 sound files.

A. *Model of Audio Steganography*

The basic model [6] of Audio steganography consists of Carrier (Audio file), Message and Password (stego- key) Carrier is also known as a cover-file, which hides the secret information. The model is demonstrated in Figure 1:

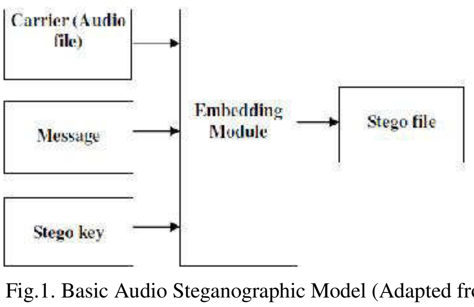


Fig.1. Basic Audio Steganographic Model (Adapted from [6])

Message is the data that the sender wants to remain it private. Message can be plain text, image, audio or any type of file. Password is known as a stego-key, which ensures that only the recipient who knows the corresponding decoding key will be able to extract the message from a cover-file. The cover-file with the secret information is known as a stego-file.

The information hiding process consists of following two steps:

i. Firstly, redundant bits in a cover-file are identified. Redundant bits are those bits that can be altered without destroying the integrity and exploiting the quality of the cover file.
ii. To embed the secret information (or data) in the cover file, the redundant bits present in the cover file is replaced by the bits of the secret information.

B. *Audio steganography Techniques:*

1) Least Significant Bit coding: Least significant bit (LSB) [7] coding is the simple, fast and popular methodology to embed information in a digital audio file. In this technique, LSB of binary sequences of each sample of digitized audio file is replaced with binary equivalent of secret message. LSB coding allows for a large amount of data to be encoded.
2) Phase Coding: Phase coding [7] addresses the disadvantages of the noise-inducing methods of audio Steganography. Phase Coding works by substituting the phase of an initial audio segment with a reference phase that represents the data (or secret information).
3) Parity Coding: Parity coding [7] is one of the robust audio Steganographic techniques. Instead of breaking a signal into individual samples, this method breaks an original signal into separate samples and embeds each bit of the secret message from a parity bit. If the parity bit of a selected region does not match the secret bit to be encoded, the process inverts the LSB of one of the samples in the region. Thus, the sender has more of a choice in encoding the secret bit.
4) Echo data hiding: Echo data hiding [7] deals with embedding of text (or data) in audio file by introducing an echo to the original signal. The data then hidden by varying three parameters of the echo: initial amplitude, decay rate, and offset. If only one echo was produced from the original signal, only one bit of information could be encoded.
5) Spread spectrum: In audio steganography, the basic spread spectrum (SS) [7] method attempts to spread secret information across the frequency spectrum of the audio signal using a code which is independent of actual signal. This is similar to a system which uses an implementation of the LSB that spreads the message bits randomly over the entire sound file.

III. How Least Significant Bit Stegano Works?

Least significant bit (LSB) steganography is the simple, fast and popular methodology to embed information in a digital audio file

The algorithm [5] for LSB modification as shown in

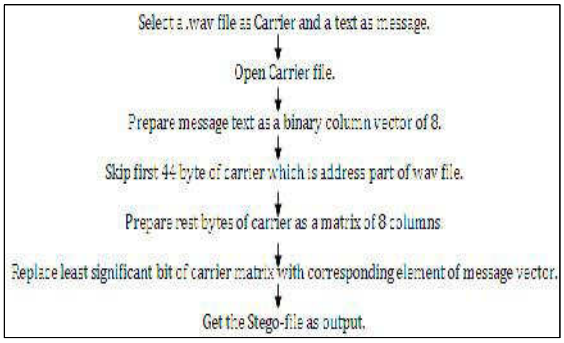


Figure 2:

Fig.2. Flowchart of LSB modification Technique for Audio Steganography (Adapted from [5])

For Example:

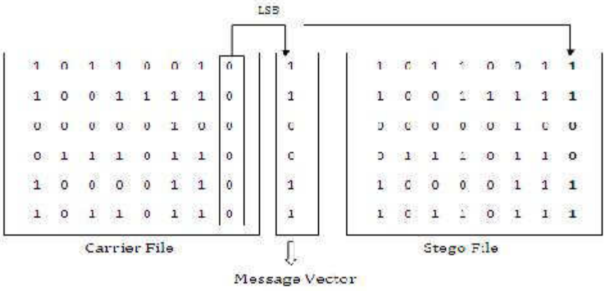


Fig.3. LSB modification procedures for Audio Steganography (Adapted from [5])

The figure 3 shows that there is reasonable no change between input carriers file and output stego-file.

IV. LITERATURE SURVEY

Djebbar et al.(2011) [8] presented a novel and adaptable techniques to enforce data security in audio steganographic system and also addresses issues related to the need to secure and preserve the integrity of data concealed in voice communications, even when the latter passes through insecure channels.

Shahadi and Jidin [9] have proposed a high capacity and high quality audio steganography algorithm The aim of this algorithm to gain a large embedding capacity and high output quality for different data types and to gain advantages of three things: Discrete wavelet packet transform, bits block matching between message and cover signals, and adaptive hiding in LSB depend on the strength of the cover samples.

Kumar and Anuradha[10] have gives a method of embedding text data (or message) in an audio file using LSB and restored correctly at the destination (or receiver) end. The WAV files may be used for designing has been done in an audio file. The WAV files have basically two components: header and data. The embedding of data should be done in data not in header because header part can create corruption in an audio file.

Gopalan and Shi [11] have given a simple technique for embedding multimedia data for large payload in audio steganography using bit modification method. In each sample of cover audio, modification of bit can increase the payload at the cost. Depending on the index of bit used to modify the sample in accordance with hidden data, the resulting stego audio signal may become detectable to incorrect retrieval of hidden data.

Qiao et al. [12] has proposed a quantitative analysis method to determine the embedding strength of MP3 steganography by extracting features from transform domain. This method is beneficial to estimating the length of secret message and locating embedding position.

Asad et al. [13] has proposed two ways to extend the conventional LSB modification technique for audio steganography to make it more secure against steganalysis and decrease the probability of secret message being extracted by attacker. The first way is to randomize bit number of host message used for embedding secret message and the second method randomize sample number containing next secret message bit.

Balgurgi and Jagtap [7] provides implementation of two level encryption of user data by combining two areas of network security i.e. steganography and cryptography. The authors also describe combination of LSB technique with XORing method increase the level of security and improve the confidentiality of audio steganography.

Basu and Bhowmik [14] have presented a method of embedding text information in an audio file that does not affect the file structure and contents of audio file. In the proposed method, first the sampling is done in an audio file and then changed the appropriate bit of each sample to embed textual information.

Sakthisudhan et al. [15] has given a method that is lesser prone to error while transmission. In the proposed method, secret message in form of audio file is embedded within another carrier audio file (.wav) .In the transmitter end, the embedding process is carried out with the carrier.

V. GAPS IN LITERATURE

By conducting the survey we have found that the most of existing researchers has neglected the issue of the audio file size. The techniques which have not used the header data proves to be inefficient when audio get little bit corrupted during transmission time. Also long audio file length increases the overall computation time. The emphases of this work are to propose a technique which put minimum effect on audio signals.

VI. CONCLUSION

In the condition of steganalysis, classification of audio documents as a bearing hidden information is not a security issue. A cover audio object can be converted into a stego-audio object via steganographic methods. In this work we have reviewed a statistical methods will be proposed which will calculate the effect of steganography on audio signals. Kumar and Anuradha (2012) [10] has proposed a technique which does not use the data headers to embed the information. The limitation of this vary approach is that system will analyses the entire wave file to find the messages; which becomes time consuming if the size of wav file is quite large. To reduce this problem in near future we will use headers to embed the message in wav files. Also we will use audio compression techniques to speed up the system. As every steganography technique comes up with some overheads and also results in increasing the size of audio signals so overheads and optimality is also considered in this thesis. Different metrics will be calculated which will be used to compare proposed optimal technique with available methods.

REFERENCES

[I] Nosrati, Masoud, Ronak Karimi, and Mehdi Hariri, " Audio Steganography: A Survey of Recent Approaches", World Applied Programming. 2012,pp: 202-205.

[II] Nosrati, Masoud, Ronak Karimi, and Mehdi Hariri (2011). An introduction to steganography methods. World Applied Programming,2011,pp: 191-195.

[III] Channalli, Shashikala, and Ajay Jadhav. "Steganography an art of hiding data." International Journal on Computer Science and Engineering. 1.3 ,2009, pp. 137-141.

[IV] Kumar, Samir, and Gupta Banik Bandyopadhyay Barnali , " LSB Modification and Phase Encoding Technique of Audio Steganography Revisited", International Journal of Advanced Research in Computer and Communication Engineering. 2012,pp: 1-4.

[V] Jayaram, P., H. R. Ranganatha, and H. S. Anupama , "Information Hiding Using Audio Steganography–A Survey. International Journal of Multimedia & Its Application"s (IJMA), 2011,pp: 86-96.

[VI] Balgurgi, Pooja P., and Sonal K. Jagtap , " Intelligent processing: An approach of audio steganography",International Conference on Communication, Information & Computing Technology (ICCICT) in India, Mumbai, India,2012,pp:23-27.

[VII] Djebbar, Fatiha,, "A view on latest audio steganography techniques", International Conference on Innovations in Information Technology 2011,pp:409-414.

[VIII] Shahadi, Haider Ismael, and Razali Jidin , " High capacity and inaudibility audio steganography scheme",7th International Conference on Information Assurance and Security (IAS),2011,pp: 104-106.

[IX] Kumar, Harish, "Enhanced LSB technique for audio steganography" , Third International Conference on Computing Communication & Networking Technologies (ICCCNT) in India,2011,pp:109-112..

[X] Qiao, Mengyu, Andrew H. Sung, and Qing hong Liu , "Predicting embedding strength in audio steganography",9th IEEE International Conference on Cognitive Informatics (ICCI),2010,pp:. 925-930.

[XI] Asad, Muhammad, Junaid Gilani, and Adnan Khalid (2011). An enhanced least significant bit modification technique for audio steganography. Paper presented at the 2011 International Conference on Computer Networks and Information Technology (ICCNIT),2011,pp:143-147.

[XII] Basu, Pramatha Nath, and Tanmay Bhowmik , " On Embedding of Text in Audio A Case of Steganography." , International Conference on Recent Trends in Information, Telecommunication and Computing (ITC),2010,pp: 203-206.

[XIII] Sakthisudhan, K., P. Prabhu, and P. Thangara , " Secure Audio Steganography for Hiding Secret information",International Conference on Recent trends in Computational Methods, Communication and Controls (ICON3C),2012,pp: 33-37.

An Empirical on Image Information Retrieval Approaches

Harpreet Kaur
Sant Baba Bhag Singh Institute of Engg. & Technology, Padhiana, Distt. Jalandhar, India

Mandeep Saini
Sant Baba Bhag Singh Institute of Engg. & Technology,
Padhiana, Distt. Jalandhar, India

Mandeep Minhas
Sant Baba Bhag Singh Institute of Engg. & Technology, Padhiana, Distt. Jalandhar, India.

***Abstract*--Image databases are becoming increasingly popular due to large amount of images those are generated by various applications and the advances in computation power, storage devices, scanning, networking, image compression, and desktop publishing. This paper provides an overview of current research in image information retrieval from image databases with two aspects of image retrieval (IR) as Text-based Retrieval and Content-based Retrieval and also make the comparison of these two approaches.**

***Keywords*--Image Retrieval, CBIR, query, Texture, Shape, Similarity.**

I. INTRODUCTION

Advances in data storage and image acquisition technologies have enabled the creation of large image datasets. In order to deal with these data, it is necessary to develop appropriate information systems to efficiently manage these collections. Image searching is one of the most important services that need to be supported by such systems. In general, two different approaches have been applied to allow searching on image collections: one based on image textual metadata and another based on image content information. [1] If the query is text-based, then the problem involves how to convert the textual query into something that can be compared with images. On the other hand, if the query is image-based, there is a different problem involving the user interface. [2].

Fig. 1 shows the architecture of image retrieval. The system has user interface, where the user enters the text and/ or sample image as a query. The textual and visual content descriptors are generated from the text query and image query. The descriptors are converted into a vector format. Similarly textual and visual descriptors are calculated and converted into vector representation for the images stored in the database. The vector, generated by the user query is then matched with the vectors stored in the database. The text and content-based methods return two independent lists of images with different weights. These two lists must be combined in a meaningful way to give the user a combined image list. [3]

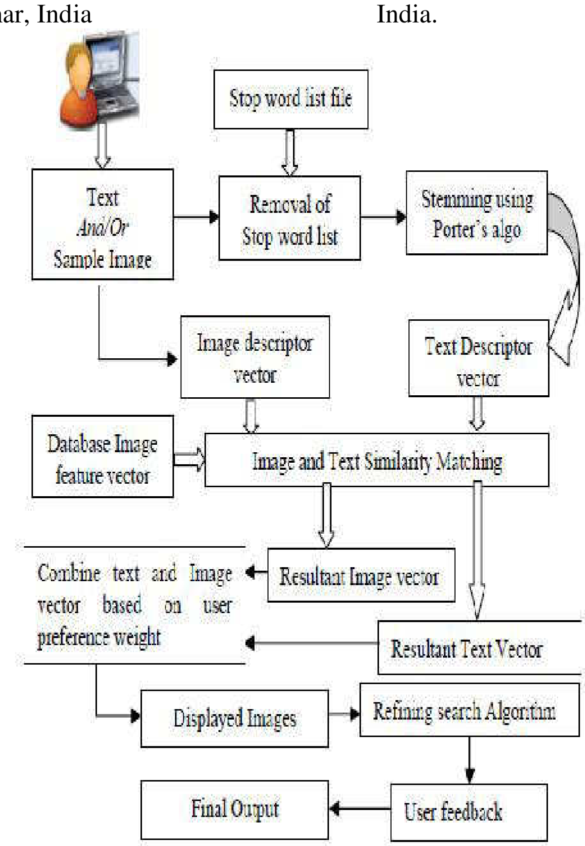


Fig. 1. Architecture of Image Retrieval [3]

II. TEXT BASED IMAGE RETRIEVAL

Text-based image retrieval uses traditional database techniques to manage images. Through text descriptions, images can be organized by topical or semantic hierarchies to facilitate easy navigation and browsing based on standard Boolean queries. Although text-based methods are fast and reliable when images are well annotated, they are incapable of searching in unannotated image collections. The generalization of the information retrieval from the text domain to the image database is, however, non-trivial. [3].

Fig. 2.Retrieved images for user query "Bike

A). *Techniques used in text retrieval:*

1). Bag of words approach

It is a simplifying representation used in information retrieval. In this a text is represented as a bag (multiset) of its words, disregarding grammar and even words order but keep multiplicity.

2). Stop words can be removed

It means very frequent words contain little information and can be removed as automatically in Google.

3). *Stemming can improve results*

It is strongly dependent on language .basically; it is the processes for reducing infected words to their stem, base or root form-generally a written word form.

4). *Spelling correction*

Google had a big success with this to correct spellings automatically.

III. CONTENT BASED IMAGE RETRIEVAL

Research on content-based image retrieval has gained tremendous momentum during the last decade and was introduced in the early 1980s. A lot of research work has been carried out on Image Retrieval by many researchers. CBIR is the most important and effective image retrieval method and widely studied in both academia and industry arena. In content based image retrieval ,users search the multimedia repository providing information about the actual contents of the image, audio, or video clip."Content-based" means that the search will analyze the actual contents of the image rather than the meta- data such as keywords, tags, and/or descriptions associated with the image. The term 'content' in this context might refer to colors, shapes, textures, or any other information that can be derived from the Image itself.

A) Color-based retrieval:

It is one of the important features of content based image retrieval. This color feature has widely been used in CBIR systems, because of its easy and fast Computation.[5] Computing distance measures based on color similarity is achieved by computing a color histogram for each image that identifies the proportion of pixels within an image holding specific values.[6]

Color is also an intuitive feature and plays an important role in image matching. The extraction of color features from digital images depends on an understanding of the theory of color and the representation of color in digital images. The color histogram is one of the most commonly used color feature representation in image retrieval.

1). Advantages:

- Speediness.
- Low demand of memory space.
- Not sensitive in changes of image size and rotation.
- Easy and fast computation. [7]

B) Texture-based retrieval:

The texture feature is another type of important and useful visual information for image retrieval. "Texture is an attribute representing the spatial arrangement of the grey levels of the pixels in a region or image". There exist different approaches to extract and represent textures. They can be classified into space-based, frequency-based models, and texture signatures. Some popular techniques i.e. Wavelet transform, co-occurrence matrix, and Gabor filters are applied to express texture features for image. [5] We usually adopt texture's statistic feature and structure feature as well as the features that based on spatial domain are changed into frequency domain. [7]

1). Advantages

- It is a powerful low-level feature for image search and retrieval applications.
- It provides computationally inexpensive properties.

C) Shape-based retrieval

Shape does not refer to the shape of an image but to the shape of a particular region that is being to be implemented.

Shape descriptors are classified into *boundary- based (or contour-based)* and *region-based* methods. This classification takes into account whether shape features are extracted from the contour only or from the whole shape region. There is three problems need to be solved during the image retrieval that based on shape feature. Firstly, shape usually related to the specifically object in the image, so shape's semantic feature is stronger than texture. [1]

1). Advantages

- Stronger.
- Helpful in pattern recognition.

IV. IMPLEMENTATION OF CONTENT BASED IMAGE RETRIEVAL

In this, the following two approaches are being used:

- System flowchart of CBIR
- Architecture of CBIR

A). *System flowchart of CBIR*

The proposed system will produce the output as images which are relevant to the query Image. The system flowchart is given below:

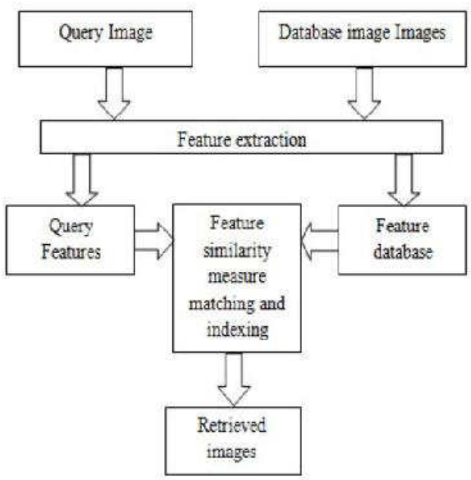


Fig. 3 System flowchart [8]

The system has the following structure-:

1*). Collection of Database*:

A database containing number of images with any one of this formats .bmp, .jpg, .tiff. is required.

2). *Query:*

The user provides a sample image or sketched figure as the query for the system.

3). *Feature Extraction:*

There are various kinds of low-level visual features to represent an image, such as color, texture, shape, and spatial relationship. Since one type of features can only represent part of the image properties, a lot of work done on the combination of these features.

Detailed about all the features are given below:

a). Color:

Color feature is one of the most widely used features in image retrieval. Colors are defined on a selected color space. Variety of color spaces include, RGB, LAB, LUV, HSV (HSL), YCrCb and the huemin- max-difference (HMMD).In this we also have the Color Structure Descriptor (CSD) which represents an image by both the color distribution of the image or image region (similar to a color histogram) and the local spatial structure of the color. The extra spatial information makes the descriptor sensitive to certain image features to which an ordinary color histogram is blind. CSD used an 8 × 8 structure to scan the total image. This descriptor counts the number of times a particular color is contained within the structuring element while the image or image region is scanned by this structuring element. It has used HMMD color space. Color Structure Descriptor Block Diagram is shown as follows:

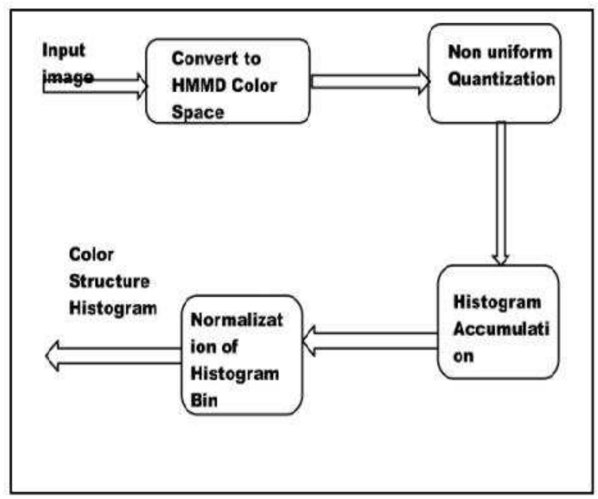


Fig.4 Block diagram of color Structure descriptor[9]

The input image of different color space is converted to HMMD Color space. HMMD color space is supported in MPEG-7. The hue has the same meaning as in the HSV space, and max and min are the maximum and minimum among the R, G, and B values, respectively. The different component is defined as the difference between max and min. The three of the four components are sufficient to describe the HMMD space (Hue, Min, Max) or (Hue, Difference, Sum). This color space is represented by double cone structure. The HMMD color space is very effective and compared favorably with the HSV color space.

b) *Texture:*

The texture feature is another type of important and useful visual information for image retrieval.. There exist different approaches to extract and represent textures. They can be classified into space- based, frequency-based models, and texture signatures. Some popular techniques i.e. wavelet transform, co-occurrence matrix, and Gabor filters are applied to express texture features for image.

c) *Shape:*

It is the most obvious requirement at the primitive level. It is seen that natural objects are primarily recognized by their shape. Two main types of shape feature are commonly used– global features such as aspect ratio, circularity and moment invariants and local features such as sets of consecutive boundary segments.

4). *Similarity Matching:*

This involves matching these features to yield a result that is visually similar. The commonly used similarity measure method is the Distance method. There are different distances available such as Euclidean distance, City Block Distance, Canberra Distance.

5). *Retrieval:*

The System retrieves and presents a sequence of images ranked in decreasing order of similarity or with the minimum distances is returned to the user. [9]

B Architecture of CBIR:

Following Figure shows a typical architecture of a content-based image retrieval system.

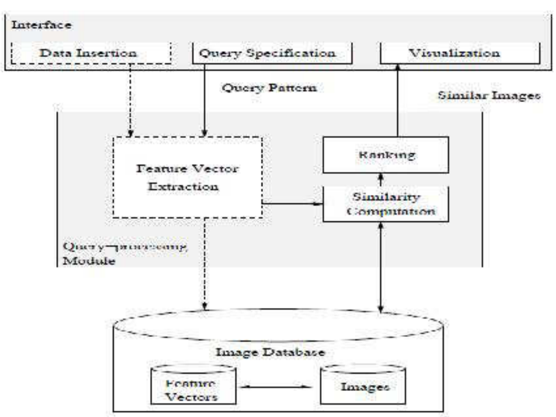


Fig.5. Architecture of CBIR [10]

1). Two main process of interface

a). Data insertion and query processing.

The data insertion subsystem is responsible for extracting appropriate features from images and storing them into the image database (see dashed modules and arrows). This process is usually performed off-line.The query processing, in turn, is organized as follows: the interface allows a user to specify a query by means of a query pattern and to visualize the retrieved similar images. The query-processing module extracts a feature vector from a query pattern and applies a metric (such as the Euclidean distance) to evaluate the similarity between the query image and the database images. Next, it ranks the database images in a decreasing order of similarity to the query image and forwards the most similar images to the interface module. Note that both the data insertion and the query processing functionalities use the feature vector extraction module.

V. COMPARISON BEWTEEN IMAGE RETRIEVAL

A). Approches

FEATURES	**TEXT-BASED IMAGE RETRIVAL**	**CONTENT-BASED IMAGE RETRIEVAL**
Introduction	It is based on the textual annotation of image.	It is based on the visual content of the image which depends upon the color, texture and shape of that image
Techniques used	It requires considerable level of human labor for manual annotation.	It uses extraction and similarity matching techniques.
Support	Difficult to support a variety of task dependent queries.	Easily supportable for all tasks.
Performance	Lack of systematization in the annotation process decrease the performance.	Performance increases due to the possibility of an automatic
Cost	expensive for large scale image database.	less expensive.
Efficiency	very low	High
Complexity	easy and simple	less complex
Database used	not used for large database	it use large database to store the images
Time	perform time consuming tasks	Less consumption of time
Accuracy	Manual annotation is not accurate	highly accurate process
Reliability	less reliable	perform reliable tasks

VI. DRAWBACKS OF TEXT BASED IMAGE RETRIEVAL

Early image retrieval was similar to text retrieval and was realized by remarks and notes to image, so all images should be remarked firstly then retrieval can be done through matching remark text .This method has some deficiencies such as subjectivity and labor-intensive nature of the keywords assigning process. Moreover, it is sometimes almost impossible to describe content of an image by words, especially for textures found in an image. Here are some more drawbacks of text-based image retrieval which in turn give rise to CBIR in practice:

- The work of image remarking is very large.
- It is difficult to select effective remark text that can represent abundant information in the image.
- Expensive tasks for large image database, and is often subjective, context-sensitive and incomplete.
- Different people may supply different textual annotations for the same image. This makes it extremely difficult to answer user queries reliably.
- The retrieval efficiency is very low.

VII. APPLICATIONS OF CBIR

A wide range of possible applications for CBIR technology has been identified (e.g. Gudivada and Raghavan [1995]). Potentially fruitful areas include:

- The military
- Architectural and engineering design
- Medical diagnosis
- Education and training [4]

A). The Military

Military applications of imaging technology are probably the best-developed, though least publicized. Recognition of enemy aircraft from radar screens, identification of targets from satellite photographs, and provision of guidance systems for cruise missiles are known examples though these almost certainly represent only the tip of the iceberg. Many of the surveillance techniques used in crime prevention could also be relevant to the military field. [4]

B). Architectural and Engineering Design

Architectural and engineering design share a number of common features – the use of stylized 2- and 3-D models to represent design objects, the need to visualize designs for the benefit of non-technical clients, and he need to work within externally-imposed constraints, often financial. Such constraints mean that the designer needs to be aware of previous designs, particularly if these can be adapted to the problem at hand. Hence the ability to search design archives for previous examples which are in some way similar, or meet specified suitability criteria, can be valuable. [4]

C). Medical Diagnosis

The increasing reliance of modern medicine on diagnostic techniques such as radiology, histopathology, and computerized tomography has resulted in an explosion in the number and importance of medical images now stored by most hospitals. While the prime requirement for medical imaging systems is to be able to display images relating to a named patient, there is increasing interest in the use of CBIR techniques to aid diagnosis by identifying similar past cases. Most development work in the PACS (picture archiving and communication systems) area is still directed towards providing basic functionality (ensuring that medical images can be successfully digitized, stored and transmitted over local area networks without loss of quality) and usability (providing user-centred interfaces and integrating image storage and retrieval with wider aspects of patient record management). However, experimental content-based retrieval systems are beginning to have some impact. Examples of this include the I^2C system for retrieving 2-D radiological images from the University of Crete [Orphanoudakis et al, 1994], and the 3-D neurological image retrieval system currently being developed at Carnegie- Mellon University [Liu et al, 1998], both developed with the aim of assisting medical staff in diagnosing brain tumors. [4]

D Education and Training

It is often difficult to identify good teaching material to illustrate key points in a lecture or self-study module. The availability of searchable collections of video clips providing examples of (say) avalanches for a lecture on mountain safety, or traffic congestion for a course on urban planning, could reduce preparation time and lead to improved teaching quality. In some cases (complex diagnostic and repair procedures) such videos might even replace a human tutor. [4]

VIII. CONCLUSION:

In this paper, we have discussed the two approaches that are used in image information retrieval from large collection of images called image databases. With the comparison between these approaches, response of CBIR is much better and fast as compared to Text Based Image retrieval having a lot of drawbacks.

REFERENCES

[1] Ricardo da Silva Torres, Alexandre Xavier Falcão, "Content-Based Image Retrieval: Theory and Applications-:" Vol. 9 , Nov 2006.

[2] Greg Bickerman, Joshua Ehrlich, Joshua Ehrlich, "Text-Based Image Retrieval Using Histograms-:", May 2010.

[3] B. Dinakaran1, J. Annapurna2, Ch. Aswani Kumar, "Interactive Image Retrieval Using Text and Image Content-:", Vol. 10, No 3, 2010.

[4] Eakins, John, Graham Margaret "Content Based Image Retrieval" University of Northumbria at Newcastle. October 1999

[5] Manimala Singha and K.Hemachandran., "Signal and Image Processing:An Content Based Image Retrieval using Color and Texture", International Journal (SIPIJ)Vol.3,No.1,Feb2012.

[6] Nidhi Singhai, Prof. Shishir K. Shandilya, "A Survey On: Content Based Image Retrieval Systems:" International journal of computer applications (09758887) Vol.4, No.2, July 2010.

[7] Swapnalini Pattanaik1, Prof. D.G.Bhalke2, "Beginners to Content Based Image Retrieval:" International journal of scientific research engineering and technology, Vol.1, issue2, pp 40-44, May 2012.

[8] Swapnalini Pattanaik, D.G. Bhalke, "Efficient Content based Image Retrieval System using Mpeg-7 Features:" International journal of computer applications (09758887) Vol.53, No.5, Sept. 2012.

A Reveiw of Gurumukhi Handwritten Text Segmentation for Broken and Touching Characters

Parika Mangla
Baba Farid College of Engineering, Bathinda, India
parikamangla27@gmail.com

Harleen Kaur
Baba Farid College of Engineering, Bathinda, India
Harleengrover@gmail.com

***Abstract*—OCR is a process of automatic computer recognition of characters in optically scanned and digitized pages of text. Segmentation of words into characters is an important challenge in OCR. This is even more challenging when we segment characters in an offline handwritten document. Touching and broken Character make this problem more complex. In this paper, we have reviewed character segmentation techniques that can segment simple characters, touching characters, broken characters. Also we reviewed the challenges effecting character segmentation and if these problems are not removed then it will create problem in recognition phase** .

Keywords— OCR,Segmentation, Character segmentation, broken character segmentation, touching character segmentation,Gurumukhi script.

I. INTRODUCTION

OCR is an acronym for Optical Character Recognition. OCR is the process of converting the scanned images of machine printed or handwritten text, symbols, numerals, letters in to the format which can be processed by the computer such as ASCII [2]. OCR is the most important part of Electronic Document Analysis systems. An OCR has variety of commercial and physical applications. It can be used for automatic reading and processing of the forms, old degraded documents, bank cheques , postal codes, forms [3]. It can prove as an aid for visually handicapped persons.

A. Steps in OCR

- Scan Input Text (Create an Image)
- Binarize the input image(Convert in 0's and 1's form)
- Extract lines from the paragraph
- Extract words from the lines
- Extract Characters from the words
- Extract features of the characters
- Identify the character (Recognition of Character)

B. Segmentation

Segmentation is an important task of any OCR system. The accuracy of OCR system mainly depends on the segmentation algorithm being used. In OCR a perfect segmentation is required before individual characters are recognized. Segmentation of handwritten text in Gurumukhi script is an uphill task primarily because of the structural features of the script and varied writing styles. Segmentation is a technique which partitions handwritten words into individual characters [2]. Segmentation include following steps: Scanned image, Noise removal, binarization, edge detection , dilation and filling and processed image for feature extraction.

C. Character Segmentation

Character Segmentation divides the words into individual characters. It's an operation that seeks to decompose an image of sequence of characters into sub images of individual symbols. A poor segmentation process produces misrecognition or rejection segmentation process carried after out only the pre processing of image [3]. After segmentation of the character features can be extracted from the segmented character. Let us see one example of character segmentation [7].

Fig.1.Word

II. CHALLENGES IN SEGMENTING CHARCER SEGMENTATION

Although segmentation is a straightforward task when dealing with typewritten or well-written characters, it can be quite difficult when the words are handwritten. The main reason for this difficulty is that the handwritten characters often overlap, touched , broken and in some cases, may be skewed. Also, the wide variations in handwriting styles make it very difficult to make generalizations for making segmentation heuristics. But for a symbol in Gurumukhi script, the same pixel values in horizontal direction might be shared by two or more characters/symbols of same word. This adds to the complication of segmentation problem in Gurumukhi script.

- Problem of broken characters
- Problem of overlapped characters
- Problem of touching characters
- Problem of Skewed characters

A. Problem of broken character

Missing character is another problem with the handwritten Gurumukhi script because during writing, some portion of the character may be missing. It is also difficult to segment the character because when we start tracing the path some portion may be missing so this leads to the error in the segmentation.

So, the missing characters result in over-segmentation. Due to presence of broken characters the performance of any OCR may further decrease.

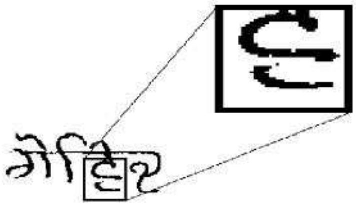


Fig.2.Word [2]

III. *PROBLEM OF TOUCHING CHARACTER*

This problem also arises due to different writing styles adopted by different people. While writing, if one character touches the other character then it becomes difficult to recognize both characters [1]. The problem in touching characters is that the characters are also attached in a zone other than the base line, so the vertical histogram of the word shows no white space between characters through which we can separate them.

Fig.3. Word [6]

IV. *PROBLEM OF OVERLAPPED CHARACTER*

This problem arises due to different writing styles of different people. In this problem one character is written above on the other characters mistakenly [8]. This is termed as overlapping of characters.

Fig.3. Word [6]

V. *PROBLEM OF SKEWED CHARACTER*

In this problem words in a line are not written into straight horizontal line but the word inclined either upward or downward which causes difficulty to detect the header line for that word.

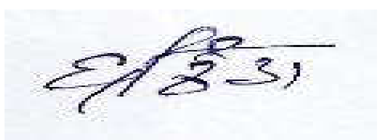


Fig.4. Word

VI. EXITING APPROACHS

In the literature, for achieving high segmentation accuracy, several segmentation techniques are proposed that can be broadly classified into three categories, namely Explicit Segmentation (Pure Segmentation), Implicit Segmentation (Recognition Based Segmentation) and Holistic (Segmentation Free) Approaches.

- Classical approach - It is an explicit segmentation. It consists of methods that divide the input image into sub images, which are then classified. The process of cutting up the word images into classifiable character sub images is termed as dissection. Many researchers in the literature adopted these techniques like (Saba et al., 2011). These techniques are used to find all the interconnections between character images (also called ligatures) and cut the word image through all the detected ligatures.
- Recognition based segmentation- In recognition based segmentation; a search is made for image components that match with the character classes in the alphabet. These are employed to integrate segmentation and recognition processes. Technique used here is HMM in word recognition. The main interest of this category of methods is that they bypass the segmentation problem: No complex "dissection" algorithm has to be built and recognition errors are basically due to failures in classification.
- Holistic approach - In the holistic approach, we attempt to recognize the word as a whole and thus avoid segmenting this into characters. A holistic (Segmentation Free) process recognizes an entire word as a unit. A major drawback of this class of methods is that their use is usually restricted to a predefined lexicon. Since they do not deal directly with letters but only with words, recognition is necessarily constrained to a specific lexicon of words.
- Hybrid approach - Hybrid approach employs dissection together with recombination rules to define potential segments [7]. In this approach, there is a continuous space of segmentation strategies rather than a discrete set of classes with well-defined boundaries.

A. *Main strategies used in character segmentation*

- Horizontal Projection (HP): For a given binary image of size X × Y, where X is the width and Y is the height of the image, the horizontal projection is defined as:

  HP(i), i = 1, 2, 3, …, Y. where HP(i) is the total number of black pixels in ith row.

- Vertical Projection (VP): For a given binary image of size X × Y, where X is the width and Y is the height of the image, the vertical projection is defined as:

VP(j), j = 1, 2, 3, …, X. where VP(j) is the total number of black pixels in jth column.

- Headline Rectangle (HR): To overcome the problem of uneven headline, the row having maximum horizontal profile is displaced with a threshold value of 6 rows above and 6 rows below it. This rectangle is taken as headline rectangle. Its height will always be 13
- Water reservoir: The water reservoir principle is as follows. If water is poured from top and bottom of the word, the cavity regions of the numerals where water will be stored are considered as reservoirs. Here by top (bottom) reservoirs we mean the reservoirs obtained when water is poured from top (bottom). (Here, water pouring from bottom we mean the water pouring from top after rotating the component by 180°).We analyze these top and bottom reservoirs for the segmentation.

VII. DISSCUSSION AND CONCLUDING REMARKS

Here in this paper we have reviewed different challenges of segmenting a character implemented through different techniques. The above problems are created in only handwritten documents not in printed documents. The problems are created due to different handwriting so we are concluding our techniques by comparing them to remove such problems. In order To detect and segment characters we will use pre-segmentation techniques by comparing our existing techniques and developed an efficient algorithm to remove broken and touching character problems .We will try to improve procedure to touching character segmentation and from the results we will conclude further recognition phase.

REFRENCES

[1] Munish kumar , Mk jindal , R.K.Sharma "segmentation of Isolated And Touching Characters in Offline Handwritten Gurumukhi Script Recognition." in IJ Information technology and computer science , 2014.

[2] Nabin Sharma, Palaiahnakote Shivakumara, Umapada Pal, Michael Blumenstein And Chew Lim Tan ,"A New Method For Character Segmentation From Multi-Oriented Video Words" In 2013 IEEE.

[3] Ashwin S Ramteke, Milind E Rane,"Offline Handwritten Devanagari Script Segmentation" in 2012

[4] Sandeep N.Kamble, Prof. Megha Kamble, " Morphological Approach for Segmentation of Scanned Handwritten Devnagari Text" in Oct-Dec, 2011, vol. 2

[5] Naresh Kumar Garg, Lakhwinder Kaur & M.K. Jindal ,"The Hazards in Segmentation of Handwritten Hindi Text" International Journal of computer Applications(0975-8887) in Sep 2011, vol 29- No.2

[6] Galaxy Bansal ,Daramveer Sharma ,"isolated handwritten words segmentation techniques in gurumukhi script" 2010International Conference in computer sceince and its applications .

[7] Likforman-Sulem, L., Zahour, A. and Taconet, B., "Text line Segmentation of Historical Documents: a Survey", International Journal on Document Analysis and Recognition, Springer, Vol. 9, Issue 2, pp.123-138, 2007.

[8] Akihiro Nomura,Kazuyuki Michishita, Seiichi Uchida, and Masakazu Suzuki," Detection and Segmentation of Touching Characters in Mathematical Expressions". Seventh International Conference on Document Analysis and Recognition 2003 IEEE.

[9] Naresh Kumar Garg , Lakhwinder Kaur and M.K. Jindal "Segmentation of Handwritten Hindi Text" International Journal of computer Applications(0975-8887) in 2010 , vol. 1-No. 4

[10] Vijay Kumar, Pankaj K. Sengar," Segmentation Of Printed Text In Devanagari Script And Gurmukhi Script" International Journal Of Computer Applications (0975 – 8887) Volume 3 – No.8, June 2010.

[11] K. Wong, R. Casey and F. Wahl "Document Analysis System ", IBM j.Res . Dev., 26(6), pp.647-656, 1982.

[12] F. Hones and J. Litcher, "Layout extraction of mixed mode documents", Machine Vision Application, vol. 7, pp. 237–246, 1994.

[13] K. Kise, W. Iwata, and K. Matsumoto, "A computational geometric approach to text line extraction from binary document images", in Proc. IAPR Workshop Document Analysis Systems, pp. 364-375, 1998.

Document Image Binarization Techniques-A Review

Ritu
Department Of Computer Science And Engineering
Guru Nanak Dev University
Amritsar, India
Reet.gangar@gmail.com

Kamaljit Kaur
Department Of Computer Science And Engineering
Guru Nanak Dev University
Amritsar, India
Kamal.aujla86@gmail.com

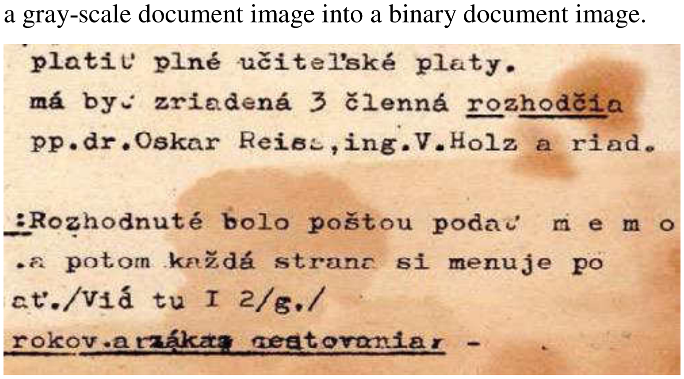
Abstract*—Image binarization is the procedure of parting of pixel values into dual-collections, black as foreground and white as background. Thresholding has found to be a well-known technique used for binarization of document images. Thresholding is further divide into the global and local thresholding technique. In document with uniform contrast delivery of background and foreground, global thresholding has found to be best technique. In degraded documents, where extensive background noise or difference in contrast and brightness exists i.e. there exists many pixels that cannot be effortlessly categorized as foreground or background. In such cases, local thresholding has significant over available techniques. The main objective of this paper is to evaluate the different image binarization techniques to find the gaps in existing techniques.

***Keywords*— Documents, Binarization, thresholding, binary image.**

I. INTRODUCTION

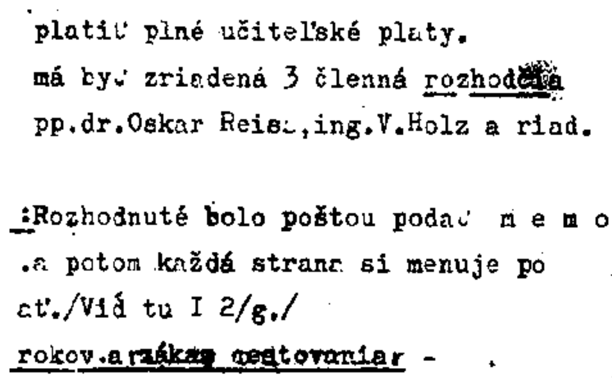
A binary image [1] is a digital image that has only two possible values for each pixel. Typically the two colors used for a binary image are black and white though any two colors can be used. The color used for the objects in the image is the foreground color while the rest of the image is the background color. Binary images [2] often arise in digital image processing as masks or as the result of certain operations such as segmentation and thresholding. Some input/output devices, such as laser printers, fax machines, and bi-levelcomputer displays, can only handle bi-level images. Binary images are produced from color images by segmentation.

Many approaches and techniques were developed to enhance documents images quality. Binarization is one of the most important pre-processing steps which consists to separate foreground and background of documents images. It converts a gray-scale document image into a binary document image.


platiť plné učiteľské platy.
má byť zriadená 3 členná rozhodčia
pp.dr.Oskar Reis.,ing.V.Holz a riad.

:Rozhodnuté bolo poštou podať m e m o
.a potom každá strana si menuje po
at./Viď tu I 2/g./
rokov.a rzákaz gestovania r -

Fig.1. Input Image


platiť plné učiteľské platy.
má byť zriadená 3 členná rozhodčia
pp.dr.Oskar Reis.,ing.V.Holz a riad.

:Rozhodnuté bolo poštou podať m e m o
.a potom každá strana si menuje po
at./Viď tu I 2/g./
rokov.a rzákaz gestovania r -

Fig.2. Binarized Image

Figure 1 has shown the input image for binarization and figure 2 has shown the binary image for the same.

Generally, they can be classified into three major types: global binarization, local binarization and hybrid binarization methods.

A. The global Thresholding Technique

Computes an optimal threshold for the entire image, these techniques need few computations and can work well in simple cases. But fails in complex backgrounds, such as non-uniform color and poor illuminated backgrounds. These methods are usually not suitable for degraded document images, because they do not have a clear pattern that separates foreground text and background.

B. The local Binarization Techniques

Set different thresholds for different target pixels depending on their neighborhood/local information. Generally, these techniques are sensitive to background noises due to large variance in case of a poor illuminated document or bleed-through degradation.

C. Hybrid Binarization Approach

Combines global and local thresholding. A first step consists in carrying out a global thresholding to classify a part of the background of the document image and keep only the part containing the foreground (graphics or text in our case). A second step aims to refine the image obtained by the previous step in order to obtain a sharper result by applying an adaptive thresholding Technique [3].

D. Dynamic Threshold Binarization

Such as iteration method defines the threshold of a pixel with the grey-level values of its own and neighboring pixels and the coordinate of the pixel. This binarization method is commonly used for the bad quality images, especially the images with single-peak histogram. However, owing to the dynamic threshold calculation, the method has high computation complexity and slow speed. Following Table shows a simple comparison of some common binarization algorithms. In the table, the dynamic threshold binarization method only has a "medium" evaluation due to its problems in the computation complexity. Therefore, there is a considerable need of improving the dynamic threshold binarization algorithm [8].

Table 1 is showing the comparative analysis of different binarization approaches.

Table 1. Image Binarization Algorithms

Algorithm	**Suitable condition**	**Shortage**	**Over all**
Itration method	Distinct double peak histogram	Non-uniform illumination	Avg
Uniform gradient	Natural scene	_	Good
Multi-threshold gradient	Non-uniform illumination	_	Very good
Texture based	Bill, text and car license	_	Very good
Differential histogram	Single-peak histogram	Complex grey-level change	Avg
Maximal variance	No histogram	Geometric structure	Good

II. RELATED WORK

Bolan et al. [1] has proposed a classification framework to combine different thresholding methods and produce better performance for document image binarization. Instead of designing a new binarization method, the self-training strategy on existing binarization methods has been used, which has improved not only the performance of existing binarization methods, but also the robustness on different kinds of degraded document images. The given framework divides the document image pixels into three sets, namely, foreground pixels, background pixels and uncertain pixels. A classifier is then applied to iteratively classify those uncertain pixels into foreground and background, based on the pre-selected froeground and background sets. Doccument image binarization methods can be categorized into two main methods: global thresholding method and, local thresholding method. The global thresholding methods are usually not suitable for degraded document images, because they do not have a clear bimodal pattern that separates foreground text and background. So the local thresholding methods are better approaches for degraded document images with non-uniform background and foreground distribution.

Sheikh et al. [2] proposed a new method for image binarization. This is a modified and improved version of the iterative partition based algorithm. In the given methodology in the authors have used the global thresholding algorithm Otsu. This algorithm has been used to calculate the threshold values for different sub-images evolved in the partitioning process. Otsu's method gives very good result for the images whose pixel values have a tendency of clustering towards two major gray levels i.e., the histogram analysis gives a maximum of two sharp peaks. So if the histogram has a maximum of two peaks, then this is logical to apply Otsu's method in that particular partition. And if the number of peaks is more than two, then Otsu's method may not give satisfactory result for that partition. The results of implementation of the algorithms compared to the other five methods in terms of two quantitative measures, namely, misclassification error and the relative foreground area error.

Zayed et al. [3] have proposed an effective document image binarization technique to work with various kind of document degradation such as, bleed-through, aging and bad illumination conditions. Hybrid binarization approach combines global and local thresholding in order to preserve textual and picture information. The objective of this work was to propose a hybrid thresholding method, which gives better binarization results independently from historical documents degradations such as, uneven illumination, smears and bleed-through degradation. They proceeded to a comparison between different global methods based on the recall criterion, a good recall rate means that the method preserves an ambiguous pixels as not classified as background. They then compared the different methods of local binarization based on the criterion of precision which gives a clear idea about the ability of the method to correctly classify a pixel. First, they compared the final outcome of different methods of binarization and then they compared the different methods for different types of degradation. The proposed hybrid method gives the best results regardless of the type of degradation.

Maythapolnun et al. [4] reviewed global binarization algorithms for improving corrupted document image, thus representing their differences and similarities, and also their benefits and drawbacks. They have been classified into three groups, which are global thresholding, local thresholding and hybrid thresholding. Algorithms for image global threshold binarization are summarized here. No particular thresholding technique could be claimed as the most excellent method. This article reviewed the global binarization algorithms for degraded images which are to be scanned. The thresholding algorithms are easy, fast and easy for digital image segmentation, which is effectively used in a large binarization of application of digitization. A number of global thresholding methods and algorithms depend on objects and background, and could have a drawbacks due to the double-sided results in historical document images that are affected by ink lose blood from paper and noisy background. Usually, document image study uses thresholding as a regular algorithm to change the gray document images to binary mode. Document image binarization is enormously important for older papers to be digitized into digital information. Upcoming work will focus on the local binarization and hybrid binarization.

Ying et al. [5] has proposed a morphology-based Otsu binarization algorithm for sediment image processing. The image segmentation technique is the key link in river sediment image processing, segmentation results have a direct impact on the count accuracy. The diversity and complexity of sediment image particle, has become a constraint major bottleneck to the establishment of river sediment image segmentation, right now there is no completely general-purpose segmentation method. Experiments show that the new algorithm to maintain the speed and good noise suppression ability, at the same time this algorithm can better retain the original image characteristics, the algorithm execution speed has been greatly improved. Images have good results after binary.

Elisa and change [6] discovered that the binarization alone is not evaluated, but rather how it interacts with the downstream processes. Recently pixel accurate “ground truth” images have been introduced for use in binarization algorithm evaluation. The choice of binarization ground truth affects the binarization algorithm design, either directly if design is by automated algorithm trying to match the provided ground truth, or indirectly if human designers adjust their designs to perform better on the provided data. The performance can vary significantly depending on choice of ground truth, which can influence binarization design choices. Mixing pixel level results with overall system performance is likely a better way to evaluate the binarization algorithm. Perhaps having available end-to-end system, allow the binarization algorithms to be tested in a ‘goal directed’ fashion in conjunction with the pixel accurate ground truth.

Zhang et al. [7] improved background gray-level based binarization algorithm to ensure the binarization quality, but also reduce the computation. The non-uniform illumination QR code image would present all white or all black error areas after global binarization processing. This paper put forward an improved background gray-level based binarization algorithm for the non-uniform illumination QR code image. Firstly, doing sub-block processing according to the size of the QR code image. On this basis, using gray-level estimation formula for computing the gray-level value of each block. Secondly, using joint interpolation algorithm to build the background gray level image. Then using the background gray-level image to adapt the original image to get the corrected image. Finally, using the Otsu algorithm for the corrected image binarization. Experiments show that the algorithm can do effective correction to uneven illumination QR code image, get a good binary image.

Qinghua et al. [8] presented an image binarization algorithm based on the differential evolution algorithm, called BAbDE. Differential Evolution algorithm is a population based stochastic algorithm for solving optimization problem. Due to its simplicity, effectiveness and robustness, Differential evolution has been successfully applied in solving optimization problems arising in many practical applications, such as parameter identification, function optimization, multi-objective optimization, optimal design, scheduling problem. This paper focuses on the performance of the Differential evolution algorithm in the image binarization processing. Numerical experiments are firstly conducted to study the setting of two control parameters, the mutation factor and the crossover probability, of BAbDE. BAbDE is then compared with Otsu’s method and Kmeans method in their thresholds, processing times and entropies. The experimental results for several common-used images show that BAbDE is practicable for image binarization, and its entropy is smaller.

Ma et al. [9] Introduced an improvement of multi-threshold dynamic binarization for bill images. The improvement scheme is based on the research of existing binarization algorithms and the motivation to solve the problem of noise jamming in bill images. An application of the improved binarization scheme was implemented. The result shows that the scheme is effective and has a well anti-noise ability. Image binarization is a common method of processing grey-level images using grey-level transformation. Nornally, it defines a threshold value to divide a grey-level image into two pixel groups. Theoretically, the group with bigger pixel values than threshold denotes the object and the group with smaller pixel values than threshold is the background. The experiment result shows that the improved algorithm has a good anti-noise capability.

Patvardhan et al. [10] proposed a Curvelet based method that is able to remove complex image background as well as white Gaussian noise and results in a better binarized document image as compared to other conventional methods. The ability of sparse representation. The proposed method is able to remove low frequency complex backgrounds and high frequency. In this paper, an effort has been made to remove white Gaussian as well complex image background to achieve better character recognition with help of curvelet transform based approach. The Gaussian noise corrupts the images when mainly transmitted over the analog transmission channels such as Television. Different types of test images are taken to test the capacity of proposed scheme and in case of all types of images; proposed scheme performs better than others.

Wagdy et al. [11] presented fast and efficient document image clean up and binarization method based on retinex theory and global threshold. The given method is fast and produces high quality results compared to the previous works. Document image clean up and binarization is an interesting research topic in image processing that used in most of document image analysis and retrieval. The given method combines the advantages of local and global thresholding by using the concept of retinex theory which can effectively enhance the degraded and poor quality document image. After that, fast global threshold is used to convert the document image into binary form.

Parker et al. [12] described an automated image enhancement approach that requires no training data. The approach is applicable to images of type written text as well as hand written text or a mixture of both. The pair of parameters used by the approach is automatically self-tuned according to the input image. Converting the information found to public knowledge occurs more quickly and cheaply if an automatic method to enhance these degraded documents is used instead of enhancing each document image by hand. The method described here converts a colour or gray scale document image to a strictly black and white document image. The conversion technique is designed to simultaneously reduce the effect of document degradation and highlight the essence of the pre-degraded document. The ultimate goal is to produce a single black and white image that makes the information in the original document as legible as possible. The parameters of our method are set automatically. No human interaction or guidance is required to determine which parameter values to use and it does not require training data of any kind. Avoiding the need for human interaction can significantly improve the throughput of image digitization and archival effects.

Sudipta et al. [13] wished for a new image binarization process that used an easy standard deviation method and gave us extremely fine outcomes for MRI of head images. The difficulty of binarization of gray MRI images because of the black background and great intensity difference has been defea by our planned method. Threshold of image was stronge-minded by standard deviation multiplied by a heuristic value. This method was extremely helpful to take out the useful objects from an image and, hence, to differentiate the brain from the black background. for MRI of brain tumor images the proposed methods got very fine outcomes and formed importent information compared to the other well-known method like Otsu, Savala, Niblack, Bernsen, Kapur and Otsu as a frame work in iterative partition. They compared thier technique with other well known popular algorithms which are shortly describe. For some of the images local thresholding method Bernsen formed acceptable results. Deviding framework method do not produce excellent results. Their proposed method was a global thresholding methods produced awfully good results for MRI of brain. From the metric ME, RAE is less in their proposed method and F-measure were greater value which was expected. Thus thier proposed method was very good and efficient algorithms for binarization of MRI of brain.

Bolan et al. [14] proposed a different document image binarization technique that addressed these problems by using adaptive image contrast. The adaptive image contrast was a mixture of the local image contrast and the local image gradient that was relaxed to content and background changes caused by dissimilar types of text degradations. In the planned technique, an adaptive contrast diagram was first constructed for an input degraded text image. The contrast map was then binarized and joined with Canny's edge map to recognize the content stroke edge pixels. The record text then segmented by a local threshold that is expected based on the intensities of searched text stroke edge pixels within a local window. The proposed method was simple, robust, and involved least constraint alteration. The proposed technique was easy and strong, only a small number of parameters were mixed up. Moreover, it worked for different kinds of degraded text images. This paper obtained an adaptive image contrast based document image binarization technique that was relaxed to different types of document degradation such as uneven illumination and document smear. The proposed method has been tested on the a mixture of datasets. The proposed technique maked utilize of the local image contrast that was evaluated based on the local maximum and minimum.

Patel et al. [15] have discussed a variety of image segmentation techniques which were developed for distinct domains and compared this techniques based on the changed parameters. As image segmentation technique is not simple task to perform no image segmentation also can give 100% correctness but quiet there was a capacity for just biginning an image segmentation technique to identify many objects which could also effort on different dimensions of images. There were a lot of image segmentation techniques existing today. Videos and images have become necessary section of human life in latest era. These images or videos contain a lot of other characters which a someone would like to take out. A number of the techniques work for particular size of image simply. It is also not understandable that whether these techniques maintain multiple object segmentation or not. Few image segmentation techniques hold up many object segmentation. In the comparative study of different image segmentation techniques it quite clear that not any of the existing image segmentation technique provides 100% accuracy. The proposed system works for unchanging threshold of 120 and it works for Indian vehicle number plates, which are as per the Indian Road Transport Office strategy. So it might not work for other number plates which are non uniform and other image objects.

III. CONCLUSION

This paper has focused on the degraded document binarization technique. Document binarization is an important application of vision processing. The main objective of this paper is to evaluating the short comings of algorithms for degraded image binarization. It has been found that each technique has its own benefits and limitations; no technique is best for every case. The main limitations of existing works are found to be noisy and low intensity images. In near future we will propose a new algorithm which will use more reliable methodology to enhance the work. We will propose a new algorithm which will use non-linear enhancement as preprocessing technique to improve the results further.

REFERENCES

[1] Su, Bolan, Shijian Lu, and Chew Lim Tan. "Combination of document image binarization techniques." In Document Analysis and Recognition (ICDAR), 2011 International Conference on, pp. 22-26. IEEE, 2011.

[2] Shaikh, Soharab Hossain, Asis Maiti, and Nabendu Chaki. "Image binarization using iterative partitioning: A global thresholding approach." In Recent Trends in Information Systems (ReTIS), 2011 International Conference on, pp. 281-286. IEEE, 2011.

[3] Zayed, Mohamed, Asma Ouari, Meriem Derraschouk, and Youcef Chibani. "An effective hybrid thresholding technique for degraded documents images binarization." In Internet Technology and Secured Transactions (ICITST), 2011 International Conference for, pp. 460-465. IEEE, 2011.

[4] Athimethphat, Maythapolnun. "A Review on Global Binarization Algorithms for Degraded Document Images." AU JT 14, no. 3 (2011): 188-195.

[5] Ying, Xiao, Yuan Bo, Zeng Dajun, Wang Bing, and Xie Fengying. "Morphology Based Sediment Particle Image Binarization Algorithm Research." In Computer Science and Electronics Engineering (ICCSEE), 2012 International Conference on, vol. 3, pp. 375-378. IEEE, 2012.

[6] Barney Smith, Elisa H., and Chang An. "Effect of "Ground Truth" on Image Binarization." 2012.

[7] Zhang, Yinghui, Tianlei Gao, Deguang Li, and Huaqi Lin. "An improved binarization algorithm of QR code image." In Consumer Electronics, Communications and Networks (CECNet), 2012 2nd International Conference on, pp. 2376-2379. IEEE, 2012.

[8] Su, Qinghua, Zhangcan Huang, Zhongbo Hu, and Xiaohong Wang. "Binarization algorithm based on differential evolution algorithm for gray images." In Fuzzy Systems and Knowledge Discovery (FSKD), 2012 9th International Conference on, pp. 2611-2615. IEEE, 2012.

[9] Ma, Chi, Qiuying Bai, and Xuechang Chen. "The new improvement of multi-threshold dynamic binarization for bill images." In Software Engineering and Service Science (ICSESS), 2012 IEEE 3rd International Conference on, pp. 67-70. IEEE, 2012.

[10] Patvardhan, C., A. K. Verma, and C. Vasantha Lakshmi. "Document image denoising and binarization using Curvelet transform for OCR applications." In Engineering (NUiCONE), 2012 Nirma University International Conference on, pp. 1-6. IEEE, 2012.

[11] Wagdy12, Marian, Ibrahima Faye, and Dayang Rohaya. "Fast and Efficient Document Image Clean Up and Binarization Based on Retinex Theory."

[12] Parker, Jon, Ophir Frieder, and Gideon Frieder. "Automatic Enhancement and Binarization of Degraded Document Images." In Document Analysis and Recognition (ICDAR), 2013 12th International Conference on, pp. 210-214. IEEE, 2013.

[13] Bolan Su, Shijian Lu, And Chew Lim Tan, " Robust Document Image Binarization Technique For Degraded Document Images", Transactions On Image Processing, Vol. 22, No. 4, April. 2013 IEEE.

[14] Roy, Sudipta, and Kingshuk Chatterjee. "A New Efficient Binarization Method for MRI of Brain Image." Signal 3,2012.

[15] Patel, Chirag, Atul Patel, and Dipti Shah. "Threshold Based Image Binarization Technique for Number Plate Segmentation." International Journal 3, no. 7,2013.

A Review: Analysis of Various Types of Watermarking Techniques

Pawandeep Kaur
Department of CSE
SBS State Technical Campus
Ferozepur, India
pawandeep234@gmail.com

Sonika Jindal
Department of CSE
SBS State Technical Campus
Ferozepur, India
sonikamanoj@gmail.com

***Abstract*— Digital watermarking is used to protect images that have been tampered maliciously and with higher accuracy. Also due to transmission of multimedia content over the internet, digital watermarks are helpful in various applications like copyright management, tamper detection, content authentication, fingerprinting and broadcast monitoring. In this paper, review of several aspects of watermarking has been defined and to check the image quality various performance evaluation metrics has also been discussed. So, our main focus is on various types of watermarking and analysis of numerous watermarking techniques such as SVD, SVD-DWT, and SVD-DCT considering the image quality, robustness and image imperceptibility.**

***Keywords*— Watermarkin, Performance Evaluation Metrices Singular Value Decomposition, DWT, DCT**

I. INTRODUCTION

The extensive range of information services and quick distribution of information has been a major factor in the development of new technology [1]. The Internet is a global system through which the use of multimedia content is rising day by day. Multimedia distribution has become a crucial way to deliver services to people all over the world. Due to the growing usage of multimedia content (image, text, audio, video, etc.) on the Internet, serious issues have emerged such as counterfeiting, forgery, fraud, and pirating of the content [2].Copyrighted material can be easily exchanged over peer-to-peer networks. In order to protect the data of the content providers, these digital contents or data can be watermarked.

Digital watermarking is the process in which secret key is used which decide the location where the watermark would be embedded [3] [4]. In detection, (often called extraction) this is applied to the signal to attempt to extract the watermark as shown in Fig: 1. If the signal was unmodified during transmission, then the watermark still is present and it may be extracted [9].

Watermarks should be robust against all possible intended and unintended attacks such as removal attacks, protocol attacks, cryptographic attacks, and geometric attacks. Watermarking can be classified using several different methods. The two broad categories of watermarking are spatial domain and frequency domain [5]. Recently (SVD) is explored as a new transform technique for watermarking and this proposed technique is more robust with minimum and no distortion when compared to the previous techniques [6].Watermarks are used in variety of applications such as copyright protection of copyright material, broadcast monitoring of content, for authentication purposes so that data can be protected from any modification, for tamper detection by making use of fragile watermarks which get destroyed if any tampering is done with digital content and to check the owner of content, they also used in fingerprinting [7].

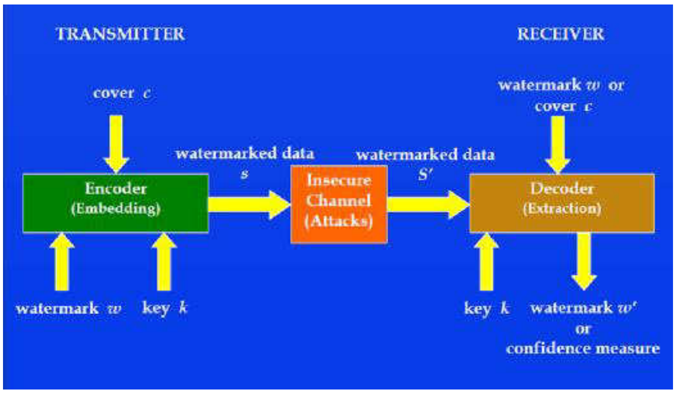


Fig. 1. Digital image watermarking process

The paper is organized as follows. Section 2 discusses the types of digital image watermarking. Section 3 describes performance evaluation metrics. Section 4 discusses a comparative analysis of DWT, DCT transform techniques of watermarking using SVD. Section 5 defines conclusion.

II. TYPES OF DIGITAL IMAGE WATERMARKING

The watermarking can be classified according to its importance and goals. So; various types of watermarking are shown in Fig: 3.

A. According to working domain

i. Spatial Domain Watermarking: In spatial domain, images are represented by pixels. Simple watermarks could be embedded by modifying the pixel values or the least significant bit (LSB) values [3]. It directly loads the raw data into the image pixels. Some of its algorithms are LSB and SSM Modulation based technique.

a) Least Significant Bit (LSB): This is the simplest algorithm in which information can be inserted into every bit of image information. Given an image with pixels, and each pixel being represented by an 8-bit sequence, the watermarks are embedded in the last (i.e. least significant bit) of selected pixels of the image [8] [9]. [10] Proposed a simple data hiding technique by simple LSB substitution. In this technique last bit of host data is randomly changed and produce the watermarked data at output.

b) Spread Spectrum Modulation (SSM): This technique is applied in the watermarking algorithms with linked information. Attached to the original image with pseudo noise signal; it is modulated by the watermark. This is done for the establishment of secure communications and jamming which provide the greatest robustness [8].

ii. Frequency Domain Watermarking: In frequency domain, image is represented in terms of its frequencies. Robust watermarks could be embedded in the transform domain of images by modifying the transform domain coefficients [11]. The most commonly used transforms are the Discrete Cosine Transform (DCT), Discrete Fourier Transform (DFT) and Discrete Wavelet Transform (DWT).SVD is also used as transform technique for image watermarking.

a) Discrete Cosine Transform (DCT): Discrete Cosine Transform is a technique for converting a signal into elementary frequency components [12]. DCT based watermarking techniques are robust compared to spatial domain techniques. Such algorithms are robust against many image processing operations like brightness, low pass filtering, contrast adjustment, blurring etc. However, their implementation is difficult and computationally they are more expensive. [13] Proposed a technique in which watermark is inserted into the cover image through using original classifier. Then watermark is detected to achieve false alarm probability using watermark detector. [14] Proposed a visible watermarking scheme which is applied to the original image

in the DCT domain and some modifications are done to make the watermark highly robust.

b) Discrete Fourier Transform (DFT): The DFT allows an image to be broken up into different frequency bands.This division making it much easier to embed watermarking information into the middle frequency bands of an image [3]. DFT is rotation, scaling and translation (RST) invariant and can be used to recover from geometric distortions. DCT is not RST invariant and hence it is not possible to overcome from geometric distortions. [15] Proposed rotation in the spatial domain causes the same rotation in the frequency domain. However, Scaling in the spatial domain causes inverse scaling in the frequency domain.

c) Discrete Wavelet Transform (DWT): Wavelet domain is a frequency domain technique in which cover image is transformed into frequency domain and then its frequency coefficients are modified in accordance with the transformed coefficients of the watermark [7]. Here the wavelet filters are used to transform the image. Many filters are available, although the most commonly used filters for watermarking are Haar Wavelet Filter, Daubechies Bi-Orthogonal Filters and Daubechies Orthogonal Filters. This technique decompose an image in basically three spatial directions i.e., horizontal, vertical and diagonal and also separating the image into four different components namely LL, HL, LH and HH. LL level is the lowest resolution level which consists of the approximation part of the cover image. The remaining three levels i.e., LH, HL, HH give the detailed information of the cover image. Fig: 2 shows the three level of decomposition using DWT. The Discrete Wavelet Transform (DWT) is currently used in a wide variety of signal processing applications, such as in audio and video compression and removal of audio noise [16].

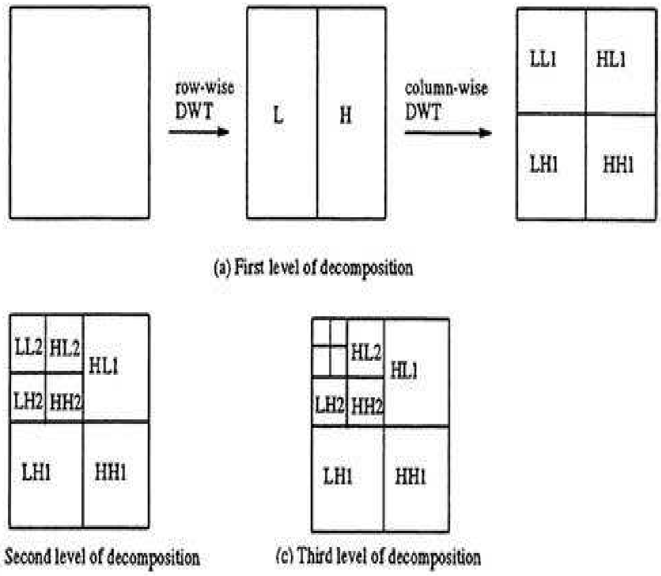


Fig. 2. Three different levels of decomposition using DWT

d) Singular Value Decomposition: SVD is a mathematical tool which decomposes a matrix into two orthogonal matrices and one diagonal matrix consisting of the singular values of the matrix [28].The SVD mathematical technique provides an elegant way for extracting algebraic features from an image. This method improves watermark robustness and resistance against many kinds of attacks. SVD is a stable and an effective method to split the system into a set of linearly independent components. A digital Image X of size MxN can be represented by its SVD as follows:

$$X = USV^T \qquad (1)$$

$$S = \quad U = [U_1, U_{22} \ldots . U_m],$$

$$V = [V_1, V_{22} \ldots V_n], \qquad (2)$$

Where U is an MxM orthogonal matrix, V is an NxN orthogonal matrix, and S is an MxN matrix with the diagonal elements represents the singular values. T denotes the transpose of the matrix.

B. According to type of document

i. Text watermarking: Watermarking of text documents provides a means of tracing documents that have been illegally copied, distributed, altered or forged [17]. This adds watermark to the PDF, DOC and other text file in order to prevent the text changes. The watermark is inserted in the font shape.

ii. Image watermarking: This is used to hide the special information into the image and to later detect and extract that special information for the author's ownership. Digital images can be produced from photographs, computer graphics, satellite pictures and medical scans [18].

iii. Audio watermarking: An audio watermark is a unique electronic identifier embedded in an audio signal. It is used to identify ownership of copyright content. It is just like a watermark on a photograph. One of the most secure techniques of audio watermarking is spread spectrum audio watermarking (SSW) [19] [20].

iv. Video watermarking: Video Watermarking is a new field in the area of multimedia. Video watermarking embeds data in the video for the purpose of Identification, annotation and copyright. Video watermarking techniques are used in variety of applications, such as broadcast monitoring, video authentication, copyright protection etc [20].

C. According to human perception

i. Visible watermarking: Visible watermarking is a technique in which a visible pattern or image is inserted into a source image. Visible watermarks change the signal altogether such that the watermarked signal is totally different from the actual signal. A good visible watermark must be difficult to remove unless exhaustive and expensive human interventions are involved. [21] suggested a visible watermarking technique in which random noise is inserted to increase the difficulty of unauthorized removal.

ii. Invisible watermarking: Invisible watermarking is the process in which a watermark can't be seen but it can be detected by some computer program. It is used in many applications, such as to indicate the ownership of originals, for a trustworthy camera, to detect alternation of images and to detect miss-appropriated images [22].Invisible watermarking is further divided into two types i.e., robust and fragile watermarking.

a) Robust watermarking: Robust watermarking is a technique in which any modification to the watermarked content will not affect the watermark. A robust watermark is mainly designed to compensate with all kind of attacks such as lossy compression, filtering, and geometric scaling that attempt to remove and destroy the watermark [24].There are many types of robust watermarking. Private watermarking is also called non blind watermarking in which the users are not authorized to detect the watermark. In this the original cover is required during detection of watermark [24]. [25] Public watermarking is also called blind watermarking in which users of the content are authorized to detect the watermark. In this there is no need of original data but secret key is required. Invertible watermarking is basically a reverse watermarking technique in which a fake original asset and a fake watermark is designed and then fake watermark is inserted into fake original asset which produces a watermarked asset similar to initial one. Quasi invertible is an extension of invertible watermarking there is no need to insert the fake original watermark into fake original asset but only the presence of fake watermark is revealed when watermark detector is applied to the digital asset.

b) Fragile watermarking: It is a technique in which watermark gets destroyed when watermarked content is modified or tampered with. To detect changes to the watermarked content, fragile watermark is designed and used in content authentication [24].

D. According to application

i. Source based watermarking: Source based watermarks are desirable for ownership identification and authentication, Unique watermark is required to identify the owner is introduced to all the copies of a particular image being distributed. This watermarking is used for authentication and

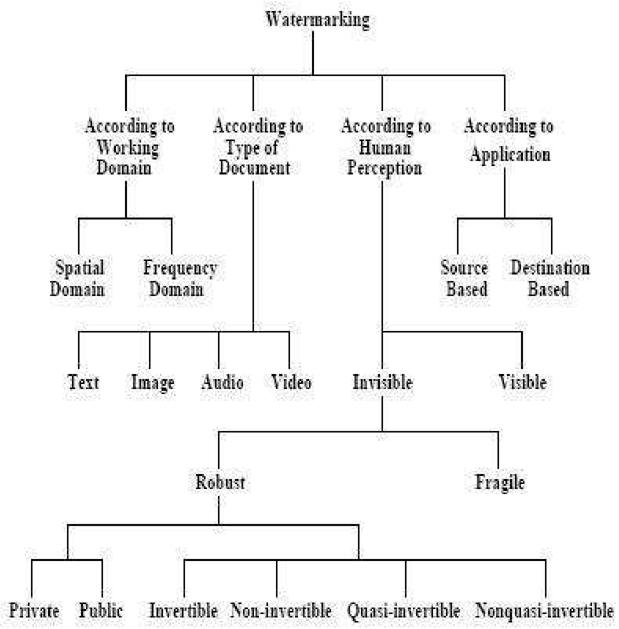


Fig. 3. Classifications of various types of digital image watermarking

to determine whether a received image or other electronic data has been tampered with [26].

ii. Destination based watermarking: [26] Defined in destination based watermarking, each distributed copy gets a unique watermark identifying the particular buyer. In the case of illegal reselling, the destination based watermark could be used to trace the buyer.

III. PERFORMANCE EVALUATION METRICS

The performance of the watermarked images must be evaluated by using some quality measures such as MSE, SNR, PSNR and BER.

i. The MSE (mean square error): [27] Defined it as average squared difference between a reference image and a distorted image. It is calculated as:

$$MSE = \frac{1}{XY}\sum_{i=1}^{X}\sum_{j=1}^{Y}(c(i,j) - e(i,j))^2 \qquad (3)$$

X represents the height and Y represents the width of the image. c(i,j) and e(i,j) are the respective pixel value of the original image and embedded image.

*i. SNR (Signal to Noise ratio):*It is defined as the ratio of signal power to the noise power. It is expressed in decibels. A ratio higher than 1:1 (greater than 0 dB) indicates more signal than noise. It measures the signal strength relative to the background noise (Unwanted signal). It is calculated by the formula given below:

=

*i. The PSNR (peak signal to noise ratio):*It is a quality metric used to determine the degradation in the embedded image with respect to the host image or also defined as ratio between maximum power of a signal and power of distorted signal [27]. It is most easily defined via the mean squared error (MSE) as:

PSNR =

L denotes the peak signal value of the cover image which is equal to 255 for 8 bit images.

ii. The BER (bit error ratio): [27] Defined it as the ratio that describes how many bits received in error over the number of the total bits received. It is often expressed as percentage and calculated by comparing bit values of embedded image and cover image.

BER =

H and W are the respective height and width of the water-marked image. P is the count number whose initial value is zero and it increments by one if there is any bit difference between cover and embedded image.

IV. A COMPARATIVE ANALYSIS OF VARIOUS TRANSFORM TECHNIQUES OF WATERMARKING USING SVD

i. Singular value decomposition(SVD): [6] proposed an SVD based watermarking technique in which watermark is added to the SVD domain of the original image.This technique provide the security and robusteness but if this is combined with encryption rightful ownership is not received by this technique.[29] proposed a technique in which a some section of watermark is embedded in each block by dividing the cover image into smaller blocks and then complete watermark is embedded by considering the cover image as a single block.So this twice watermark embedding technique is useful to acheive robustness to attacks.Further in most researches [30] [31] [32] developed various watermarking techniques in which only the singular values of watermark are embedded into the host image, which are not more robust and might be affected by several attacks.In order to solve the ownership problem in [31], the author tried to embed not only the singular values of the watermark, but also the principal component of watermark into the host image.Also scaling factor plays an important role to control the transparency and robustness of the watermarked image.[33] [34] developed a technique in which original image is divided into blocks and watermark is inserted in SVs of each block, which provide the more robustness against variety of attacks such as noise, cropping and compression of images.

ii. DWT-SVD Based Watermarking: DWT technique is used in many watermarking techniques in which modifications are mainly done on DWT coefficients.[35] proposed a technique in which original image is divided into four bands and SVD is applied on each band and (SVs) were modified to embed the same watermark. This technique provides robust against variety of attacks.[36] proposed a new technique in which watermark is hidden in full frequency bands of color images and to measure the quality of watermark images two performance evaluation metrices are used such as peak signal to noise ratio (PSNR) and normalized correlation (NC) which provided robustness for various attacks including salt and pepper noise and Gaussian noise, cropping and JPEG compression. To protect the ownership of multimedia content [37] developed a new technique in which DWT and IDWT transformation are performed to obtain four frequency images and then by using SVD, watermark is embedded in high frequency bands.This technique is suitable to resist geometric attacks.[38] presented SVD-DWT semi blind reference watermarking in which original image is transformed into wavelet domain .By using directive contrast and wavelet coefficients reference sub image is formed and watermark embedded into reference image modifying the singular values of reference image using the singular values of the watermark.This technique proved a protection against variety of attacks.

iii. DCT-SVD Based Watermarking: DCT is a transformation that transforms a signal from spatial domain to frequency domain.[39] proposed a technique in which original bitmap is converted into frequency domain so that DCT matrices can be obtained from its block. To achieve the randomness, the pseudorandom bit generator is used that determines in which color component the bits of logo image are embedded.[40] Proposed a new technique in which only the SVs of a recognized pattern are embedded into the original image. Local peak signal to noise ratio (LPSNR) is used to achieve the maximum robustness without loss the quality of original image.[41] proposed a new modified DCT SVD watermarking which is robust against various geometric attacks. This technique is RST invariant but did not provided the good correlation coefficient's values. [42] Presented robust hybrid watermarking scheme based on DCT and SVD in which DCT coefficients are mapped into four quadrants in the form of zig-zag pattern and SVD is applied to each quadrant. The singular values of quadrant are modified by singular values of the DCT-transformed visual watermark. Watermark embedded into lowest and highest frequencies are resistant to variety of attacks. So, this technique provides more robustness and reliability.

V. CONCLUSION

This paper provides literature review on digital watermarking .We surveyed various types of watermarking based on the robustness and imperceptibility of watermark. To check the image quality, various performance metrics are used which gives the better performance of watermarked images. Also, we studied the various techniques in transform domain using a new transform technique SVD, which improves the watermark resistant against variety of attacks.

ACKNOWLEDGMENT

The author would like to thank Dr. Satvir Singh, Associate Professor, Department of Electronic & Communication Engineering, SBS State Technical Campus, Ferozepur (Punjab), India, for sparking in me the enthusiasm and initiative to discover and learn. I am truly thankful to him for guiding me through the entire paper and being as a motivator in this learning curve.

REFERENCES

[1] M. M. Yeung, “Digital watermarking,” Communications of the ACM 41 (7) (1998), pp. 31–33.

[2] C. M. Wolak, “Digital watermarking,” Preliminary Proposal, Nova South-eastern University, United States.

[3] V. M. Potdar, S. Han, E. Chang, “A survey of digital image watermarking techniques,” in: Industrial Informatics, 2005. INDIN’05. 2005 3rd IEEE International Conference on, IEEE, 2005, pp. 709–716.

[4] E.-H. Fu, “Literature survey on digital image watermarking,” Lectures notes on EE381K Multidimensional Signal Processing.

[5] G. C. Langelaar, I. Setyawan, R. L. Lagendijk, “Watermarking digital image and video data. a state-of-the-art overview,” Signal Processing Maga-zine, IEEE 17 (5) (2000),pp. 20–46.

[6] R. Liu, T. Tan, “An svd-based watermarking scheme for protecting rightful ownership,” Multimedia, IEEE Transactions on 4 (1) (2002),pp. 121–128.

[7] I. Cox, M. Miller, J. Bloom, M. Miller, “Digital watermarking,” Morgan Kaufmann, 2001.

[8] J. Liu, X. He, “A review study on digital watermarking, in: Information and Communication Technologies,” 2005. ICICT 2005. First International Conference on, IEEE, 2005, pp. 337–341.

[9] R. Chandramouli, N. Memon, “Analysis of lsb based image steganography techniques,” 2001. Proceedings. 2001 International Conference on, Vol. 3, IEEE, 2001, pp. 1019–1022.

[10] C.-K. Chan, L.-M. Cheng, “Hiding data in images by simple lsb substitution,” Pattern recognition 37 (3) (2004), pp. 469–474.

[11] F. Y. Shih, S. Y. Wu, “Combinational image watermarking in the spatial and frequency domains,” Pattern Recognition 36 (4) (2003), pp. 969–975.

[12] M. Jiansheng, L. Sukang, T. Xiaomei, “A digital watermarking algorithm based on dct and dwt,” in: Proceedings of the 2009 International Sympo-sium on Web Information Systems and Applications, 2009, pp. 104–107.

[13] B. Tao, B. Dickinson, “Adaptive watermarking in the dct domain,” in: Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on, Vol. 4, IEEE, 1997, pp. 2985–2988.

[14] S. P. Mohanty, K. R. Ramakrishnan, M. S. Kankanhalli, “A dct domain visible watermarking technique for images,” in: Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on, Vol. 2, IEEE, 2000, pp. 1029–1032.

[15] V. Solachidis, L. Pitas, “Circularly symmetric watermark embedding in 2d dft domain,” Image Processing, IEEE Transactions on 10 (11) (2001), pp. 1741–1753.

[16] E. Brannock, M. Weeks, R. Harrison, “Watermarking with wavelets: simplicity leads to robustness,” in: Southeastcon, 2008. IEEE, IEEE, 2008, pp. 587–592.

[17] J. Brassil, S. Low, N. Maxemchuk, L. OGorman, “Hiding information in document images,” in: Proceedings of the 29th Annual Conference on Information Sciences and Systems, Citeseer, 1995, pp. 482–489.

[18] J. K. Su, F. Hartung, B. Girod, “Digital watermarking of text, image, and video documents,” Computers & Graphics 22 (6) (1998), pp. 687–695.

[19] I. J. Cox, J. Kilian, T. Leighton, T. Shamoon, “Secure spread spectrum watermarking for images, audio and video,” in: Image Processing, 1996. Proceedings., International Conference on, Vol. 3, IEEE, 1996, pp. 243– 246.

[20] R. T. Paul, “Review of robust video watermarking techniques,” IJCA Special Issue on Computational Science 3 (2011), pp. 90–95.

[21] G. W. Braudaway, K. A. Magerlein, F. C. Mintzer, “Protecting publicly available images with a visible image watermark,” in: Electronic Imaging: Science & Technology, International Society for Optics and Photonics, 1996, pp. 126–133.

[22] F. Mintzer, G. W. Braudaway, M. M. Yeung, Effective and ineffective digital watermarks,” in: Image Processing, 1997. Proceedings., International Conference on, Vol. 3, IEEE, 1997, pp. 9–12.

[23] E. T. Lin, E. J. Delp, “A review of fragile image watermarks,” in: Proceedings of the Multimedia and Security Workshop (ACM Multimedia’99) Multimedia Contents, Citeseer, 1999, pp. 25–29.

[24] D. Huang, H. Yan, “Interword distance changes represented by sine waves for watermarking text images, Circuits and Systems for Video Technology,” IEEE Transactions on 11 (12) (2001), pp. 1237–1245.

[25] P. H. Wong, O. C. Au, Y. M. Yeung, “Novel blind multiple watermarking technique for images, Circuits and Systems for Video Technology,” IEEE Transactions on 13 (8) (2003), pp. 813–830.

[26] C. I. Podilchuk, W. Zeng, “Image adaptive watermarking using visual models,” Selected Areas in Communications, IEEE Journal on 16 (4) (1998), pp. 525–539.

[27] A. K. Singh, N. Sharma, M. Dave, A. Mohan, “A novel technique for digital image watermarking in spatial domain,” in: Parallel Distributed and Grid Computing (PDGC), 2012 2nd IEEE International Conference on, IEEE, 2012, pp. 497–501.

[28] K.-L. Chung, W.-N. Yang, Y.-H. Huang, S.-T. Wu, Y.-C. Hsu, “On svd-based watermarking algorithm,” Applied Mathematics and Computation 188 (1) (2007) , pp. 54–57.

[29] E. Ganic, N. Zubair, A. M. Eskicioglu, “An optimal watermarking scheme based on singular value decomposition,” in: Proceedings of the IASTED International Conference on Communication, Network, and Information Security, Vol. 85, 2003.

[30] G. Bhatnagar, B. Raman, “A new robust reference watermarking scheme based on dwt-svd,” Computer Standards & Interfaces 31 (5) (2009) ,pp. 1002– 1013.

[31] C. Jain, S. Arora, P. K. Panigrahi, “A reliable svd based watermarking scheme,”.

[32] E. Ganic, A. M. Eskicioglu, “Robust dwt-svd domain image watermarking: embedding data in all frequencies,” in: Proceedings of the 2004 Workshop on Multimedia and Security, ACM, 2004, pp.

166–174.

[33] R. Ghazy, N. El-Fishawy, M. Hadhoud, M. Dessouky, F. A. El-Samie, “An efficient block-by-block svd-based image watermarking scheme,” in: Radio Science Conference, 2007. NRSC 2007. National, IEEE, 2007, pp. 1–9.

[34] R. A. Ghazy, N. A. El-Fishawy, M. M. Hadhoud, M. I. Dessouky, F. E.-S. Abd El-Samie, “Performance evaluation of block based svd image watermarking,” Progress In Electromagnetics Research B 8 (2008) 147–159.

[35] E. Ganic, A. M. Eskicioglu, “Robust dwt-svd domain image watermarking: embedding data in all frequencies,” in: Proceedings of the 2004 Workshop on Multimedia and Security, ACM, 2004, pp. 166–174.

[36] V. Santhi, A. Thangavelu, “Dwt-svd combined full band robust watermarking technique for color images in yuv color space,” International Journal of Computer Theory and Engineering 1 (4) (2009), pp. 424–429.

[37] P. Saxena, S. Garg, A. Srivastava, “Dwt-svd semi-blind image watermark-ing using high frequency band,” in: Proceedings of, 2012, pp. 138–142.

[38] G. Bhatnagar, B. Raman, “A new robust reference watermarking scheme based on dwt-svd,” Computer Standards & Interfaces 31 (5) (2009), pp. 1002– 1013.

[39] S. E. Baba, L. Z. Krikor, T. Arif, Z. Shaaban, “Watermarking of digital images in frequency domain,” International Journal of Automation and Computing 7 (1) (2010), pp.17–22.

[40] F. Huang, Z.-H. Guan, “A hybrid svd-dct watermarking method based on lpsnr,” Pattern Recognition Letters 25 (15) (2004), pp .1769–1775.

[41] S. R. Haque, “Singular value decomposition and discrete cosine transform based image watermarking,” Master’s thesis. Sweden: Department of Interaction and System Design, School of Engineering, Blekinge Institute of Technology.

[42] A. Sverdlov, S. Dexter, A. M. Eskicioglu, “Robust dct-svd domain image watermarking for copyright protection: embedding data in all frequencies,” in Proc. of European Signal Processing Conference, 2005.

Steganography: Review Study & Comparative Analysis

Navneet Kaur
Dept. of Computer Science and Engineering,
SBS State Technical Campus,
Ferozepur (Punjab), India
navneet18kaur@gmail.com

Sunny Behal
Dept. of Computer Science and Engineering,
SBS State Technical Campus,
Ferozepur (Punjab), India
sunnybehal@rediffmail.com

***Abstract*—Steganography is the science that involves communicating secret data in an appropriate carrier which is used from hundreds of years. The communication can be kept secret by embedding the message into carrier or by special tools such as invisible ink, microdots etc. In digital age new techniques of hiding the data inside the carrier are invented which are known as digital steganography. Nowadays, the carrier of message can be an image, audio, video or text file. In this paper we provide the comparison of available steganography techniques to researchers for the motive of in-depth study of the subject. The literature review of subject will helps to develop new techniques by analyzing and comparing the existing techniques.**

Keywords— Steganography, Cryptography, Least Significant Bit, Discrete Cosine Transform, Discrete Wavelet Transform.

I. Introduction

Steganography is the art and science of covered writing (hide in plain sight) and its techniques are in use from hundreds of years. Digital Steganography is the technique of securing digitized data by hiding it into another piece of data. Today, in digital age the easy access to any form of data such as audio, videos, images and text make it vulnerable to many threats [1]. The data can be copied for purpose of copyright violation, tampered with or illegally accessed without the knowledge of owner. Therefore, the need of hiding secret identification inside different types of digital data is required such that owner can prove copyright ownership; identify attempts to tamper with sensitive data and to embed annotations. The main task of the field of steganography is the storing, hiding, and embedding of secret data in all types of digital data. The main goal of steganography is to communicate securely in a completely undetectable manner [2] such that no one can suspect that it exist some secret information. Unlike cryptography, which secures data by transforming it into another unreadable format, steganography makes data invisible by hiding (or embedding) them in another piece of data [3]. Thus cryptography is science of overt secret writing while steganography as covert secret writing. The cover, host or the carrier is the target media in which information is hidden so that other person will not notice the presence of the information. The modified cover, including the hidden data, is referred to as a *stego-object* which can be stored or transmitted as a message [4]. The secret information can be embedded in various types of covers. If information is embedded in a cover text (text file), the result is a stego-text object. Similarly, it is possible to have cover audio, video and image for embedding which result in stego-audio, stego-video and stego-image respectively. Nowadays, the combinations of steganography and cryptography methods are used to ensure data confidentiality [5] and to improve information security. Table 1 shows the comparison of various secret communication techniques used nowadays.

TABLE I. Comparison of Secret Communication Techniques

Secret Communication Techniques	Confidentiality	Integrity	Un-removability
Encryption	Yes	No	Yes
Digital Signatures	No	Yes/No	No
Steganography	Yes/No	Yes/No	Yes

The proposed paper provide a systematic survey of existing Steganography research by categorizing existing methods according to the certain features and analyzing the advantages of these features. The motive of the paper is to provide researchers with in-depth study of subject. The proposed paper is organized in 4 sections. Section 2 deals with the classification of various types of steganography. The section 3 demonstrates the comparative study of existing steganography methods. At last, we summarize the paper in section 4.

II. Classifications of Steganography

Figure 1 represents of the generic embedding and decoding process in steganography. In the process, a secret message is embedded inside a cover object to produce the stego-object which will extracted by target person with help of the secret key. Without key the message cannot be decoded therefore the key acts as a security mechanism during communication. The first step of embedding information is achieved by passing both the secret message and the cover object into the encoder. Then encoder selects one or several protocols to embed the information into the cover object. The selection of protocols

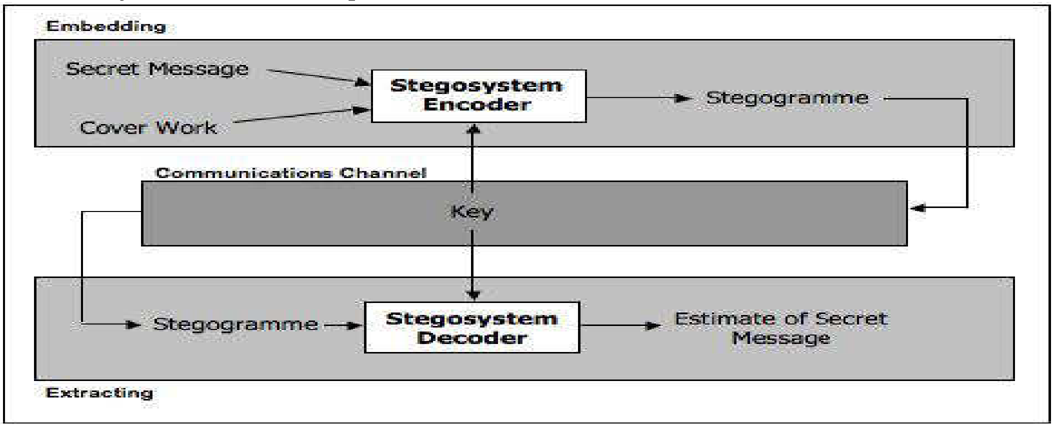


Fig. 1 Block diagram of Steganography process

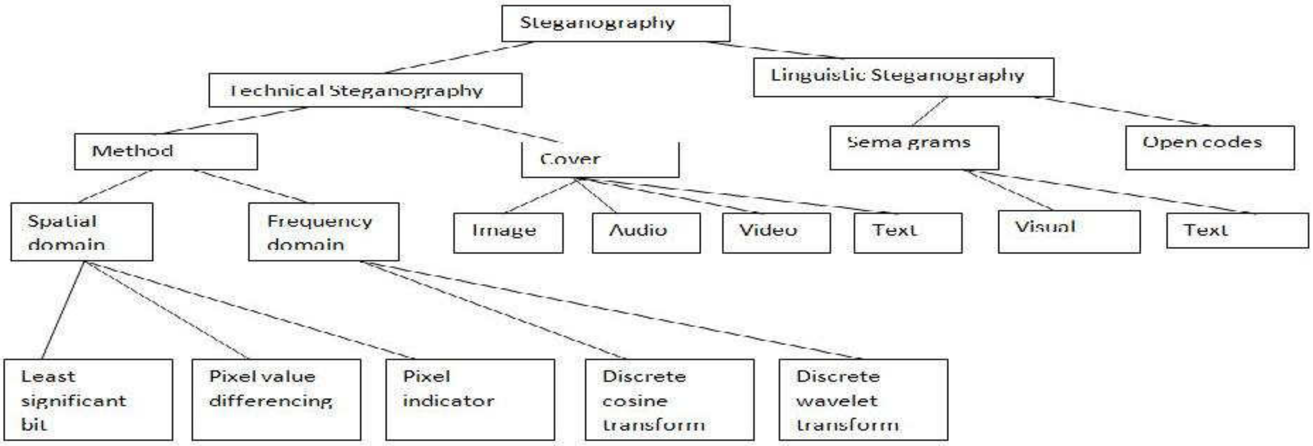


Fig. 2: Block diagram of Steganography process

depends on information to be sent and embedded object properties. For example, one can use an image based protocols to embed information inside images. The stego-object then sent to destination through some communication channel. At the other side, a secret key is required to retrieving back the message that has been embedded inside the stego-object. That key is checked against the stego-object before the decoding process. If the key is matched, only then the original message can be retrieved. Figure 2 shows a common taxonomy of steganographic techniques which are going to be discussed in this paper.

A. Linguistic Steganography: Linguistic technique is used to hide the message within the cover text in non-obvious way such that the presence of message is imperceptible to an outsider [6]. It is divided into two types:

Semagrams: It uses only symbols and signs to hide the information. It is further categorized into two ways:

i. Visual Semagrams**:** A visual semagram uses physical objects used every day to convey a message. For example: the positioning of items on a particular website.

ii. *Text Semagrams***:** This type is used to hides a message by modify the appearance of the carrier text, or by changing font size and type, or by adding extra space between words and by using different flourished in letters or handwritten text.

B. *Open Code*: In this approach the message is embedded in legitimate paraphrases of cover text in the way such that it appears not obvious to an unsuspecting observer. It can be achieved by two ways viz., Jargon which is understood only by a group of peoples and Cipher which uses some concealed ciphers to hide a message openly in the carrier medium. A subset of jargon codes can be cue codes, where pre-arranged phrases convey certain meaning.

i. *Technical Steganography*: Technical steganography uses special *tools, devices or scientific* methods to hide a message. In this type one can use invisible ink, microdots, computer based methods or various hiding places to keep message secret.

ii. *Cover*: The cover object is the carrier of the message which can be an image, video, audio, text, or some other digital media [7]. The information is encoded by changing various properties of cover object.

C. *Steganography:* In this approach the cover text is produced by generating random character sequences, changing words within a text, using context-free grammars or by changing the formatting of an existing text to conceal the message. The cover text generated by this approach can qualify for *linguistic steganography* if text is linguistically-driven. Although these text-based methods has its own unique characteristics for cover text but suffers from various problems from both a linguistic and security standpoint [8] [9] [10].

D. *Image Steganography*: This Steganography technique is more popular in recent years than other steganography possibly because of the flood of electronic image information available with the advent of digital cameras and high-speed internet distribution. In Image steganography we can hide message in pixels of an image. An image steganographic scheme is one kind of steganographic systems, where the secret message is hidden in a digital image with some hiding method [10]. Someone can then use a proper decoding procedure to recover the hidden message from the image. The original image is called a cover image in steganography, and the message-embedded image is called a stego image [11] [2]. For example: It can involve the concealing of information in the naturally occurred noise within the image and thus provides a good cover from the observers. Noise refers to the imperfections inherent in the process of rendering an analog image as a digital image. The process of hiding the information in image steganography involves:
   i) Data Hiding: First we need to select a novel image steganography technique to hide the message inside the image. Researchers have been proposed various techniques in this field as given in [12].
   ii) Data Embedding: Then we select the cover images for hiding the information. The selection of image/s depends upon the selection of technique and amount of data to be hiding. This process is known as data embedding. To ensure the integrity and confidentiality of the data a secret key is used to encode the information in stego-image.
   iii) Data Extracting: A secret key is required to retrieving back the message that has been embedded inside the stego-image. Without the secret key, the data cannot be retrieved from the image. Therefore the key is checked against the image before the decoding process. If the key is matched, only then the original message can be retrieved.

E. *Audio Steganography*: Audio steganography allows the hiding of messages in audio files by targeting the "noise" present in it. It also provides good cover as humans can't distinguish between audio frequencies much clearly. It is emerging area of steganography that relies on using an audio source as a space to hide information [8] [13].

F. *Video Steganography*: Video Steganography is a new field in the area of multimedia. It helps to embed data in the video for the purpose of identification, annotation and copyright. These techniques are used in variety of applications, such as broadcast monitoring, video authentication, copyright protection etc [8].

Method: In spatial domain, images are represented by pixels. Simple watermarks could be embedded by modifying the pixel values or the least significant bit (LSB) values [14]. It directly loads the raw data into the image pixels. Some of its algorithms are LSB, SSM Modulation based technique.

Spatial Domain: In this technique only the least significant bits of the cover object is replaced without modifying the complete cover object. It is a simplest method for data hiding but it is very weak in resisting even simple attacks such as compression, transforms [8].

i) *Least Significant Bit (LSB):* This is the most common, simple approach for embedding data in a cover image. The least significant bit (8th bit) of one or all of the bytes inside an image is changed to a bit of the secret message. When we use 24-bit image, three color bits components are used which are red, green, blue, each byte store 3 bits in every pixel. An 800 × 600 pixel image, can thus store a total amount of 1,440,000 bits or 180,000 bytes of embedded data. For example: a grid for 3 pixels of a 24-bit image can be as follows:

   (00101101 00011100 11011100)
   (10100110 11000100 00001100)
   (11010010 10101101 01100011)

   When number 200 is embedded into the least significant bits of this part of the image in binary form, the resulting grid will be as follows:

   (00101101 00011101 11011100)
   (10100110 11000101 00001100)
   (11010010 10101100 01100011)

   The number was embedded into the first 8 bytes of the grid, only the 3 underlined bits needed to be changed according to the embedded message. On average, only half of the bits in an image will modified to hide a secret message using the maximum cover size. Since there are 256 possible intensities of each primary color, changing the LSB of a pixel results in small changes in the intensity of the colors. These changes cannot be visible by the human eye due to the message hidden. In these consecutive bytes of the image data – from the first byte to the end of the message – are used to embed the information. Commonly LSB makes use of BMP images, since they use lossless compression. But now for other image file formats, LSB has also been developed [4].

ii) *Pixel Value Differencing:* It provides both high embedding capacity and outstanding imperceptibility for the stego-image; this segments the cover image into non overlapping blocks containing two connecting pixels and it modifies the pixel difference in each pair for data embedding [8].

iii) *Pixel Indicator*: This method gives the stego images of better quality than the traditional method while maintaining a high embedding capacity and it also uses concept of hiding the data using the difference between the pixel values [14].

A) *Frequency Domain*: It is more complex way than traditional approach to hide information in an image. Transformations are used on the image to hide information. Transform domain embedding can be termed as a domain of embedding techniques in frequency domain; where image is represented in terms of its frequencies.

i) *Discrete Cosine Transformation:* These methods convert the uncompressed image into JPEG compressed type [7]. It is based on data hiding used in the JPEG compression algorithm to transform successive 8 x 8-pixel blocks of the image from spatial domain to 64 DCT coefficients each in frequency domain [8]. The main advantage of this method is its ability to minimize the block like appearance resulting when boundaries between the 8 x 8 sub-images become visible (known as blocking artifact).

ii) *Discrete Wavelet Transformation*: This is the new idea in the application of wavelets; where information is stored in the wavelet coefficients of an image, instead of changing bits of the actual pixels. It splits the signal into set of basic functions. There are two types of wavelet transformation: one is continuous and other is discrete [15]. It also performs local analysis and multi-resolution analysis. DWT transforms the object in wavelet domain and then processes the coefficients and performs inverse wavelet transform to show the original format of the stego object [16].

III. COMPARISON OF DIFFERENT STEGANOGRAPHIC TECHNIQUES

Steganography Techniques	Cover Media	Embedding Technique	Advantages	Disadvantages
Mahato, *et al*., [17]	Text	Inside the source code of HTML.	Less suspicion because HTML tags and attributes are used to hide the information which is checked rarely by anybody.	Higher time complexity of the algorithm.
Geetha, *et al*.,[18]	Image	LSB: 262,144 bits, Canny edge detection method, Gaussian filter, Multiple Error Replacement, Variable Embedding Ratio.	Maintain visual quality by storing information at edges of the image which is insensitive to human eyes with highest embedding capacity.	
Shirali-Shahreza *et al*.,[19]	Text	Uses English words which have different terms in British English and American English.	Less suspicious	Easy to break once idea is exposed to third party.
Samidha and Agrawal[20]	Image	Physical location and, intensity value of pixels are considered for randomly selection of pixels	Provide good security as pixels are selected from image randomly in which information is stored randomly selected bits.	
Thenmozhi and Chandraseka ran [21]	Image	Henon Map and 2-D DWT.	More random and non-predictable, provides high data capacity and good invisibility cover	
Sui and Luo [22]	Text	Inside the source code of HTML.	Achieves high efficiency and security by changing the case of letters in the tags.	It is difficult to develop secure system using HTML because it is still immature in theory and applications.
Shirali-Shahreza [23]	Audio	Adaptive wavelet domain steganography, LSB	Imperceptible from original audio and achieves High capacity at lower error rate.	
Chae and Manjunath [24]	Video	Block DCT, MPEG-2 coding	Higher rate of data embedding and robustness to motion compensated coding like MPEG-2.	Prone to attacks as data is embedded frame-by-frame, higher computational cost.
Talip, *et al*., [25]	Text	Hybrid approach - Crypto-	Useful for the local community	The SKT and CCM

		Steganography.	to protect message from compromised.	tables need to be improved.
Mare, *et al.*,[26]	Image	Smart LSB pixel mapping.	Size of jump table for extraction is reduces and leaves more space for secret data, Higher security.	Jump table cannot be store in nosy areas.

IV. CONCLUSION

This paper provides literature review on steganography techniques. As steganography becomes widely used in computing, there are issues that need to be resolved. There are a wide variety of different techniques with their own advantages and disadvantages. We surveyed various types of steganography in this paper. We studied the various techniques, Least Significant Bit (LSB), DCT (Direct Cosine Transform), DWT (Discrete Wavelet Transform) which helps to improve the security.

REFERENCES

[1] D. Artz, "Digital steganography: hiding data within data," in IEEE Internet Computing, vol. 5, pp. 75-80, Jun 2001.

[2] M.M., Amin. M. Salleh, S. Ibrahim, M.R. Katmin, M.Z.I, Shamsuddin, "Information hiding using steganography," in Proc. 4th National Conference on Telecommunication Technology, 2003.

[3] S. Channalli, A. Jadhav, "Steganography An Art of Hiding Data," in International Journal on Computer Science and Engineering, vol. 1(3), pp: 137-141, 2009.

[4] Morkel, Tayana, J.H.P. Eloff, and M.S. Olivier. "An overview of image steganography." ISSA. 2005.

[5] Y.Y. Chung, F.F. Xu, F. Choy, "Development of video watermarking for MPEG2 video," in IEEE Region 10 Conference, TENCON, pp: 1 - 4, 2006.

[6] N. Singh, B.S. Bhati and R. S. Raw. "Digital image Steganalysis for computer forensic investigation," in Computer Science and Information Technology (CSIT), pp: 161-168, 2012.

[7] A.L. Shatnawi, M. Atallah and B.M. AlFawwaz, "An Integrated Image Steganography System with Improved Image Quality," in Applied Mathematical Sciences, vol. 7, pp: 3545-3553, 2013.

[8] Bhattacharyya, Souvik and G. Sanyal, "A Robust Image Steganography using DWT Difference Modulation (DWTDM)," in International Journal of Computer Network & Information Security, vol. 4, pp: 27-40, 2012.

[9] K. Bennett, "Linguistic Steganography: survey, analysis, and robustness concerns for hiding information in text," in Center for Education and Research in Information Assurance and Security, 2004.

[10] H. Singh, P.K. Singh, K. Saroha, "A Survey on Text Based Steganography" in Proc. of the 3rd National Conference; INDIACom-2009 Computing For Nation Development, 2009.

[11] N. Jain, S. Meshram and S. Dubey, "Image Steganography Using LSB and Edge–Detection Technique," in International Journal of Soft Computing and Engineering (IJSCE), pp: 2231-2307, 2012.

[12] R. Ibrahim and T.S. Kuan, "Steganography Algorithm to hide secret message inside an Image," in Computer Technology and Application journal, pp: 102-108, 2011.

[13] Jenkins, Neil, and J. E. Martina, "Steganography in audio," University of Cambridge CST Part II Dissertation, 2009.

[14] S.A. Sohag, M.K. Islam, M.B. Islam, "A Novel Approach for Image Steganography Using Dynamic Substitution and Secret Key," American Journal of Engineering Research, vol. 02, pp: 118 – 126, 2013.

[15] S. Rubab and M. Younus, "Improved Image Steganography Technique for Colored Images using Wavelet Transform," in International Journal of Computer Applications, vol. 39, pp:29-32, 2012.

[16] B.B. Gupta and S.K. Bandvopadhvay, "A DWT Method for Image Steganography," in International Journal, vol. 3, 2013.

[17] S. Mahato, D.K. Yadav, and D.A. Khan, "A modified approach to text steganography using hypertext markup language," in third international conf. on Advanced computing and communication technologies, pp: 40-44, 2013.

[18] C.R Geetha, S. Basavaraju, and C. Puttamadappa, "Variable load image steganography using multiple edge detection and minimum error replacement method," in IEEE conf. on Information & communication technologies, pp: 53-58, 2013.

[19] M.H. Shirali-Shahreza and M. Shirali-Shahreza, "A new synonym text steganography," in International conf. on Intelligent Information Hiding and Multimedia Signal Processing, pp: 1524 – 1526, 2008.

[20] D. Samidha, and D. Agrawal, "Random image steganography in spatial domain," in International conf. on Emerging trends in VLSI, Embedded System, Nano electronics and Telecommunication System, pp: 1-3, 2013.

[21] S. Thenmozhi, and M. Chandrasekaran, "A novel technique for image steganography using nonlinear chaotic map," in 7th international conf. on intelligent systems and control, pp: 307-311, 2013.

[22] X. G. Sui, and H. Lilo, "A new .steganography method based on hypertext," in Proc. Radio Science Conf., pp:181-184, 2004.

[23] S. Shirali-Shahreza, and M.T. Manzuri-Shalmani, "Adaptive Wavelet Domain Audio Steganography with High Capacity and Low Error Rate," in International Conference on Information and Emerging Technologies, pp: 1 - 5, 2007.

[24] J.J. Chae, B.S. Manjunath, "Data Hiding In Video," in Proc. International Conf. on Image Processing, pp: 311 – 315, 1999.

[25] M. Talip, A. Jamal, and G. W. Qiang, "A proposed steganography method to uyghur script," in International Conf. on Cyberenabled Distributed Computing And Knowledge Discovery, pp: 125-128, 2012.

[26] S. F. Mare, M. Vladutiu, and L. Prodan, "High capacity steganographic algorithm based on payload adaptation and optimization," in 7th IEEE international symposium on Applied computational intelligence and informatics, pp: 87-92, 2012.

Study of JPEG Compression with Different DCT Methods

Padma Devi
Department of ECE
CTIEMT, Jalandhar
padma.devi6@gmail.com

Ramanand Harijan
Department of ECE
BMGI, Gurgaon
ramanand332@gmail.com

Harpal
Department of ECE
SGIET, Gurgaon.
er.harpaltanwar@gmail.com

Gagandeep Singh Walia
Department of ECE
CTIEMT, Jalandhar.
ergaganwalia2@gmail.com

***Abstract*-In these days, as the size of images are increasing, compression has become very important ad common way to minimize the amount of storage necessary for images. The purpose of image compression is to reduce the size to achieve savings in storage and communication cost and at the same time, to keep most of the information contained in the original image. In JPEG standard which is frequently used for still images, Discrete Cosine Transform (DCT) is most popular lossy technique used to decompose the image into low frequency and high frequency coefficients. This allows further image compression operation to occur. In this paper we compare different methods of calculating DCT having different complexity and different number of resources used. Out of them using an early detection algorithm in an distributed arithmetic architecture can achieve reduction in computation and speed up the image compression more efficiently.**

Keywords: Image compression, JPEG Encoder, 2D-DCT, compression artifacts.

I. INTRODUCTION

Image compression is an important topic in the digital world. The need of image compression has grown steadily, and compression techniques have been studied for years and will continue to improve. Image compression technique can be broadly classified into two categories: lossless and lossy. In lossless method, the exact original data can be recovered while in lossy scheme a close approximation of the original data can be obtained. As the compression ratio for lossless and lossy compression are around 3:1 and 100:1 respectively, lossy image compression techniques are used in most application these days [7]. Image compression has been and continuous to be crucial to growth of multimedia computing and also it is the natural tech for handling the increased spatial resolution of today's imaging sensors and evolving broadcast television. The main purpose of image compression is to reduce the size of the representation and at the same time, to keep most of the information contained in the original image.

The JPEG image compression standard [3] was developed by Joint Photographic Expert Group and is frequently used for still image. The standard is very flexible and many algorithms can be used for both still images and video application. Discrete cosine transform (DCT) proposed by Ahmed et al in 1974 [1] is the basis for many image and video compression algorithm, especially the still image compression standard JPEG in lossy mode [3]. DCT is an orthogonal transform consisting of a set of basic vectors that are sampled cosine functions. The performance of the DCT is very similar to the Karhunen_loeve transform (KLT) for highly correlated data [1]. DCT transform the information from time or space domain to the frequency domain, such that we can achieve several application goals like memory saving, compact representation and fast transmission.

II. IMAGE COMPRESSION

The term Compression means the process of reducing the amount of data required to represent a given quantity of information. Image Compression main aims to reduce the number of bits required to represent an image by removing the redundancies which makes it one of the most useful and commercially successful technologies in the field of Digital Image Processing .Three principle types of data redundancies that can be identified are:

A. Coding redundancy:

Coding redundancy consists of variable length code words selected as to match the statistics of the original source. In the case of Digital Image Processing, it is the image itself or the processed version of its pixel values. Examples of image coding schemes that explore coding redundancy are the Huffman codes and the Arithmetic coding technique.

B. Spatial redundancy:

Spatial redundancy is sometimes called Interframe redundancy, Geometric redundancy or Interpixel redundancy. Here, because the pixels of most 2-D intensity arrays are correlated spatially that is each pixel is similar to or independent on neighboring pixels, information is unnecessarily replicated in the representation of the correlated pixels. Examples of this type of redundancy include Constant area coding and many Predictive coding algorithms

C. Irrelevant information:

Most 2-D intensity arrays contain information that is ignored by the human visual system.Image and video compression techniques aim at eliminating or reducing any amount of data that is psychovisually redundant. Most of the image coding algorithms in use today exploit this type of redundancy, such as the discrete cosine transform based algorithm at the heart of the JPEG encoding standard.

The usual steps involved in compressing an image are specifying the rate and distortion parameters for the target image, dividing this image data into various classes, based on their importance. Then dividing the available bit budget among these classes, such that the distortion is minimum. Next step involves quantizing each class separately using the bit allocation information followed by encoding of each class using an entropy coder and write to the file

III. COMPRESSION ARTIFACTS

When an aggressive data compression technique is applied to an image it results in compression artifact that removes some information that cannot be stored in the available data-rate because of its complexity, or the algorithm that is used may determined the data incorrectly to be of little subjective importance but which is objectionable to the user. A compression artifact is a particular class of data error that is usually the consequence of quantization in lossy image compression. Artifacts are often as a result of the latent errors inherent in lossy data compression. Some common artifacts are:

A. Blocking Artifacts:

It is a distortion or discontinuities due to the presence of high frequency components at the block boundary that appears in compressed image and result in abnormally large pixel blocks. It is also called "macro blocking", it occurs when the encoder is allocated with the low bandwidth. In lossy compression, as higher the compression rate, the more the content is removed.

B. Colour Distortion:

Colour distortion occurs mostly in uniform colour areas. Human eyes are not as sensitive to colour as to brightness, in image compression mostly the detailed chromium colour information is disposed, while retaining luminance colour information. This process is called "chroma sub sampling", which means that the colour image is split into a brightness image and two colour images. The brightness (luma) image is stored at the original resolution, whereas the two colour (chroma) images are stored at some lower resolution. The compressed images look slightly dim, with less brilliant colour.

C. Ringing Artifacts:

In digital image compression, ringing artifacts appear as spurious signal ("rings") near sharp transitions in a signal. In normal viewing, they look like "rings" near the edges. They occurred in low bit rates image coding. Due to the band limited signals in frequency domain that passes through a low pass filter it result in ringing artifacts. Whereas in the time domain, ringing artifacts are caused by the ripple in the sinc function. It is also called Gibbs phenomenon.

D. Blurring Artifacts:

Blurring artifacts are diffuse parts in images. Blurring of an image generally caused due to the attenuation of high frequency particles. Blurring makes the image smoother than originally. It occurs in low bit rate image coding.

IV. THE PROCESS FLOW OF JPEG IMAGE COMPRESSION

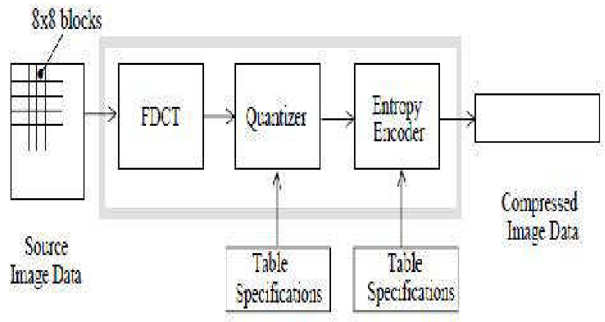


Fig.1.DCT based encoder

The following are the general overview of the JPEG compression process:

A. The image is partitioned into non-overlapping blocks of 8 × 8 pixels and the pixel values are converted to signed integer format from unsigned integer format.

B. For colour images colour space conversion is performed by transform the RGB input image into a luminance and chromium space colour such as YCbCr. After colour space conversion expansion and down sampling is performed to reduce the sampling rate of colour information (Cb and Cr) because chrominance component are less important to human eye [5].

C. DCT is applied to each block of pixel using a row column decomposition technique.

D. After DCT, the image is described in terms of its frequency domain in great details. Since human eye cannot pick out any changes in very bright and very dim colour image can be compressed further by rounding the changes in frequency that are distinguishable to the eye to zero. Each 8 × 8 block is compressed through quantization which discards the 2-D DCT high frequency and small amplitude coefficients.

E. The coefficients in the array of compressed block are stored in a drastically reduced amount of space by encoding DCT coefficient in the zigzag sequence as shown in fig.2.

F. The entropy encoding uses differential coding, run length coding (RLE), huffman coding and variable length coding to decrease the number of bits used to represent the image [3].

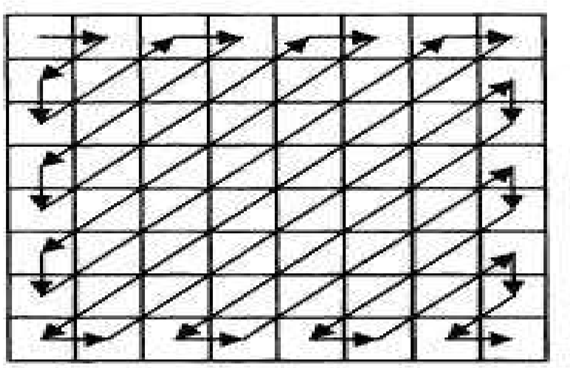


Fig.2. zigzag order of 8 × 8 block

V. 2D-DCT OVERVIEW

Equation of 2D-DCT

$$F(u,v) = \frac{C(u)C(v)}{4}\sum_{j=0}^{7}\sum_{k=0}^{7} f(j,k) os\left[\frac{(2j+1)u\pi}{16}\right] cos\left[\frac{(2k+1)v\pi}{16}\right]$$

$C(u) = \frac{1}{\sqrt{2}}$, for u=0, and 1 for u = 1, 2… 7

$C(v) = \frac{1}{\sqrt{2}}$, for v=0, and 1 for v = 1, 2… 7

DCT is related to the Discrete Cosine Transform (DFT). DCT based architectures have gained popularity and are mostly used in the image processing. The DCT is computationally intensive and has a high degree of computational complexity. The DCT unit is the foundation for achieving image compression by concentrating most of the image in the lower spatial frequencies. In Fast 2D-DCT algorithms, since a cosine depends only on the position of samples in the 8×8 block, their value can be pre computed and can be written as a matrix multiplication. T = C X C ' where C is the matrix of the values of the cosines and C ' is the transpose of C. X is the 8 × 8 pixel data matrix.

The direct implementation of the 1D-DCT method every 1D-DCT require 8 multiplier and 7 additions for each sample. Since there are 64 samples in an 8×8 block, the 2D-DCT will require 64×8 multiplications and 64×7 additions in ach direction. As a result it requires 2×64×8 =1024 multiplication and 896 additions for each 8×8 block [3].

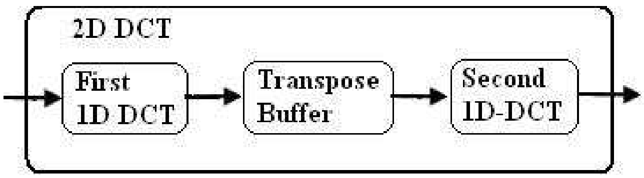


Fig.3. 2D-DCT architecture

Since the 2D-DCT is a separable function, the computations can be done in two steps as shown in fig 3, by applying a 1D DCT in one dimension, (row wise) and then applying another 1D-DCT in the other dimension (column wise) on the output of the first 1D-DCT, instead of the direct computation. This simple decomposition reduces the complexity in computing 2D-DCT coefficients by a factor of 4 [5].

VI. ANALYSYS OF VARIOUS 2D-DCT METHODS

There are several papers proposing the row column decomposition approach and high level design of 2D-DCT architecture. Among them, Agostini *et al* [5] propose a fast 2D-DCT algorithm using six stage pipelined architecture that is based on work presented in [10]. It employs row column decomposition technique to calculate the 2D-DCT by two 1D-DCT calculation and a transposition buffer. It requires 464 additions and 80 multiplications for the 2D-DCT calculation of 8×8 block. It uses Wallace tree multipliers that decomposes the multiplication in shift and add operations.

H. Osman *et al* [11] presents a JPEG Encoder architecture that requires minima FPGA resources without degrading the quality if image. They have implemented the DCT processing block such that there is no special need of a zigzag unit, hence it reduces the FPGA resources and improve performance speed. The comparison of resource utilization is given in Table 1.

Antonino *et al* [6] present a fast 2d-DCT hardware accelerator design for a FPGA based SoC. It targeted mainly to a mixed HW/SW design and has the main objective of defining the interface to/from the processor. It uses seven stage pipelined architecture and for a single pipeline it requires 19 adder/subtractor and 4 multipliers.

Table I. Comparison of resource utilization.

Resource	H. Osman [11]	A. Tumeo [12]	A.Tumeo [6]
Slice Flip-Flops	339	3450	3431
4-Input LUTs	821	5656	2618
Slices	593	2932	2823

Trainor *et al* [9] had proposed a distributed arithmetic architecture that has used parallelism and pipelining .Lee's algorithm for fast DCT computation of 8×8 block require 192 multiplications and 464 additions and the fast algorithm proposed by Cho and Lee required 96 multiplication and 466 additions for 8×8 DCT computation [3].

M. A. BenAyed *et al* [8] presents a multiplier less new technique for 1D-DCT architecture based on the combination of Arithmetic Distribution and Loeffler algorithm.

H. Osman *et al* [11] presents a JPEG Encoder architecture that requires minima FPGA resources without degrading the quality if image. They have implemented the DCT processing block such that there is no special need of a zigzag unit, hence it reduces the FPGA resources and improve performance speed.

Qihui zhang *et al* [7] proposed a distributed arithmetic and memory oriented architecture for 2D-DCT using separability property. In this the 2D-DCT is computed via two 1D-DCT and a transposition memory which transpose the result of first 1D-DCT. It uses an architecture that requires adders and registers only and no multiplier are required in 2D-DCT calculation. This architecture is based on the previous work of M.T. Sun et al [2] in which implementation of 16×16 DCT chip is presented using a concurrent architecture.

Nishida *et al* [13] proposed a zero value prediction for fast DCT calculation but this scheme result in degradation of image visual quality. Jin li *et al* [4] presents an algorithm to detect zero quantized DCT coefficients. This algorithm can be used to reduce the computational complexity by predicting the DCT coefficients as zero in advance and skip the DCT and quantization computations and hence improving the encoding efficiency for fast encoding in JPEG compression without any degradation in the quality of image.

VII. CONCLUSION

In this paper 2D-DCT algorithm and comparison between different schemes are presented. While comparing the different methods for computing DCT, the critical points are complexity of computation, number of FPGA resources used and the degradation of quality of image. From the various architecture studied in the literature we conclude that using a detection algorithm to predict zero-quantized DCT coefficients [4] in an area efficient high performance VLSI architecture for DCT based on the distributed arithmetic can be useful to get an area efficient, improved encoding efficiency, high speed operation architecture for the JPEG Encoder having minimal utilization of the FPGA resources.

REFERENCES

[1] N. Ahmed, T. Natarajan, K. R. Rao, "Discrete Cosine Transform", IEEE Transaction on computers, vol.23, pp.90-93, January 1974.

[2] M. T. Sun, T. C. Chen, and A. M. Gottlieb, "VLSI Implementation of a 16×16 Discrete Cosine Transform," IEEE Transaction on Circuits and System, vol.36, no.4, pp.610-617, April 1989.

[3] Tinku Acharya and Ping-Sing Tsai, "JPEG 2000 standard for image compression", A John Wiley & Sons Incorporation, Publication Hoboken, New Jersey, 2005.

[4] Jin Li1, Jarmo Takala, Moncef Gabbouj, and Hexin Chen "A detection Algorithm for zero-quantized DCT coefficients in JPEG" IEEE international conference on Acoustic, speech and signal processing, Las Vegas, 2008.

[5] L.V. Agostini, I.S. Silva, and S. Bampi., "Pipelined fast 2d DCT architecture for JPEG image compression", In Integrated Circuits and Systems Design, 2001, 14th Symposium on, pp. 226–231, Pirenopolis, Brazil, 2001.

[6] Antonino Tumeo, Matteo Monchiero, Gianluca Palermo, and Fabrizio Ferrandi, "A Pipelined Fast 2D-DCT Accelerator for FPGA-based SOCs", IEEE Computer Society Annual Symposium on VLSI,2007.

[7] Qihui Zhang and Nan Meng, "A Low Area Pipelined 2-D DCT Architecture for JPEG Encoder", 52nd IEEE International Midwest Symposium on Circuits and Systems, pp.747-750, 2009.

[8] M. A. BenAyed, L. Dulau, P. Nouel, Y. Berthoumieu, N. Masmoudi, P. Kadionik, and L. Kamoun, "New Design Using VHDL Description For DCT Based Circuits", proceedings of the Tenth International conference on Microelectronics, pp.87-90, monastir, Tunisia, 1998.

[9] D. W. Trainor, J. P. Heon, and R. F. Woods, "Implementation of the 2D DCT using a Xilinx XC6264 FPGA," IEEE Workshop on Signal Processing Systems - Design and Implementation, Issue , 3-5 pp.541 – 550, 1997.

[10] M. Kovac and N. Ranganathan. "JAGUAR: A Fully Pipeline VLSI Architecture for JPEG Image Compression Standard". Proceedings of the IEEE, vol. 83, no. 2, pp. 247-258, 1995.

[11] Hossam Osman and Waseim Mahjoup, "JPEG Encoder for Low-Cost FPGAs", International Conference on Computer Engineering & System, pp 406-411, Cairo, Nov' 2007.

[12] Antonino Tumeo, Matteo Monchiero, Gianluca Palermo, Fabrizio Ferrandi and Donatella Sciuto, " A self- Reconfigurable Implementation of the JPEG Encoder", IEEE International Conference on Application specific Systems Architectures and Processors, pp.24-29, Montreal, July 2007.

[13] Y. Nishida, K. Inoue and V. G. Moshnyaga, "A Zero-Value Prediction Technique for Fast DCT Computation," IEEE Workshop on Signal Processing System, pp.165-170, Aug. 2003.

A Review on Image Enhancement Techniques

Geetika Mahajan
Department of CSE
Sai Institute of Engineering and Technology
Amritsar (Punjab), India
mahajangeetika89@gmail.com

Aman Arora
Department of CSE
Sai Institute of Engineering and Technology
Amritsar (Punjab), India
aman_study@yahoo.com

***Abstract* —This paper has focused on the review of the several vision enhancement methods. Digital image enhancement has proven to be one of the most significant image applications; since it has capability to enhance the visibility of images. It enhances the quality of poor pictures. Distinctive procedures have been proposed so far for improving the quality of the digital images. To enhance picture quality image enhancement can specifically improve and limit some data presented in the input picture. It is a kind of vision system which reductions picture commotion, kill antiquities, and keep up the informative parts. Its object is to open up certain picture characteristics for investigation, conclusion and further use. The main objective of this paper is to explore and discover the limitations of the existing image enhancement strategies.**

***Keywords*—Image enhancement, Histogram equalization, Discrete Wavelet Transform.**

I. INTRODUCTION

The methods for improving the quality of digital images are known as image enhancement techniques. It is relatively simple, for example, to make an image light or dark or to enlarge or reduce contrast. Sophisticated image enhancement software also supports several filters for changing images in a variety of ways. The main purpose of image enhancement is to process a given image so that the outcome is more appropriate than the original image for a definite use.

It sharpens image features such as edges, boundaries, or contrast to build a graphic display more useful for display and analysis. The enhancement doesn't raise the inbuilt information content of the data, but it increases the active range of the selected features so that they can be detected simply. Image enhancement methods can be based on either spatial or frequency domain techniques.

A. Spatial domain methods:

In the spatial domain method, the pixel composing of image facts are measured and the different procedures are directly applied on these pixels. The image processing functions in the spatial domain may be expressed as

$$G(x,y) = T[f(x,y)] \qquad (1)$$

Where $f(x, y)$ is the input image, $G(x, y)$ is the processed output image and T represents an operation on 'f' defined over some neighborhood of (x, y). Sometimes T can also be used to operate on a set of input images.

Fig. 1. Image (a) before (b) after enhancement

i. Histogram Equalization (HE):

It is a great point processing enhancement method that seeks to optimize the contrast of an image at all points. It advances image contrast by destruction or equalizing the histogram of an image. A histogram is a table that basically counts the number of times a value appears in some data set.

Fig. 2. (a) Original Image (b) Result of HE [3]

Histogram equalization uses Cumulative Distribution Function (CDF) as the research table. For example, for an N-bit image, histogram h, normalized CDF is given by:

$$\hat{C}j = \sum_{i=0}^{j} \hat{h}i. \; j \quad \{0,1,\ldots 255\} \qquad (2)$$

CDF gives that what proportion of samples in an image are equal to or less than value j. Normalized CDF must be rescaled to [0,255] and is then used as the research table. It increases the monotonocity. Slope of CDF is vertical where there is a group of information in the source and is horizontal where there is little information in the source. CDF of completely equalized image is straight line with slope 1.

Histogram equalization is the straight forward method used to accomplish enhanced quality images in black and white color scale in different application areas such as medical image processing that includes X-ray ,MRIs and CT scans ,object tracking ,speech recognition etc. The chief benefit is that it is easy and efficient. The main two disadvantages are: the destruction property; not often utilized in purchaser electronics products such as TV because it may extensively change the original brightness and cause adverse artifacts.

ii. Bi-Histogram Equalization (BHE):

The purpose of the bi-histogram equalization is to conserve the mean brightness of a certain image. The input image is decomposed into two sub-images based on their means and the resulting equalized sub-images are enclosed by each other about the input mean. In hardware implementation, this method requires additional complex hardware than the typical Histogram Equalization (HE). For effective use of this technique, an attempt to decrease the difficulty should be ended. Many applications can be made achievable by utilizing this technique in the field of consumer electronics such as TV, VTR (Video Tape Recorder), Camcorder.

iii. Dualistic Sub-image Histogram Equalization (DSIHE):

A histogram has been separated into sub-histograms based on their median values such that one include high intensity pixels and another include low intensity pixels and equalize each part separately with HE technique. At last, we obtain the outcome after the processed sub-images are composed into one image. HE can modify the original brightness and cause adverse artifacts. BHE and DSIHE conserve the original brightness. There are images that are not well enhanced by BHE and DSIHE, they need more conservation.

iv. Recursive Mean Separate Histogram Equalization (RMSHE):

It recursively decomposes the input histograms based on their respective means. It is the more usual enhancement and improves the brightness conservation. In other words, it is the simplification of HE and BHE in conditions of brightness conservation.

v. Gain Controllable Clipped Histogram Equalization (GC-CHE):

Gain Controllable Clipped Histogram Equalization managed noise extension as well as conserving the original brightness of the image. It is an explanation of BHE and RMSHE methods. The clipping rate is adaptively prohibited to enhance the contrast with conserving the mean brightness. Based on the clipping level, histogram of an image is

clipped and the clipped segment is then re-distributed to the inclusive active range by locally regulating the clipping gain.

Fig. 3. Results for jar: (a) Original image(b)Enhanced image by applying HE methods(c)Enhanced image by applying BHE method(d)Enhanced image by applying DSIHE method [5]

Though enhancing the contrast of a low light-level image, the contrast elevation is familiar to solve the noise extension problem according to the input image and balance contrast using gain control methods.

B. Frequency domain methods:

Frequency domain methods are based on modifying the spectral transform of an image. It transforms the image to its frequency representation. It also computes inverse transform back to the spatial domain. The high frequencies correspond to pixel values that change rapidly across the image (e.g. text, texture, leaves, etc.) and strong low frequency components correspond to large scale features in the image (e.g. a single, homogenous object that dominates the image).

i. Adaptive Intensity Transfer function

Adaptive intensity enhancement uses a particularly calculated nonlinear transfer function which is able of dropping the intensity of bright regions and at the same time enhancing the intensity of dark regions.

ii. Discrete Wavelet Transform (DWT)

It is the realization of the wavelet transform with discrete set of the wavelet scales and translations obeying few definite rules. In other words, this transform decomposes the signal into equally orthogonal set of wavelets which is the key distinction from the CWT (Continuous Wavelet Transform)**.**The information provided by CWT is extremely disused as far as the renewal of the signal is concerned and requires a considerable amount of computation time and resources. The DWT on the other hand, provides satisfactory information both for analysis and synthesis of the original signal with a significant decrease in the computation time. The DWT is easier to implement as compared to CWT.

iii. Dominant Brightness level analysis

Contrast enhancement method is based on the dominant brightness level analysis and adaptive intensity transform for remote sensing images. This algorithm computes the brightness- adaptive intensity transform function by means of the low frequency luminance element in discrete wavelet domain transform and transfer intensity values according to the transfer function. In this method discrete wavelet transform (DWT) decompose input image into set of band called HH, HL, LH and LL. LL sub band has explanation of low-,middle-,high-intensity layers by the log average luminance. Adaptive intensity transfer function estimation is done using the knee transfer function and gamma adjustment function is based on the dominant brightness level of each layer. After this process the enhanced image is obtained by using inverse DWT (IDWT).

iv. Contrast Enhancement

The output of the resolution enhancement module will be taken as the input to this module. The key steps included in this process are as follows: First we will apply DWT on the input image and in parallel we improve its contrast using GHE and get its DWT. Now determine the hanger (U), aligner (V) and singular value matrix (SVM) for the LL sub-bands obtained above. Find the greatest element in both the SVMs and take their ratio ( ). Now determine the new STM and estimate the new LL sub-band. The estimated LL sub-band and the HF components of real input image are now used to re-produce the contrast enhanced image.

C. Gamma Correction

Gamma defines the brightness of a computer display. It determines how bright the output of the display will be. Therefore, it is used to change the output levels of a monitor. While the gamma setting affects the brightness of a display, it is not identical to the brightness. This is because gamma adjustments are not linear, like brightness levels are.

Fig. 7. (a) Image before gamma correction (b) Image after gamma correction

Instead, the gamma setting applies a function to the input levels, which produces the final output level. You can imagine this function as a curved line instead of a straight one. Therefore, the gamma setting affects both the brightness and the contrast of the display. The reason gamma correction is used is because the input signal, or voltage, sent to a monitor is not sufficient to create a bright image. Therefore, if the gamma is not changed, the images on your screen would be dark and hard to see. By applying gamma correction, the brightness and contrast of the display are enhanced, making the images show brighter and more natural looking.

D. Remote Sensing Images

Remote sensing images are representation of parts of the earth surfaces as seen from space. Remote sensing means the sensing of the earth's surface from space by utilization of the properties of electromagnetic waves emitted, reflected or diffracted by the sensed objects for the reason of improving natural resources management, land use and the safety of the environment. It is the gain of information regarding an object without being in physical contact with it.

Fig. 8. Remote sensing image [7]

II. LITERATURE SURVEY

A novel contrast enhancement approach [1] based on dominant brightness level analysis and adaptive intensity transformation for remote sensing images has been used. This method has effectively enhanced any low-contrast images acquired by a satellite camera and also suitable for other various imaging devices such as consumer digital cameras, photorealistic 3-D reconstruction systems and computational cameras. A straightforward and efficient color image enhancement method [2] for endoscopic images has been proposed. The proposed image enhancement method works by two interconnected steps: image enhancement at gray level and color reproduction. The proposed method has been applied on any RGB images collected from any white light endoscopic devices.

A new optical transfer function-based micro image enhancement algorithm [3] has been proposed. The optical transfer function-based micro image enhancement algorithm can generate a better micro image enhancement effects. The Laplacian pyramid of the image processing algorithm [4] for image enhancement is an important field of image processing and critical technologies. The Laplacian pyramid is ever-present for decomposing images into several scales and is commonly used for image analysis. This method always produces high-quality results in the process of image detail enhancement.

A method that enhances an image with natural contrast[5] has been proposed. The aim of contrast enhancement is achieved using parameters `a', `b', `c' and `k', where `a', `b', `c' and `k' are constants with the new extension in their range. Local contrast enhancement increases the gray level of original image on the basis of light and dark edges. This method has applied on m×n size of an original gray scale image. The local mean and local standard deviation of entire image, minimum value and maximum value of the image are used to statistically describe digital image.

A multifractal theory analysis of infrared image and the multifractal uniqueness of edges in the infrared images [6] have been proposed. The image enhancement method is more accessible and highlights the human eyes susceptible image area. The difficulty of low visibility and blurry edge has been resolved well and the enhancement image is more appropriate to human's observing.
The homomorphic enhancement [7] that can remove the control of irregular explanation in frequency domain algorithm has been proposed. They have been offered a hybrid algorithm to enhance the image.. Through the hybrid algorithm to enhance the infrared image, not only enhanced the infrared image of the details, but the outline of the image has also been smooth. Finally, the enhanced image is better than other algorithm of results. The theory of retinex and method of image enhancement [8] based on experiments and study of vision has been proposed. For X-Ray medical images, they have been improved the global and local image enhancement by Retinex algorithm.
A method including three steps: image preprocessing, image recognition and enhancement [9] have been proposed. The study of recognition and enhancement of the traffic sign for the computer generated image is very less. For recognition and enhancement of the traffic sign in the Computer generated image, the complexity is how to correctly identify and enhance the traffic sign, and maintaining other information of object do not been altered. The adopting stepwise refinement method attempts to solve the problem. The proposed idea of stepwise refinement provides a procedural reference for complex image recognition.

The dental panoramic radiography [10] is one of dental imaging. The digitized film-based image to digital image has been essential to let image enhancements in order to improve the interpretability quality of information in the image. Estimation of the preference image quality is performed based on objective principle. A method to improve the enhancement outcome with image fusion method [11] has been proposed. Some unusual evaluation methods and fusion policies are discussed and compared. They have shown that the fusion improves the enhancement results.

An industrial X-ray image enhance algorithm [12] based on histogram and wavelet has been proposed. Compared with other image enhancement algorithms, this algorithm can recover the global image contrast successfully as well as defeat the observable artifacts of X-ray image, the x-ray image turn into more clear, and an enhanced perceptual image is acquired for the image quality identify and identical.

An enhancement algorithm based on multi-scale Retinex [13] which can make up the lack of conventional wavelet algorithm. The standard and recognition methods of multi-scale Retinex and wavelet has been calculated. The panchromatic and multicolor remote sensing image enhancement has been passed out based on the two methods and remote sensing images could get superior enhancement quality, so multi-scale Retinex is a better method for sensing image enhancement.

III. GAPS IN EARLIER WORK

The survey has shown that the most of image enhancement techniques has certain limitations. Following are the main limitations in earlier work.

1. It has been found that the most of the existing techniques are based upon the transform domain methods; which may introduce the color artifacts and also may reduce the intensity of the input remote sensing image.
2. Due to transform domain methods Gaussian random noise may be presented in the output images.
3. Remote sensing and underwater images has been neglected by many researchers in earlier work.

IV. CONCLUSION AND FUTURE DIRECTIONS

Image enhancement plays a significant role in vision applications. Recently much work is done in the field of remote sensing images enhancement. Many techniques have been proposed so far for enhancing the remote sensing images. It has been found that the most of the existing techniques are based upon the transform domain methods; which may introduce the color artifacts and also may reduce the intensity of the input remote sensing image.

To overcome the short comings of the available techniques in near future we will modify the existing transform domain method using adaptive gamma correction to enhance the results further. However no implementation is considered in this work so in near future suitable simulation tool will be used.

REFERENCES

[1] Lee, Eunsung,"Contrast enhancement using dominant brightness level analysis and adaptive intensity transformation for remote sensing images."Geoscience and Remote Sensing Letters, IEEE 2013 pp: 62-66.

[2] Imtiaz, Mohammad Shamim, TareqHasan Khan, and Khan Wahid. "New color image enhancement method for endoscopic images." , International Conference on Advances in Electrical Engineering (ICAEE) IEEE, 2013.pp:234-237.

[3] Sun, Yaqiu, and Xin Yin. "Optical transfer function-based micro image enhancement algorithm." International Conference on Communications Workshops (ICC), IEEE, 2013.pp:343-348

[4] Teng, Yanwen, Fuyan Liu, and Ruoyu Wu. "The Research of Image Detail Enhancement Algorithm with Laplacian Pyramid." Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing. IEEE, 2013.pp:1024-1026.

[5] Goel, Savita, Akhilesh Verma, and Neeraj Kumar. "Gray level enhancement to emphasize less dynamic region within image using genetic algorithm." Advance Computing Conference (IACC), 2013 IEEE 3rd International. IEEE, 2013.pp:167-172.

[6] Tianhe, Yu, et al. "Enhancement of infrared image using multi-fractal based on human visual system." Measurement, Information and Control (MIC), 2012 International Conference on. Vol. 2. IEEE, 2012.pp:110-115.

[7] Zhang, Chaofu, Li-ni Ma, and Lu-na Jing. "Mixed Frequency domain and spatial of enhancement algorithm for infrared image." Fuzzy Systems and Knowledge Discovery (FSKD), 2012 9th International Conference on. IEEE, 2012.pp:198-201.

[8] Weizhen, Sun, Li Fei, and Zhang Qinzhen. "The Applications of Improved Retinex Algorithm for X-Ray Medical Image Enhancement." Computer Science & Service System (CSSS), 2012 International Conference on. IEEE, 2012.pp:121-125.

[9] Yaping, Li, et al. "The recognition and enhancement of traffic sign for the computer-generated image." Digital Home (ICDH), 2012 Fourth International Conference on. IEEE, 2012.pp:134-139.

[10] Juliastuti, E., and L. Epsilawati. "Image contrast enhancement for film-based dental panoramic radiography." System Engineering and Technology (ICSET), 2012 International Conference on. IEEE, 2012.

[11] Fang, Xiaoying, et al. "A method to improve the image enhancement result based on image fusion." Multimedia Technology (ICMT), 2011 International Conference on. IEEE, 2011.pp:765-769.

[12] Jin, Li, et al. "Industrial X-ray image enhancement algorithm based on adaptive histogram and wavelet." Strategic Technology (IFOST), 2011 6th International Forum on. Vol. 2. IEEE, 2011.pp:453-457

[13] Henan, Wu, et al. "Remote sensing image enhancement method based on multi-scale Retinex." Information Technology, Computer Engineering and Management Sciences (ICM), 2011 International Conference on. Vol. 3. IEEE, 2011.pp:342-346.

A Review of Skew Detection Approaches in Image Processing

Neha Kumari Thakur
Guru Nanak Dev University
Amritsar, India
thakur.neha262@gmail.com

Amit Chhabra
Guru Nanak Dev University
Amritsar, India
chhabra_amit78@yahoo.com

***Abstract*—Nowadays, document image processing has become an important technology in the automation of office documentation tasks. Automatic document scanners are an essential component of systems capable of those tasks. While scanning the document, skew is unavoidably introduced into the document image. This skew has a considerable effect on document analysis, document understanding, and character segmentation etc. Therefore, detecting the skew of a document image and correcting it are important issues in realizing a practical document reader. In this paper, a survey of various techniques for skew detection is presented.**

Keywords—Skew, Hough Transform,Nearest Neighbor, Cross Correlation.

I. INTRODUCTION

In this era, a large amount of information is still stored and presented through paper media, as people feel easier to receive paper documents. Even so, there is a trend to share information in electronic form. To change paper documents into electronic form, usually three steps are involved: scanning and digitization, page layout analysis, and character recognition.

The digitalization process of documents produce images rotated at small angles with respect to the original image. The skew introduced makes it difficult to visualize the images by human users. It also increases the complexity of automatic image recognition, degrades the performance of optical character recognition tools, increases the space needed for image storage, etc. Therefore, it is very difficult, to prevent skew produced in scanning operations. It is thus preferred to detect and correct the skew in the document image at the initial stage to nullify the effect of skew to the subsequent processing steps. Thus, skew correction is an important part of any document processing system. Skew in scanned document can be of two types and is shown below [1]:

1. Clockwise skew (Positive skew)

2. Anti-clockwise skew (Negative skew)

Fig. 1. a) Clockwise Skew [1]

Fig. 1. b) Anticlockwise Skew [1]

There are many approaches previously proposed for skew detection and correction in image processing. In recent times, many different methods came into existence. Some of the methods for skew detection are shown diagrammatically in the next section and are described below.

II. METHODS FOR SKEW DETECTION

There are various methods for skew detection and correction and they can be categorized into five groups:

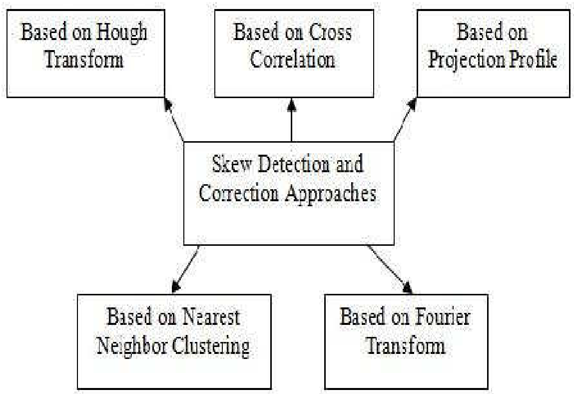


Fig. 2. Classification of Skew Detection Approaches

A. *Skew detection using Hough transform*

Kwag et al. [2] proposed a fast skew estimation and correction algorithm for English and Korean documents based on a Block Adjacency Graph (BAG) representation. BAG is one of the efficient data structures for extracting information related to connected components, the image rotation for skew correction is performed using the block information in BAG. The skew estimation algorithm uses a coarse/refine strategy based on Hough transformation of connected components in the image. The skew correction algorithm generates a non-skew image by rotating the blocks, rather than individual pixels. The approach is shown in fig. 3 below.

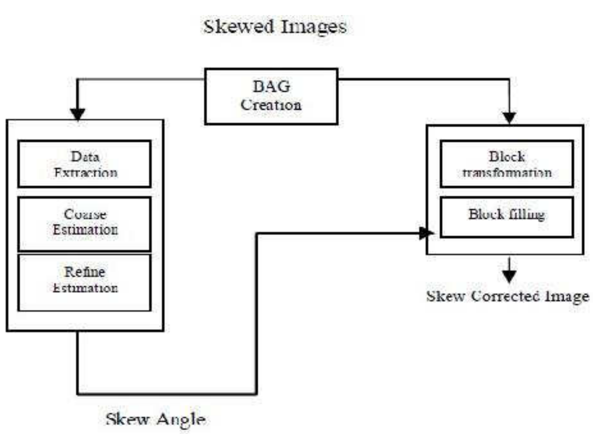


Fig. 3. Block diagram of the proposed skew estimation and correction [2]

Manjunath et al. [3] proposed a method based on thinning and Hough transform to determine the slew angle. It works in two steps. In the first step, the characters selected from the document are blocked and then, thinning is performed on the blocked region. However, the other methods based on this method fail for document images containing text with picture. The approach is shown in the fig. 4 below.

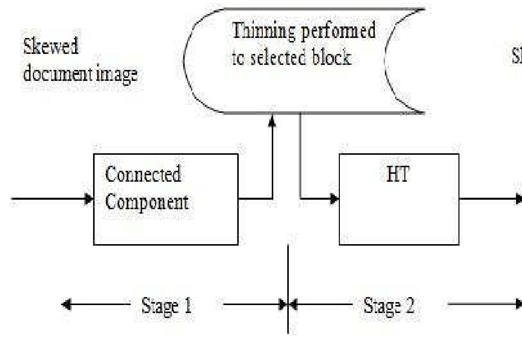


Fig. 4.Block diagram of the proposed method [3]

C. Singh et al. [4] studied that Hough transform provides a robust technique for skew detection in document images, but suffers from high time complexity which creates problem for detecting skew in large documents. The skew detection and correction process is divided into three stages: a preprocessing stage using block adjacency graph, voting process using Hough transform and de-skewing image using rotation. This algorithm is effective for documents with Roman scripts, but does not provide good results for Indian scripts. They also found that time taken for the whole process is less than one second even for large documents.

D. Kumar et al. [5] proposed a modified form of Hough transform. Here, Hough transform (HT) is applied to the set of pixels such that the total time taken by the algorithm gets reduced keeping the accuracy of the process intact. The spectrum of the HT space is divided i.e., angle of skew can be 0 degree to 45 degree into one-tenths, hence getting the portion in which the resultant skew lies. Then only that portion is further inquired by diving it into one-tenths and so on. This way the algorithm reaches the solution rapidly as compared to the classical HT.

B. *Based on cross correlation*

Chen et al. [6] proposed a modified cross correlation skew detection algorithm that uses horizontal and vertical cross correlation simultaneously to deal with vertical layout text which is commonly used in Chinese and Japanese document. Instead of calculating correlation of entire image, they randomly selected small regions to speed up the process. The region verification stage and auxiliary peaks further made the method more reliable and robust.

Lin et al. [7] described a method that uses fuzzy c-regression algorithm (FCRM) to estimate the skew angles of document images with high accuracy. It works in four steps, starting with analysis of image using run length smoothing and black and white transition to produce parallel straight lines and in the end FCRM clustering algorithm is applied to estimate the skew angle. The skew estimation is valid for various classes of documents obtained by scanning or copying including general documents, magazines, and newspapers. The system works very well for such input images. The average errors are within 0.2°.

C. Projection profile

Li et al. [8] described a novel document skew detection algorithm based on wavelet decompositions and projection profile analysis. Firstly, the skewed document images are decomposed by using the wavelet transform. The matrix containing the absolute values of the horizontal sub-band coefficients, which preserves the text's horizontal structure, is then rotated through a range of angles. It can save computation time and improve the estimation accuracy. A projection profile is computed at each angle, and the angle that maximizes a criterion function is regarded as the skew angle. The advantage of proposed algorithm is that it works well on various documents including Chinese, Japanese, English, diagrams, etc.

D. Nearest Neighbor based approach

Liolios et al. [9] proposed an improved version of nearest neighbor clustering for document skew detection. A heuristic is used which attempts to group all the connected components that belong to the same line of text, into one cluster. The larger clusters further result in better estimation of skew angle. The skew detection accuracy of this improved method is better, when compared to the classical approach, with no change in the order of complexity.

Lu et al. [10] described a method for documents skew estimation in which size restriction is introduced in the detection of nearest neighbor pairs. Then, the chains with a largest possible number of nearest-neighbor pairs are selected, and their slopes are calculated to give the skew angle of document image. The proposed method can successfully detect skew angles of different documents, without the skew angle limitation, and without the requirement of predominant text area. It can also deal with different text orientations appearing on the same image.

Iuliu et al. [11] introduced a generic, scale-independent algorithm capable of accurately detecting the global skew angle of document images within the range [-90°, 90°]. By using the same framework, they extended the algorithm for Roman script documents so as to cope with the full range [-180°, 180°) of possible skew angles. Also, the improved algorithm is very fast and requires no explicit parameters. They also performed experiments on a combined test set comprising around 110 000 real-life images which shows the accuracy and robustness of the proposed method.

E. Based on Fourier transform

Lowther et al. [12] proposed a technique that is based on calculating the average 2D Fourier transform of blocks in a document image and using the Radon transform to find the peak in the directional spectrum. The technique firstly divides the image into sub parts, Fourier analysis has to be done in which the full Fourier spectrum is computed for each block (using FFT) and represented with the origin at the center. Then the mean of the Fourier calculation of each block is determined for applying the radon transformation. A donut shaped mask is applied to the averaged Fourier spectrum to remove DC and very low spatial frequency components and make the radial bands of uniform bandwidth in all directions.

Lastly, modified form of Radon transform is applied to the mean Fourier block in the range [-90°, 90°] in angle increments of 1º. The peak of the Radon transformation is taken as the approximate angle of skew .The technique has also shown robustness to documents highly corrupted by noise or having a large number of graphical elements.

Apart from the approaches discussed above there are many more other approaches also.

Brodic et al. [13] described an algorithm that creates a rectangular hull around all the text characters of the document. Then it combines nearby rectangular hull to form objects. After this, mathematical morphology is applied on it, the biggest object is selected. The rectangular hull gravity center forms reference points on these objects is used as a base for calculation of the initial skew rate. The least square method is used and initial skew rate is calculated. Due to the correctness of this method, it can be integrated partly or completely with the existing algorithms for handwritten text parameter extraction.

Papandreou et al. [14] proposed a new document skew detection approach based on vertical projections and bounding box minimization criterion. It gives better results as compared to skew detection algorithms based on horizontal projections because it is noise and wrap resistant. For these reasons, this algorithm can be efficiently applied to historical machine printed documents.

Panwar et al. [15] introduced a new technique for skew normalization. It is based on orthogonal projection of the segmented line with respect to X-axis. The orthogonal projection approach detects the exact skew angle, and corrects it efficiently; it does not need to calculate the baseline explicitly as the orthographic projection with maximum length is the horizontal axis. Using this technique skew of any angle can be detected and hence can be applied to any language writing style. For e.g. the English language is written left to right but in case of Urdu language we write right to left. All such cases can be handled by this approach, when rotation angle varies from 0 degrees to 90 degrees.

Rezaei et al. [16] described a method for skew detection in two steps, dimension reduction and skew estimation. The first step involves feature transformation and feature selection, i.e. initial set of features transform to other set and then optimal set of features are selected based on an objective function. At the end, a criterion function for skew estimation is obtained. Then, the angle corresponding to the maximum or the minimum value of the function is usually considered as the skew angle.

III. CONCLUSION

There are number of techniques available for correcting the skew in a document image; some of them are discussed above in the paper like Hough transform, projection profile, nearest neighbor, cross correlation etc. Each of these techniques has some limitations; some of them provide speed but are suitable for small text, some give accurate results but are slow in speed and some are costlier if they provide speed and accuracy. The research is still going on how to remove the limitations in these techniques.

REFERENCES

[1] Kaur, Mandip, and Simpel Jindal. "An Integrated Skew Detection And Correction Using Fast Fourier Transform And DCT."

[2] Kwag, H. K., S. H. Kim, S. H. Jeong, and G. S. Lee. "Efficient skew estimation and correction algorithm for document images." Image and vision computing 20, no. 1,2002,pp:25-35

[3] Manjunath Aradhya, V. N., G. Hemantha Kumar, and P. Shivakumara. "Skew Detection Technique for Binary Document Images based on Hough Transform."International Journal of Information Technology 3, no. 3 ,2007

[4] Singh, Chandan, Nitin Bhatia, and Amandeep Kaur. "Hough transform based fast skew detection and accurate skew correction methods." Pattern Recognition 41, no. 12, 2008, pp: 3528-3546.

[5] Kumar, Deepak, and Dalwinder Singh. "Modified Approach of Hough Transform for Skew Detection and Correction in Documented Images." International Journal of Research in Computer Science 2, no. 3, 2012, pp: 37-40.

[6] Chen, Ming, and Xiaoqing Ding. "A robust skew detection algorithm for grayscale document image." In Document Analysis and Recognition, 1999. ICDAR'99. Proceedings of the Fifth International Conference on, pp. 617-620. IEEE, 1999.

[7] Lin, Chin-Teng, Kan-Wei Fan, Chang-Mao Yeh, Her-Chang Pu, and Fang-Yi Wu. "High-Accuracy Skew Estimation of Document Images." International Journal of Fuzzy Systems 8, no. 3, 2006.

[8] Li, Shutao, Qinghua Shen, and Jun Sun. "Skew detection using wavelet decomposition and projection profile analysis." Pattern recognition letters 28, no. 5 (2007): 555-562.

[9] Liolios, Nikos, Nikos Fakotakis, and G. Kokkinakis. "Improved document skew detection based on text line connected-component clustering." In Image Processing, 2001. Proceedings. 2001 International Conference on, vol. 1, IEEE, 2001, pp. 1098-1101.

[10] Lu, Yue, and Chew Lim Tan. "Improved nearest neighbor based approach to accurate document skew estimation." In Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on, IEEE, 2003, pp. 503-507.

[11] Konya, Iuliu, Stefan Eickeler, and Christoph Seibert. "Fast seamless skew and orientation detection in document images." In Pattern Recognition (ICPR), 2010 20th International Conference on, IEEE, 2010, pp. 1924-1928.

[12] Lowther, Scott, Vinod Chandran, and Subramanian Sridharan. "An accurate method for skew determination in document images." In Digital Image Computing Techniques and Applications, vol. 1, 2002, pp. 25-29.

[13] Brodi , Darko. "The evaluation of the initial skew rate for printed text." Journal of Electrical Engineering 62, no. 3,2011,pp: 134-140.

[14] Papandreou, A., and Basilios Gatos. "A Novel Skew Detection Technique Based on Vertical Projections." In Document Analysis and Recognition (ICDAR), 2011 International Conference on, IEEE, 2011, pp. 384-388.

[15] Panwar, Subhash, and Neeta Nain. "A Novel Approach of Skew Normalization for Handwritten Text Lines and Words." In Signal Image Technology and Internet Based Systems (SITIS), 2012 Eighth International Conference on, IEEE, 2012, pp. 296-299.

[16] Rezaei, Sepideh Barekat, Abdolhossein Sarrafzadeh, and Jamshid Shanbehzadeh. "Skew Detection of Scanned Document Images." InProceedings of the International MultiConference of Engineers and Computer Scientists, vol. 1. 2013.

Comparative Analysis of Vision Enhancement Schemes

Gurpreet Kaur
Department of Computer Science
CT Institute of Technology
Jalandhar, India
preeti.k.dhillon@gmail.com

Gunjan Sethi
Department of Computer Science
CT Institute of Technology
Jalandhar, India
gunjan.ctit@gmail.com

***Abstract*--Image enhancement is one of the most popular algorithms used in vision applications for improving the visibility of the digital images. Recently much work is done in the different fields like medical, remote sensing, military applications etc. to improve the visibility of digital images. Many algorithms have been proposed so far for enhancing the digital images. This paper has review different image enhancement techniques. It has been found that the most of the existing researchers have neglected many issues; i.e. no technique is accurate for different kind of circumstances. The existing methods have neglected the use of illuminate normalization to reduce the problem of poor brightness which will be presented in the image due to poor weather conditions. It is also found that the color artifacts which will be presented in the output image due to the transform domain methods; also neglected by the most of the researchers.**

Keywords--Image enhancement, Histogram equalization, Discrete Wavelet Transform, illuminate normalization, color artifacts.

I. INTRODUCTION

The methods for improving [1] the quality of digital images are known as image enhancement techniques. It is relatively simple, for example, to make an image light or dark or to enlarge or reduce contrast. Sophisticated image enhancement software also supports several filters for changing images in a variety of ways. The main purpose of image enhancement is to process a given image so that the outcome is more appropriate than the original image for a definite use.

It sharpens image features [2] such as edges, boundaries, or contrast to build a graphic display more useful for display and analysis. The enhancement [3] doesn't raise the inbuilt information content of the data, but it increases the active range of the selected features so that they can be detected simply. Image enhancement methods [4] can be based on either spatial or frequency domain techniques.

In the spatial domain method [4], the pixel composing of image facts are measured and the different procedures are directly applied on these pixels. The image processing functions in the spatial domain may be expressed as

$$G(x.y) = T[f(x,y)] \qquad (1)$$

Where $f(x, y)$ is the input image, $G(x, y)$ is the processed output image and T represents an operation on 'f' defined over some neighborhood of (x, y). Sometimes T can also be used to operate on a set of input images.

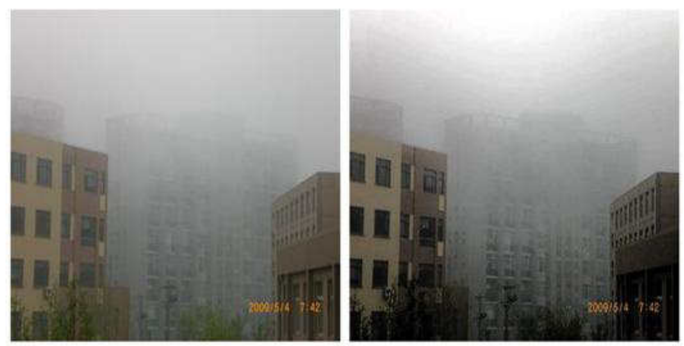


Fig. 1. Image (a) before (b) after enhancement

A. *Histogram Equalization (HE)*

It is a great point processing enhancement [5] method that seeks to optimize the contrast of an image at all points. It advances image contrast by destruction or equalizing the histogram of an image. A histogram is a table that basically counts the number of times a value appears in some data set. For an 8-bit image, there will be 256 promising samples in the image and the histogram will only count the number of times that each sample value really occurs in the image. The general shape of a histogram does not express much valuable information. The extension of the histogram relates straight to image contract -narrow histogram distributions are representative of low contrast images; wide histogram distributions are representative of higher contrast images.

Histogram [5] of an underexposed image will have a comparatively narrow distribution with a peak that is considerably shifted to the left and of an overexposed image will have a narrow distribution with a peak that is significantly shifted to the right.

Fig 2 (a) Original Image (b) Result of HE.

Histogram [6] is a means of improving the local contrast of an image without changing the global contrast to a considerable amount. This process is particularly helpful in images having large regions of related tone such as an image with a very light backdrop and dark forefront. Histogram equalization can depict hidden details in an image by stretching out the contrast of local regions and hence making the differences in the regions more observable. Histogram equalization uses Cumulative Distribution Function (CDF) as the research table. For example, for an N-bit image, histogram h, normalized CDF is given by:

$$\hat{C}j = \sum_{i=0}^{j} \hat{h}i, \quad j \quad \{0,1, \ldots\ldots\ldots 255\} \qquad (2)$$

CDF gives that what proportion of samples in an image are equal to or less than value j. Normalized CDF must be rescaled to [0,255] and is then used as the research table. It increases the monotonicity. Slope of CDF is vertical where there is a group of information in the source and is horizontal where there is little information in the source. CDF of completely equalized image is straight line with slope 1.

A good quality histogram [6] is which covers all the probable values in the gray scale used. This histogram suggests that the image has fine contrast and details in the image may be observed effortlessly. Histogram equalization is the straight forward method used to accomplish enhanced quality images in black and white color scale in different application areas such as medical image processing that includes X-ray ,MRIs and CT scans ,object tracking ,speech recognition etc. The chief benefit is that it is easy and efficient. The main two disadvantages are: the destruction property; not often utilized in purchaser electronics products such as TV because it may extensively change the original brightness and cause adverse artifacts.

B. Bi-Histogram Equalization (BHE)

The purpose of the bi-histogram equalization [7] is to conserve the mean brightness of a certain image. The input image is decomposed into two sub-images based on their means and the resulting equalized sub-images are enclosed by each other about the input mean. In hardware implementation, this method requires additional complex hardware than the typical Histogram Equalization (HE). For effective use of this technique, an attempt to decrease the difficulty should be ended. Many applications can be made achievable by utilizing this technique in the field of consumer electronics such as TV, VTR (Video Tape Recorder), Camcorder.

C. Adaptive histogram equalization

Histogram equalization emphasize [15] only on local contrast instead of overall contrast. Adaptive histogram equalization overcomes from this problem, this technique applicable for overall techniques. Histogram equalization uses similar transformation resultant from the image histogram to transform all pixels. This works well when the distribution of pixel values is similar throughout the image [3].

a) Original image b) Output AHE

Fig3. The results of adaptive histogram equalization (a) original image (b) output results of adaptive histogram equalization.

However when the image contains regions that are extensively lighter and darker, the contrast in those regions will not be adequately enhanced. Adaptive histogram equalization equation computed as

If (x,y) is a pixel of intensity i from the image, then we note with $m_{+,-}$ the mapping of right upper $x_{+,-}$,$m_{+,+}$ the mapping of right lower$x_{+,+}$, $m_{-,+}$the mapping of left lower $x_{-,+}$ and $m_{-,-}$ the mapping of the left lower $x_{-,-}$ then

$$m(i) = a[bm_{-,-}(i) + (1-b)m_{+,-}(i)] + [1-a][bm_{-,+}(i) + (1-b)m_{+,+}(i)] \qquad (3)$$

Where

$$a = \frac{y - y_-}{y_+ - y_-}, b = \frac{x - x_-}{x_+ - x_-}$$

On this by transforming each pixel with a conversion function obtained from a neighborhood region adaptive histogram equalization improves.

D. Contrast limited adaptive histogram equalization

The methods that prevent the limiting the amplification called contrast limited adaptive histogram equalization. This technique is differing from above in its contrast liming. The contrast limiting procedure has to be applied for each neighborhood from which a transformation function is derived in contrast limiting adaptive histogram equalization [13].

a) Original image b) Output CLAHE

Fig4. The results of contrast limited adaptive histogram equalization (a) output image (b) output result of CLAHE

The contrast amplification in the neighborhood of a given pixels value is given by the slope of the transformation function. This is proportional to the slope of the neighborhood cumulative distribution function and therefore to the value of the histogram at the pixel value. The general equation for contrast limited adaptive histogram equalization is

$$N_{aver} = \frac{N_{CR-Xp} \times N_{CR-Yp}}{N_{gray}} \qquad (4)$$

Where N_{aver} is average number of pixels, N_{gray} is number of gray level in the contextual region, $N_{CR-Xp}$is the number of pixel in the X-dimension in the contextual region, N_{CR-Yp} is the number of pixel in Y-dimension in the contextual region.

II. LITERATURE SURVEY

A novel contrast enhancement [1] approach based on dominant brightness level analysis and adaptive intensity transformation for remote sensing images. Firstly, they perform the discrete wavelet transform (DWT) on the input images and then decompose the LL sub-band into low, middle and high intensity layers using log-average luminance. Adaptive intensity transfer function is estimated using the knee transfer function and the gamma adjustment function. The resulting enhanced image is obtained by using the inverse DWT. This method can effectively enhance any low-contrast images acquired by a satellite camera and also suitable for other various imaging devices.

Maximum entropy value [2] indicates the maximum enhanced image. The proposed method has been applied on any RGB images collected from any white light endoscopic devices. The optical transfer function-based [3] micro image enhancement algorithm can generate a better micro image enhancement effects. The Laplacian pyramid [4] is ever-present for decomposing images into several scales and is commonly used for image analysis. It has been described the necessary view of the Laplacian pyramid decomposition, and analysis using user-defined threshold values to differentiate between the image detail and edges of the disadvantages, and advise to use the global information directly to obtain the threshold value method. It always produces high-quality results in the process of image detail enhancement.

Local contrast enhancement [5] increases the gray level of original image on the basis of light and dark edges. This method has applied on m×n size of an original gray scale image. The local mean and local standard deviation of entire image, minimum value and maximum value of the image are used to statistically describe digital image. The characteristics of each pixel [6] in the image and its multiracial spectrum have been calculated. Then pixels are classified by Human visual system, which is more susceptible to the edge structure of image. From softness area to edge area, the pixels are weighted and enhanced. The edge pixels has been classified and enhanced in accordance with the susceptibility of human visual system to the edge shape of an infrared image.

The image enhancement method [6] is more accessible and highlights the human eyes susceptible image area. The difficulty of low visibility and blurry edge has been resolved well and the enhancement image is more appropriate to human's observing.

A technique [7] that is encouraged by retinex theory and histogram rescaling techniques, which is natural rendering of color image based on retinex. Retinex theory is most attractive approaches for image enhancement and color consistency. It applies five steps of image processing, namely global mapping using a circle function, luminance enhancement using modified one filter retinex, histogram rescaling for luminance channel, a map-based image enhancement and finally a histogram rescaling. The integration of on-filter retinex and histogram rescaling also improves natural appearances of image.

Images that have poor contrast; due to fog or for any other reason has been considered in [8]. Contrast limited adaptive histogram equalization (CLAHE) method has been limited the noise and enhancing the image contrast. According the method original image converted into the RGB to HIS and then processed by CLAHE. The resultant image has shown significant improvement in fog degraded images. On the comparison with other method, this method is simple and faster.

A approach for overcoming the traditional noise reduction method [9] have proposed a new strategy. It firstly remove the noise from flat as well as edge regions. Wavelet basis and Gaussian low pass filters has used for directional transform. This framework reduces the noise without losing sharp details. This approach suits for real time application.

There are [10] so many types of image enhancement techniques that makes the image results better that associate to the person visual system. It includes the two techniques bilateral tone Adjustment and Saliency Weighted Contrast Enhancement both combined in image enhancement framework. The saliency-weighted Contrast enhancement integrates the notion of image saliency into an easy filter-based contrast enhancement technique. By using the luminance component in this saliency weighted contrast enhancement achieves extra performance. It proved that to achieve higher contrast enhancement with slight sound and huge image quality.

The new satellite image contrast enhancement technique [11] which is based on the DWT and singular value decomposition is proved to be efficient. Compared with conventional image equalization methods like standard general histogram equalization and local histogram equalization, as well as state-of-the-art methods such as brightness preserving dynamic histogram equalization and singular value equalization.

The image contains [12] information that is sometime not clear for human. Enhancement not only enhances the details that hidden in the scene and increases the recognition of interested targets. For executing the histogram projection independently contrast enhancement segmented into the sub-blocks. It enhanced the local details and conserve image brilliance to avoid blocking effect and wash-out effect. It has effective, efficient and flexible. The combination of frost filter and median filter [13] on CLAHEresulted images will help to remove the speckle noise and give better output.

The contrast enhancement histogram equalization [14] is an efficient method but it's not efficient for preserving the mean brightness of images. To overcome this problem weighted average multi segment histogram equalization method is proposed by using Gaussian filter for contrast enhancement of natural images. Weighted average multi segment histogram equalization give better results rather than multi-histogram equalization method when contrast enhancement along with brightness preservation is desired and also reduce the effect of noise that is present in the image.

Self-adaptive plateau histogram equalization [15] takes the threshold level by self that impossible in clipping and plateau histogram equalization. Self-adaptive plateau histogram equalization method is complicated and sometimes fails in execution. To overcome above problem modified self-adaptive plateau histogram equalization with mean threshold used. To overcome the self-adaptive plateau histogram method problems and also detect the local maximum and global maximum but, instead of median threshold value, mean threshold value is used for histogram modification. It enhances the image without introducing unwanted artifacts and gives better contrast enhancement and brightness preserving.

III. GAPS IN EARLIER WORK

The survey has shown that the most of image enhancement techniques has certain limitations. Following are the main limitations in earlier work.

1. It has been found that the most of the existing techniques are based upon the transform domain methods; which may introduce the color artifacts and also may reduce the intensity of the input remote sensing image.
2. Due to transform domain methods Gaussian random noise may be presented in the output images.
3. Remote sensing and underwater images has been neglected by many researchers in earlier work.

IV. CONCLUSION AND FUTURE DIRECTIONS

Image enhancement is one of the most popular algorithms used in vision applications for improving the visibility of the digital images. Many algorithms have been proposed so far for enhancing the digital images. This paper has review different image enhancement techniques. It has been found that the most of the existing researchers have neglected many issues; i.e. no technique is accurate for different kind of circumstances. The existing methods have neglected the use of illuminate normalization to reduce the problem of poor brightness which will be presented in the image due to poor weather conditions. It is also found that the color artifacts which will be presented in the output image due to the transform domain methods; also neglected by the most of the researchers.

So in near future we will use dark channel prior as the post processing function to enhance the results further. To overcome the short comings of the available techniques in near future we will modify the existing transform domain method using adaptive gamma correction to enhance the results further. However no implementation is considered in this work so in near future suitable simulation tool will be used.

REFERENCES

[1] Lee, Eunsung, et al. "Contrast enhancement using dominant brightness level analysis and adaptive intensity transformation for remote sensing images."Geoscience and Remote Sensing Letters, IEEE 10.1 (2013): 62-66.

[2] Imtiaz, Mohammad Shamim, TareqHasan Khan, and Khan Wahid. "New color image enhancement method for endoscopic images." Advances in Electrical Engineering (ICAEE), 2013 International Conference on. IEEE, 2013.

[3] Sun, Yaqiu, and Xin Yin. "Optical transfer function-based micro image enhancement algorithm." Communications Workshops (ICC), 2013 IEEE International Conference on. IEEE, 2013.

[4] Teng, Yanwen, Fuyan Liu, and Ruoyu Wu. "The Research of Image Detail Enhancement Algorithm with Laplacian Pyramid." Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing. IEEE, 2013.

[5] Goel, Savita, AkhileshVerma, and Neeraj Kumar. "Gray level enhancement to emphasize less dynamic region within image using genetic algorithm." Advance Computing Conference (IACC), 2013 IEEE 3rd International. IEEE, 2013.

[6] Tianhe, Yu, et al. "Enhancement of infrared image using multi-fractal based on human visual system." Measurement, Information and Control (MIC), 2012 International Conference on. Vol. 2. IEEE, 2012.

[7] Chen, Shaohua, and AzeddineBeghdadi. "Natural rendering of color image based on retinex." Image Processing (ICIP), 2009 16th IEEE International Conference on. IEEE, 2009.

[8] Xu, Zhiyuan, Xiaoming Liu, and Na Ji. "Fog removal from color images using contrast limited adaptive histogram equalization." Image and Signal Processing, 2009. CISP'09. 2nd International Congress on. IEEE, 2009.

[9] Kim, Sangjin, "Wavelet-domain color image enhancement using filtered directional bases and frequency-adaptive shrinkage." Consumer Electronics, IEEE Transactions on 56.2: pp.1063-1070, 2010.

[10] Ke, Wei-Ming, Chih-Rung Chen, and Ching-Te Chiu. "BiTA/SWCE: Image enhancement with bilateral tone adjustment and saliency weighted contrast enhancement." Circuits and Systems for Video Technology, IEEE Transactions on 21.3: pp.360-364, 2010.

[11] Demirel, Hasan, CagriOzcinar, and GholamrezaAnbarjafari. "Satellite image contrast enhancement using discrete wavelet transform and singular value decomposition." Geoscience and Remote Sensing Letters, IEEE 7.2: pp.333-337, 2010.

[12] Liu, Bin, et al. "Contrast enhancement using non-overlapped sub-blocks and local histogram projection." Consumer Electronics, IEEE Transactions on 57.2: pp. 583-588, 2011.

[13] Josephus, ChelsySapna, and S. Remya. "Multilayered Contrast Limited Adaptive Histogram Equalization Using Frost Filter." Recent Advances in Intelligent Computational Systems (RAICS), 2011 IEEE. IEEE, 2011.

[14] Khan, MohdFarhan, Ekram Khan, and Z. A. Abbasi. "Weighted average multi segment histogram equalization for brightness preserving contrast enhancement." Signal Processing, Computing and Control (ISPCC), 2012 IEEE International Conference on. IEEE, 2012.

[15] Raju, Aedla, G. S. Dwarakish, and D. Venkat Reddy. "Modified self—Adaptive Plateau Histogram Equalization with mean threshold for brightness preserving and contrast enhancement." Image Information Processing (ICIIP), 2013 IEEE Second International Conference on. IEEE, 2013.

Extracting Low Level Features in Content Based Image Retrieval Systems using Open CV Environment

Bhavneet Kaur
Department of CSE
SBS State Technical Campus
Ferozepur, India
Sethibhavneet37@gmail.com

Sonika Jindal
Department of CSE
SBS State Technical Campus
Ferozepur, India
sonikamanoj@gmail.com

Richa Jindal
Department of CSE
PEC Technical University
Chandigarh, India
hrricha@gmail.com

***Abstract*—Feature extraction is a significant and also the initial phase of Content Based Image Retrieval (CBIR) where the features of an image are extracted using a particular technique and thereafter stored in the form of a feature vector. This work basically shows the feature extraction of images using OPEN CV platform. OPEN CV is a library of programming functions mainly aimed at real-time computer vision. We have used its inbuilt functions and libraries and extracted the features that include key points, contours and texture of images. The output to the feature extraction can be taken as the input at the next level of CBIR.**

Keywords—Feature extraction,SIFT,SURF,OPEN CV.

I. INTRODUCTION

The Content based image retrieval is a process of fetching the image from a database of image based on the query produced by the user. It is composed of various phases that include feature extraction, similarity matching and indexing [4].Feature extraction is one of the important phases of CBIR without which the whole process can be considered to be impossible. At this phase features that comprises of the data or content of the image are extracted and stored for taking as an input to the next phase. Figure 1 shows the various phases of CBIR where being the emphasis of our work , the feature extraction phase is highlighted.

There are several techniques followed under feature extraction where SIFT and SURF is one of them as discussed in this work. For the implementation part OPEN CV platform is put into use. It is an environment provided for image processing implementation using C++ interface. Open CV (Open Source Computer Vision Library) is a library of programming functions mainly aimed at real-time computer vision, developed by Intel, and now supported by Willow Garage and Itseez [1]. The main goal behind the launch of OPEN CV was to get a portable and an optimised code with programming library functions available for free. OPEN CV was developed by Intel Microprocessor Research Lab which was distributed under a BSD style license and allows free commercial or research use.

The paper is organized into various sections: Section II explains the feature extraction and further the Sub Section A types of features as broadly classified into level 1, level 2 and level 3 features. In the section III the two major techniques under feature extraction process namely SIFT and SURF are covered. The next section shows the implementation phase of our work.

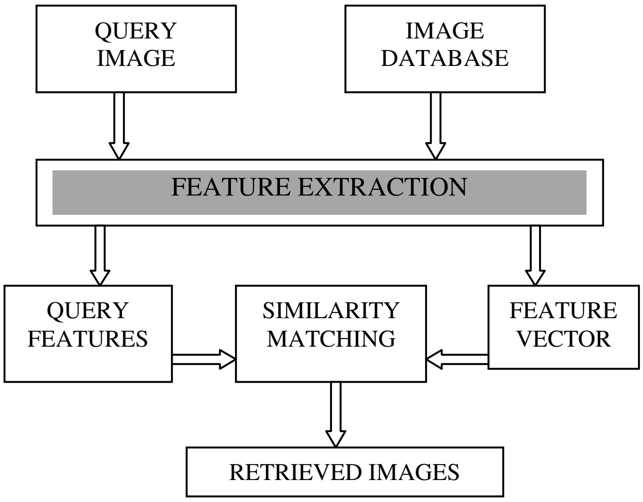


Fig. 1. PHASES OF CBIR

II. FEATURE EXTRACTION

This is an initial phase of CBIR where the features stored in an image are extracted and further stored in a feature vector. Features are basically the data or content stored in an image that describes the image. This process transforms the content of the images into the form of content features. The features thus extracted are utilized further in selection and classification phase [2].We can consider the image to be a two dimensional array taken as{F(x,y):x=1,2,…..,X, y=1,2,….,Y}.

For colored images (x,y) is taken as color value at pixel and that for gray scale it is taken as grayscale intensity value of that pixel[2].

Feature extraction phase performs the grid creation used for the identification of the features on an array and then computes the intensity value for each feature [3].In CBIR systems the images are stored at their corresponding indices [4].After the feature extraction process the features are stored using the feature vector as a data structure and therefore retrieval of images is done by performing the similarity matching of these feature vectors [4].The distance ‘D’ calculated between the feature vectors of query image ‘Q’ and image database ‘T’ is defined as,

$$D\ (\text{Feature}\ (Q), \text{Feature}\ (T)) \leq t \qquad (1)$$

A. Types of Features

Features are basically the data or the content stored by the image that gives a proper description of the image. They are functions defined for one or more measurements taken for the certain characteristics of the object [4].The features are represented using the feature descriptors. Features are of following types:

i. *Level 1 features:* These are basically based on the whole image or on the part of an image as well.

- Color: The color features are extracted and studied by using the color histograms. Here the matching is done on the basis of the color histograms which resemble to the color histogram of the query image most closely [6].Other representations for color feature are color moments, color correlograms, coherence vectors [8].
- Texture: It is a powerful regional descriptor that supports the image retrieval process. It helps in classification, recognition and finding the similarities between the images from multimedia databases [2].Similarity done on the basis of texture helps in distinguishing between areas of images with similar color [6].Texture can be represented using structural that represent texture as texels (texture element, or texture pixel) that are arranged regularly on surface based on particular arrangement [8] and statistical methods that includes Fourier power spectra, co-occurrence matrices, shift-invariant principal component analysis (SPCA), Tamura features, Wold decomposition, Markov random field, fractal model, and multi-resolution filtering techniques such as Gabor and wavelet transform, etc[2].
- Shape: Shape features comprises of aspect ratio, circularity, Fourier descriptors, moment invariants, consecutive boundary segments, etc. Shape descriptors are used for the representation of shape features [7].
- Spatial location: It is used in region classification. This particular feature may be used to differentiate when color and texture features are same. It is defined as upper, bottom, top according to the location of the region in an image [7].

III. *Level 2 features:* These are also known as high level features. They give a semantic or a rational meaning to the features of an image. There is a semantic gap between low level and high level features [9] which is defined as a space of disappointment between the high level features of CBIR and low level features that are used for query purpose. The low level features can be translated into their semantic interpretation by using various techniques. The semantic information basically extracts the logic of the low level features like color, texture etc as given by the user and processes it further and finally sets a relevance feedback for the user and concludes whether the fetched output is as desired or not. The relevance feedback is obtained in several iterations. It basically reduces the semantic gap in CBIR systems [10].Following are some major techniques under representation of level 2 features:

- SFL-Semantic Feature Layer: This technique is basically a design semantically related classes containing low level features and also includes additional knowledge like modeling information, Domain knowledge etc. This approach is implemented for high level features describing human world properties and for evaluation of 300 queries [9].
- Fuzzy production rules: This technique takes the extracts low level features like contrast of hue, light-dark contract, shape features and texture features and convert them into high level semantic features using fuzzy production rules those are derived from an image mining technique. The list of structures containing the content for the high level semantic features is created on the basis of Dempster-Shafer theory [11].
- Semantic grouping algorithm: It is a statistical algorithm for keyword annotations based on user feedback. This is an intelligent system that can gradually learn the user’s searching habits in terms of semantic relations among concepts and further uses this information for improvement of the retrieval performances [12].

IV. *Level 3 features:* These features are used for automatic image retrieval [6]. These are used to study the semantic relations of user’s search. Here semantic grouping of keywords is done using a particular algorithm. Keywords are generated from textual and manual annotations of images [12].

V. TECHNIQUES FOR FEATURE EXTRACTION

There are two famous techniques used under feature extraction named SIFT and SURF.

A. SURF

SURF stands for Speeded up Robust Features. It is an algorithm following multiple stages and thus speeds up the feature extraction process in-case of images. The SURF descriptor is based upon the Hessian based scale-space pyramid for searching of image feature points [13].It is a method that is perform ant scale and rotation-invariant interest point detector and descriptor. It is much better than other feature extraction techniques in terms of repeatability, distinctiveness, speed as well as robustness. It is better since it depends upon integral images for image convolutions [14].It is the speeded up version of SIFT feature extraction algorithm. Surf is three times faster than SIFT and same can be viewed in terms of performance also [15].It works well for the images with blurring and rotation but does not succeed well in handling [16].

SURF also helps in automatic image annotation where it selects the correct number of features and the features itself for annotation [16].Medical image annotation can also be done using SURF descriptor and the SVM classifier. There is a Fast-Hessian detector used to perform feature extraction and feature matching is performed using SVM with a quadratic kernel and on comparison with SVM classification it resulted into the improved classification of lung images with 96% accuracy. It is counted as a strong tool for medical image annotation [17].

B. SIFT

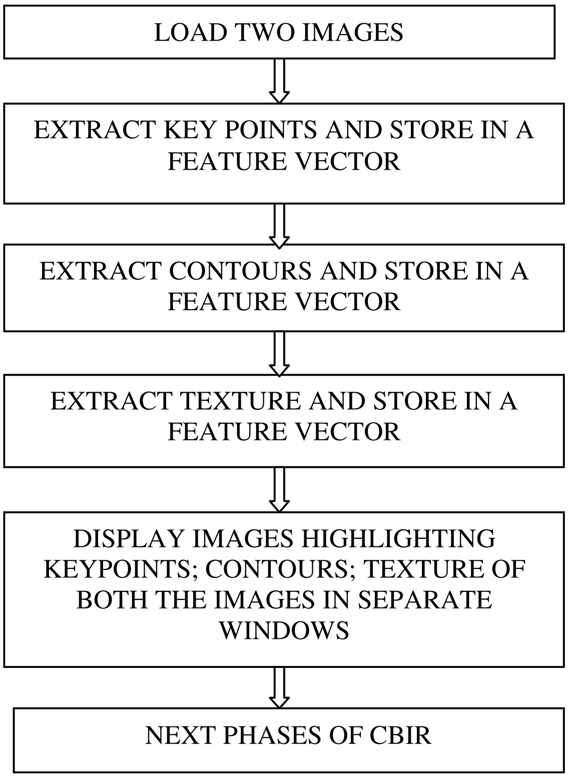
SIFT stands for Scale Invariant Feature Transform. It was proposed by David Lowe in 1999.Its name conveys its meaning as the content of the image is converted to scale invariant coordinates relative to local features [18].It is one of the techniques in image processing and computer vision that is used to capture the scene semantics [19].The major steps covered under SIFT algorithm are as follows [20]:

1. Find scale space extrema.
2. Key point localization and filtering.
   - Improve key points and throw out bad ones.
3. Orientation assignment.
   - Remove effects of rotation and scale.
4. Create descriptor.
   - Using histograms of orientations.

First two steps are the parts of Detector and last two are part of Descriptor [21].The algorithm has points or features that are local minima and maxima in a scale space and further composed of differences of Gaussians also known as DOG. Normalized histograms of the gradient directions around those points are used for describing them [21].The visual contents of a query image and database images can be extracted and further described using 128-dimensional SIFT feature vectors. Here indexing and matching of these SIFT features can be done using KD-tree and Approximate Nearest Neighbor (ANN) Search algorithms [22].

VI. IMPLEMENTATION

This section shows the implementation part of our work where we have extracted some common features from two test images using OPEN CV platform. It is an environment that provides c++ interface for image processing algorithms. The OPEN CV library is used by a large number of companies like Intel, IBM, Microsoft, SONY, Siemens, Google, and research centres Stanford, MIT, CMU, Cambridge, INRIA etc [23]. It is a open source computer vision library in C or C++ and optimized for real time applications. Figure 2 shows the system architecture of the work that has been done. The libraries and files can be downloaded and installed from [25].



1. We load two different images taken in jpeg format.
2. Extract features using SURF in OPENCV.
3. Display the different images where three different features are highlighted separately.

There are three basic features extracted:

VII. *Key points:* Key points are basically those points in an object of a particular image that give the description of features in that object of that image. Such a description can be used to locate that object in an image containing many other objects [24] . The key points are detected using SURF detector. The key points are detected and its descriptors are calculated and finally those extracted key points are stored in a vector array. After this the key points are drawn and then displayed in the form of a new image.There are various other algorithms available in OPEN CV for feature detection like SIFT, ORB, FAST.

VIII. *Contours:* The contours are the outlines of an image. It includes edges and boundaries. They are extracted by declaring and defining a function in OPEN CV that detects the edges of the image using canny edge detector. Then the contours are found and drawn from an initial point up to a final point thus creating a hierarchy of points. Hierarchy is a vector that contains the information about the image topology. [25].Finally, the contours are shown in the form of a new image.

IX. *Texture:* The texture is extracted by using al algorithm provided in OPEN CV.As a result the textured data will be printed on the new image. Therefore, as an output we get texture of the original image in the form of a new image.

X. EXPERIMENT RESULTS

We first laod the images and display them using the imshow function of OPEN CV.The images taken are gray scale. Then we have extracted the features of both the images simultaneously one after the other and these feeatures thus extracted are stored in a feature vectors.Each feature is stored in its corresponding feature vectyor.Figure 3 shows the two test images taken whose features are to be taken and below them there corresponding extracted features are shown in series where at first keypoints are shown,next are the contours and at the end we have extracted the texture.The images store a lot of content and are complex to store since there are a number of matrices that are formedSo, the hardware configurations do also matter while implementing image processing algorithms.

The output of the feature extraction phase can be taken as the input to the next phases of CBIR that may include indexing and similarity matching.

XI. CONCLUSION

The paper shows the most significant phase of CBIR that is feature extraction and how we have extracted out the features using OPEN CV platform. The features include key points, contours and texture. The extracted features are stored in a feature vector and further can be taken as an input to the next phases of CBIR. This is a sequential implementation and to obtain better speed ups and effective execution time we would perform it in parallel using Graphics Processing Unit and CUDA.

ACKNOWLEDGMENT

The authors would like to thank Dr.Satvir Singh and Mr. Sarabjit Singh from Shaheed Bhagat Singh,State Technical Campus,,Ferozepur, India for their valuable assistance in formulating this work.

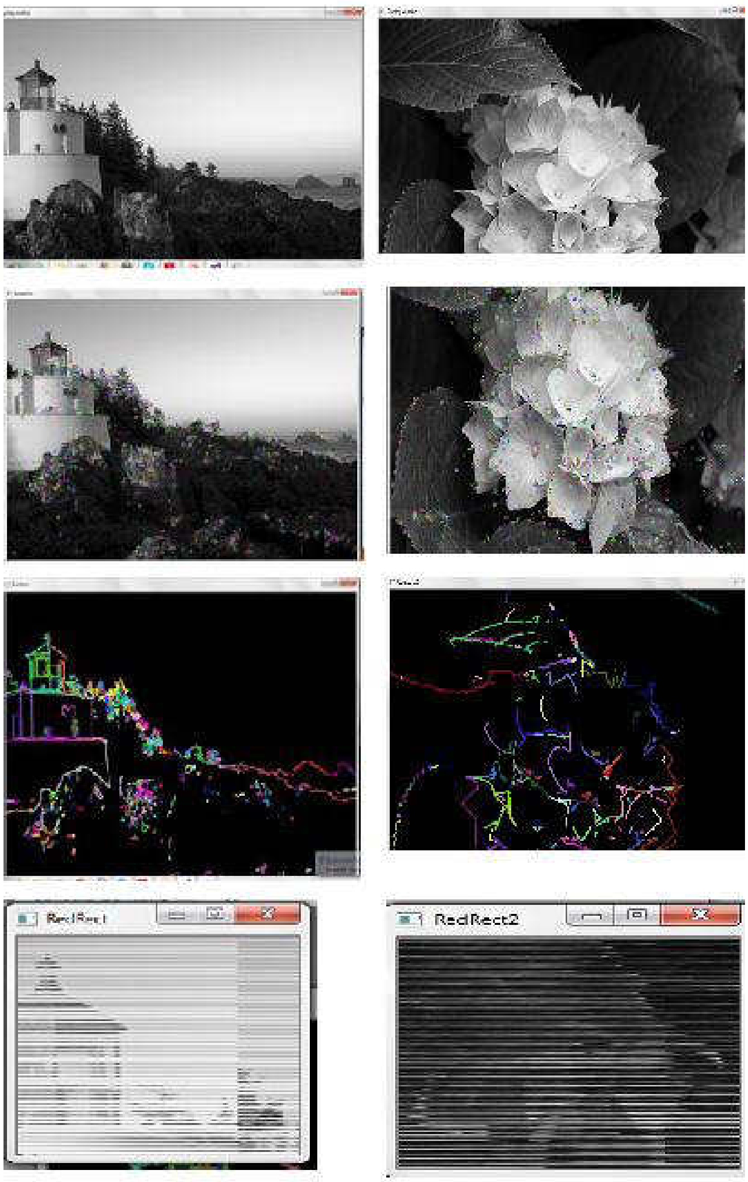


Fig. 3. Results

REFERENCES

[1] http://en.wikipedia.org/wiki/OpenCV

[2] Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems, Ryszard S. Chora´s, INTERNATIONAL JOURNAL OF BIOLOGY AND BIOMEDICAL ENGINEERING,2007,pages ={6--16}.

[3] Feature Extraction and Normalization Algorithms for high density Oligonucleotide Gene Expression Array Data,Erich E Schadt,Cheng Li,Journal of Cellular Biochemistry Supplement,2001,pages={120 - 125}.

[4] Content based Image Retrieval with Graphical Processing Unit,Bhavneet Kaur,Sonika Jindal,ITC 2014,2014.

[5] Feature Extraction with Graphics Processing Unit, Bhavneet Kaur, Sonika Jindal,ICAET 2014 ,2014.

[6] Content based image retrieval,John Eakins,Margret Graham,1999.

[7] A suvey of content-based image retrieval with high level semantics,Ying Liu,Dengsheng Zhang,Elsevier,2007,pages={262--282}.

[8] Content-based Image Retrieval: Feature Extraction Techniques and Applications, Amandeep Khokher, Rajneesh Talwar, iRAFIT2012,2012.

[9] Semantic Feature Layers in Content Based Image Retrieval:Implementation of Human World Features,Horst Eidenberger,IEEE,ICARCV 2002, pages={174-179}.

[10] Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval, Rui, Y., Huang, T., Ortega, M. and Mehrotra, IEEE Transactions on Circuits and Systems for Video Technology,1998,pages={644--655}.

[11] Content Based Image Retrieval Using High Level Semantic,Prof. Sharvari Tmane,2008

[12] CBIR: From Low-Level Features to High-Level Semantics,Xiang Sean Zhou*, Thomas S. Huang, Electronic Imaging, International Society for Optics and Photonics2000,pages={426--431}.

[13] Fast Scale Invariant Feature Detection and Matching on Programmable Graphics Hardware, Nico Cornelis, Luc Van Gool, Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on,2008,pages={1--8}.

[14] http://www.vision.ee.ethz.ch/~surf/

[15] www.opencv.org/Surf/

[16] Automatic image annotation using SURF descriptors, Muhammed, A.V. ,Kumar, G.S. ,Sreeraj, M. ,IEEE,2012.

[17] ANNOTATION OF MEDICAL IMAGES USING THE SURF DESCRIPTOR, Anna Wojnar, Ant´onio M. G. Pinheiro, Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on,2012,IEEE,pages={130--133}.

[18] SIFT applied to CBIR, Jurandy Almeida, Ricardo da S. Torres and Siome Goldenstein, Revista de Sistemas de Informacao da FSMA n,2009,pages={41--48}.

[19] Qi Zhang, Yurong Chen, Yimin Zhang, Yinlong Xu,SIFT Implementation and characterization for Multi-Core Systems, Workload Characterization, 2008. IISWC 2008. IEEE International Symposium on,IEEE,2008pages={14--23}.

[20] Distinctive image features from scale-invariant keypoints. David G. Lowe, International Journal of Computer Vision, 60, 2 (2004), pp. 91-110

[21] CBIR in Cultural Databases for Identification of Images: A Local-Descriptors Approach, Eduardo Valle, Matthieu Cord, Sylvie Philipp-Foliguet,2008.

[22] Application of Image SIFT Features to the Context of CBIR, Xu Wangming, Computer Science and Software Engineering, 2008 International Conference on,IEEE,2008, page={552--555}.

[23] Pisarevsky, V.: 2007, Intel Corporation, Software and Solution

[24] homepages.inf.ed.ac.uk

[25] www.opencv.org.

Survey on Various Image Segmentation Techniques

Navdeep Singh
Department of Computer Science
Guru Nanak Dev University
Amritsar, India
navdeep.gsp@gmail.com

Amit Chhabra
Department of Computer Science
Guru Nanak Dev University
Amritsar, India
amit.cse@gndu.ac.in

***Abstract*— Image Segmentation is a technique that divides image into unique regions. The goal of segmentation is to make the image simplified so that it can be analyzed efficiently. Many segmentation techniques are available but none of them satisfies the global properties. Computer applications like Pattern Recognition, Object Recognition, and Automatic Traffic control are based on analysis of image segmentation. Type of segmentation can be used as per requirements of the application. This survey addresses many segmentation techniques and their limitations and probable solutions to recover them.**

Keywords—Image segmentation, edge linking, seed pixel.

I. INTRODUCTION

Image segmentation is division or partitioning of an image into meaningful structures. Image segmentation is an essential step in image analysis, pattern recognition, object recognition, and visualization etc. This technique is finding broad applications in underwater acoustics, space born remote sensing, medical imaging systems, non-distractive material analysis, as well as spoken language processing.

It is a process of partitioning a digital image into segments. Each segment is set of pixels also known as super pixels. Algorithms of image segmentation are basically based on two properties i.e. discontinuity and similarity of intensity values. First property is based on the areas or pixels that are quite different like edges of the objects in the image whereas second property works in the areas where the pixels are similar. This method is used in partitioning of image into regions that are similar to predefined criteria. [1]

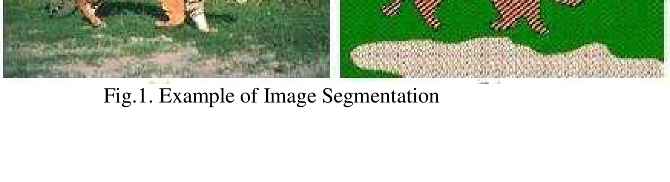


Fig.1. Example of Image Segmentation

Image segmentation is used in medical applications broadly. The amount of data to be analyzed exceeds the limit that humans can handle. Image segmentation is used in medical applications such as counting cells, classifying cells etc. Image segmentation provides more reliable results.

II. IMAGE SEGMENTATION TECHNIQUES

Image segmentation [3] is of the important part of image analysis. The basic idea of image segmentation is to distinguish the objects out of an image. It is all about differentiating background of the image and separating interesting objects out of foreground. Each pixel in an image is represented by its intensity level. Four basic categories of image segmentation are:

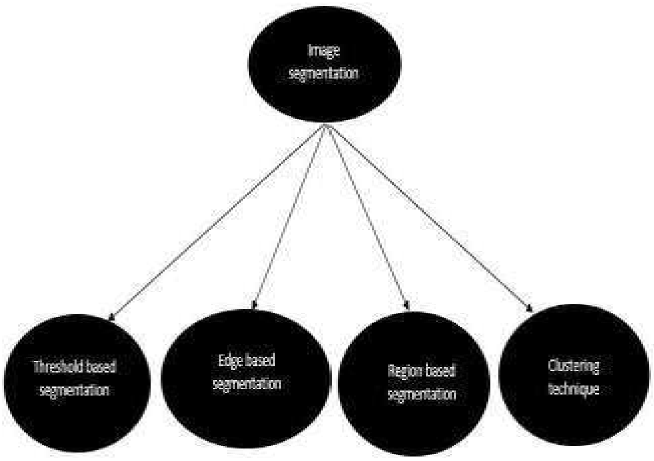


Fig. 2. Types of image segmentation

A. *Threshold based segmentation*

Threshold based technique [11] is most frequently used technique for segmentation of images. Threshold technique maps a grey-valued image to a binary image. After threshold operation, the image has been segmented into two segments, identified by the pixel values 0 and 1 respectively. Multiple methods exist to select an initial threshold value to be used in segmentation task. The most common method is to set the threshold value interactively i.e. the user manipulating the value and reviewing the result until a satisfying segmentation has been obtained.

Salem Saleh Al-Amri et al. [11] applied various threshold methods to compare with one another and find the best technique among them. Different threshold methods were applied on three satellite images in order to compare all the threshold methods. Histogram dependent technique and edge maximization technique were more effective than others such as mean method and p-tile method etc.

K. Wei et al. [12] found that current image segmentation techniques require lot of computational cost and are time consuming in order to perform image segmentation. It is a problem for real time applications. Therefore they proposed a new threshold based segmentation method. PSO technique was used to search an optimal threshold for the segmentation process. They implement the proposed hybrid method on Matlab 7.0. Results shown that their algorithm took 25 times less time as compare to traditional algorithm. It is good for real time applications.

B. Edge based segmentation

An object is fully represented by its edges; the segmentation of an image into separate objects can be achieved by finding the edges of those objects. A typical approach to segmentation using edges is (1) compute an edge image, containing all edges of an original image, (2) process the edge image so that only closed object boundaries remain, and (3) transform the result to an ordinary segmented image by filling in the object boundaries. The major problem in this approach is in edge linking; because there are some areas where is gap and no edge is detected. [5]

Yu Xiaohan et al. [15] proposed a new image segmentation technique based on edge detection and region growing segmentation methods. Their hybrid method helps the segmentation process to avoid from errors that occurred when both techniques used in a separate manner. Region growing is used to find the edge pixels in the image, while edge detection is used for finding edges of the objects. Experiments were conducted on 3D image data. Gaussian technique is used for smoothing after edge detection. Results have shown that their technique is better in order to preserve more edge information.

Yong-Ren Huang et al. [6] proposed a new concept to integrate the conventional image segmentation techniques in order to accomplish the reasonable segmentation results. An automatic seed selection algorithm using histogram was developed for both scale and color vector. And the luminance and chrominance were utilized in the image as a guidance to optimize the region growing and region merging. For texture regions elimination, the region distribution and the global edge information were employed to identify the region with texture characterization to obtain segmentation results. This new technique showed more accuracy of segmentation and region classification than proposed techniques.

Ying-Tung Hsiao et al. [5] proposed a novel approach for edge-based image segmentation. Image segmentation and object extraction play an important role in supporting content-based image coding, indexing, and retrieval. Authors proposed an image segmentation algorithm by integrating mathematical morphological edge detector with region growing technique. The images are first enhanced by morphological closing operations, and then detect the edge of the image by morphological dilation residue edge detector. After region growing is implemented, homogeneous regions are merged together by applying the process of region merging algorithm.

C. Region based segmentation

In edge based segmentation, objects were found by locating their boundaries. Whereas in region based segmentation the approach is of finding the object region instead of its edges. In this technique, any pixel of a region is selected and then homogeneous pixels are gathered to form a region until the boundary is met. Many merging methods of segmentation use a method called region growing to merge adjacent single pixel segments into one segment. Region growing needs a set of starting pixels called seeds. The region growing process consists of picking a seed from the set merging suitable neighbors to the seed. [2]

I Karoui et al. [13] proposed an unsupervised image segmentation method using level set methods and texture statistics. They claim that their method is different from other methods since it doesn't restrict to first order grey features and it doesn't assume independent variable. The implementation includes feature selection step to readjust the weights of each feature to get the segmentation. In experiment stage, filter response histogram is used to calculate the number of distributions. Wavelet is used to compute the energy of image wavelet of each band. PDE is used to re-initialize the level sets. Results have shown for a zebra image as correct segmentation.

Civahir Cigla et al. [14] presented a new image segmentation method, and tried improve the normalized cut image segmentation method. They used image with weighted un-directed graph, whereas nodes represent the regions, and weights between nodes represent the intensity match of neighboring regions. Their modified normalized cut method has solved the problem of over segmentation in which extra regions are created for image. Experiments are conducted on images of cow, mosaics, and multi-resolution NC image. The results shown that proposed method improve the segmentation.

Mahbubul Islam Chowdhury et al. [2] reported method for image segmentation that combines region growing and edge detection. Existing schemes that use region-based processing provide unambiguous segmentation, but they often divide regions that are not clearly separated. Algorithm begins by using region growing to produce an over-segmented image. Then modify the over-segmented output of the region growing using edge criteria such as edge strength, edge smoothness, edge straightness and edge continuity. Two techniques - line-segment subtraction and line-segment addition - have been used. In the subtraction technique, the weakest edge is removed at each step. In addition technique, the strongest edge is used to seed a multi-segment line that grows out from it at both ends. The addition technique produces better results than the subtraction technique where small curly regions produced by less significant regions in segmentation were not eliminated.

D. Clustering technique

Clustering [16] is also called agglomerative segmentation technique. It is used to denote techniques that are primarily used in exploratory data analysis of high-dimensional measurement patterns. Clustering methods attempt to group together patterns that are similar in some sense. This goal is very similar to what we are attempting to do when we segment an image, and indeed some clustering techniques can readily be applied for image segmentation.

Kai li et al. [9] discussed that when the spectral clustering algorithm deals with large size images, in order to get the affinity matrix, it costs a lot of computation time. To solve this problem, authors have proposed a fast image segmentation algorithm based on spectral clustering. Large size image is separated into several smaller images which are segmented in advance, and combine the segmentation results of each smaller image. A point is randomly selected in the integrated results to constitute the feature data of the large size image. Then the results are integrated to see the improvement and computation time is reduced. This algorithm gave better effectiveness than other image segmentation methods.

Yong-mei Zhou et al. [16] introduced new clustering algorithm with region-based image segmentation technique. Their method extract color, texture of each pixel of the image and then make the clusters on the basis of those features using mean-shift clustering approach, label the each region, and at the end make segments of image on the basis of these labels. Matlab 7.0 was used to implement their algorithm. Their experimentation shows that their method present better results in term of time complexity and segmentation.

TABLE I. COMPARISON OF IMAGE SEGMENTATION TECHNIQUES USED

Method used	**Description**	**Time complexity**	**User inter-action**	**Limitation**
[15] [6] used edge detection segmentation technique.	This method was more accurate as automatic seed selection algorithm was used and for text elimination region distribution was used. This gave better segmentation as well as better region classification.	More time complexity as edge linking is done in this as an extra task.	No	Less regions were obtained in the image back ground and time complexity was more than average time as for text elimination different algorithms were used.
[13] [14] [2] used region growing technique.	Algorithm begins by using region growing to produce an over-segmented image. Then modify the over-segmented output of the region growing using edge criteria such as edge strength, edge smoothness, edge straightness and edge continuity.	Less time complexity.	No	Regions that are not clearly separated do not give proper segmentati on results. Many small regions are segmented again and again where execution time of the algorithm could have been saved.

[16] [9] used clustering technique	Dividing a large size image into smaller ones and then segmenting them and combining results saves computation time. This algorithm gave better image segmentation and effective time complexity as compared to others.	More time consuming as pattern matching is very complex.	Yes	The algorithm was not applied for the segmentation of texture part of the image as this task is more time consuming in segmentation of an image.
[12] [11] used threshold technique	Different threshold methods were applied on three satellite images in order to compare all the threshold methods. Histogram dependent technique and edge maximization technique were more effective than others such as mean method and p-tile method etc.	More time complexity because threshold value is chosen repeatedly and user interaction is involved in order to select accurate threshold value.	Yes	Threshold value is to be set again and again as automated threshold value was not used in this algorithm.

III. CONCLUSION AND FUTURE WORK

Image segmentation is important part of image processing. Various segmentation techniques conclude that all are applicable as per the conditions provided. Recent research in field of image segmentation is presented in this paper. Considering time complexity, region based segmentation takes less time but from point view of implementation none of the technique provides best segmentation therefore hybrid solution of two or more techniques is best approach for image segmentation.

REFERENCES

[1] Agrawal, Pankaj, S. K. Shriwastava, and S. S. Limaye. "MATLAB implementation of image segmentation algorithms." In Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on, vol. 3, pp. 427-431. IEEE, 2010

[2] Chowdhury, Mahbubul Islam, and John A. Robinson. "Improving image segmentation using edge information." In Electrical and Computer Engineering, 2000 Canadian Conference on, vol. 1, pp. 312-316. IEEE, 2000.

[3] Felzenszwalb, Pedro F., and Daniel P. Huttenlocher. "Efficient graph-based image segmentation." International Journal of Computer Vision 59, no. 2 (2004): 167-181.

[4] Gennart, Benoit A., and Roger D. Hersch. "Computer-aided synthesis of parallel image processing applications." In SPIE's International Symposium on Optical Science, Engineering, and Instrumentation, pp. 48-61. International Society for Optics and Photonics, 1999.

[5] Hsiao, Ying-Tung, Cheng-Long Chuang, Joe-Air Jiang, and Cheng-Chih Chien. "A contour based image segmentation algorithm using morphological edge detection." In Systems, Man and Cybernetics, 2005 IEEE International Conference on, vol. 3, pp. 2962-2967. IEEE, 2005.

[6] Huang, Yong-Ren, and Chung-Ming Kuo. "Image segmentation using edge detection and region distribution." In Image and Signal Processing (CISP), 2010 3rd International Congress on, vol. 3, pp. 1410-1414. IEEE, 2010.

[7] Krishnakumar, Y., T. D. Prasad, K. V. S. Kumar, P. Raju, and B. Kiranmai. "Realization of a parallel operating SIMD-MIMD architecture for image processing application." In Computer, Communication and Electrical Technology (ICCCET), 2011 International Conference on, pp. 98-102. IEEE, 2011.

[8] Lan, Yi-hua, Cun-hua Li, Yong Zhang, and Xue-feng Zhao. "A novel image segmentation method based on random walk." In Computational Intelligence and Industrial Applications, 2009. PACIIA 2009. Asia-Pacific Conference on, vol. 1, pp. 207-210. IEEE, 2009.

[9] Li, Kai, and Xinxin Song. "A Fast Large Size Image Segmentation Algorithm Based on Spectral Clustering." In Computational and Information Sciences (ICCIS), 2012 Fourth

International Conference on, pp. 345-348. IEEE, 2012.

[10] Zhang, Hui, Quanyin Zhu, and Xiang-feng Guan. "Probe into Image Segmentation Based on Sobel Operator and Maximum Entropy Algorithm." In Computer Science & Service System (CSSS), 2012 International Conference on, pp. 238-241. IEEE, 2012.

[11] Al-amri, Salem Saleh, and Namdeo V. Kalyankar. "Image segmentation by using threshold techniques." arXiv preprint arXiv:1005.4020 (2010).

[12] Wei, Kaiping, Tao Zhang, Xianjun Shen, and Jingnan Liu. "An improved threshold selection algorithm based on particle swarm optimization for image segmentation" In Natural Computation, 2007 ICNC 2007 Third International Conference on, vol. 5, pp. 591-594. IEEE, 2007.

[13] Karoui, I., R. Fablet, J. M. Boucher, and J-M. Augustin. "Unsupervised region-based image segmentation using texture statistics and level-set methods." InIntelligent Signal Processing, 2007. WISP 2007. IEEE International Symposium on, pp. 1-5. IEEE, 2007.

[14] Cigla, Cevahir, and A. Aydin Alatan. "Region-based image segmentation via graph cuts." In Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on, pp. 2272-2275. IEEE, 2008.

[15] Xiaohan, Yu, and Juha Ylä–Jääski. "A new algorithm for image segmentation based on region growing and edge detection." In Proceeding IEEE International Symposium on Circuits & Systems, vol. 1, pp. 516-519. 1991.

[16] Zhou, Yong-mei, Sheng-yi Jiang, and Mei-lin Yin. "A Region-Based Image Segmentation Method with Mean-Shift Clustering Algorithm." In Fuzzy Systems and Knowledge Discovery, 2008. FSKD'08. Fifth International Conference on, vol. 2, pp. 366-370. IEEE, 2008.

Track-1

Technical Session 5

NEURAL NETWORK/FUZZY LOGIC

A Systematic Way of Hybrid Model Design and Comparative Analysis of EBGM and Eigen Values for Biometric Face Recognition using Neural Network

Jagmeet Singh Brar
Dept. of CSE
Govt. Polytechnic College, GTB Garh,
Moga (Punjab, India)

Sonika Jindal
Dept. of CSE
Shaheed Bhagat Singh College of Engg. & Tech.,
Ferozepur (Punjab, India)

***Abstract*—We set ourselves the task of recognizing persons from single images by reference to a gallery, which also contained only one image per person. Our problem was to address image variation due to differences in facial expression, head pose, position, and size . Automatic recognition is a vast area of computer vision, reaching from face detection, face localization, face tracking, extraction of face orientation and facial features and facial expressions. These will need to tackle some technical problems like illumination, poses and occlusions. we will focus on two recently used techniques in face recognition. The two techniques are eigenvalue and elastic bunch graph matching. In this paper, we have used EBGM to accurately recognize the faces by creating a mean image and matching the selected image with it whether the selected face be in any pose provided the main features like eyes, nose, mouth, etc. be clearly visible in order to recognize the face.**

Keywords-EBGM, Voice Recoginition, Speech.

I. BIOMETRICS SYSTEM

Although there has always been a need to identify individuals, the requirements of identification have changed in radical ways as populations have expanded and grown increasingly mobile. This is particularly true for the relationships between institutions and individuals, which are crucial to the well-being of societies, and necessarily and increasingly conducted impersonally—that is, without persistent direct and personal interaction.

Importantly, these impersonal interactions include relationships between government and citizens for purposes of fair allocation of entitlements, mediated transactions with e-government, and security and law enforcement. Increasingly, these developments also encompass relationships between actors and clients or consumers based on financial transactions, commercial transactions, provision of services, and sales conducted among strangers, often mediated through the telephone, Internet, and the World Wide Web. Biometric technologies have emerged as promising tools to meet these challenges of identification, based not only on the faith that “the body doesn’t lie,” but also on dramatic progress in a range of relevant technologies. These developments, according to some, herald the possibility of automated systems of identification that are accurate, reliable, and efficient. Biometrics (or biometric authentication) refers to the identification of humans by their characteristics or traits. Biometrics is used in computer science as a form of identification and access control. It is also used to identify individuals in groups that are under surveillance.

A. *Types of Biometric System*

i. Bertillonage:

The first type of biometrics came into form in 1890, created by an anthropologist named Alphonse Bertillon. He based his system on the claim that measurement of adult bones does not change after the age of 20. The method consisted of identifying people by taking various body measurements like a person’s height, arm length, length and breadth of the head, the length of different fingers, the length of forearms, etc. using calipers.

ii. Fingerprint Recognition:

Involves taking an image of a person's fingertips and records its characteristics like whorls, arches, and loops along with the patterns of ridges, furrows, and minutiae.

iii. Face recognition:

Technique records face images through a digital video camera and analyses facial characteristics like the distance between eyes, nose, mouth, and jaw edges. These measurements are broken into facial planes and retained in a database, further used for comparison.

iv. Iris recognition:

Analyzes features like rings, furrows, and freckles existing in the colored tissue surrounding the pupil. The scans use a regular video camera and works through glasses and contact lenses. The image of the iris can be directly taken by making the user position his eye within the field of a single narrow-angle camera. This is done by observing a visual feedback via a mirror. The isolated iris pattern obtained is then demodulated to extract its phase information.

v. Voice Recognition:

Combines physiological and behavioral factors to produce speech patterns that can be captured by speech processing technology. Inherent properties of the speaker like fundamental frequency, nasal tone, cadence, inflection, etc. are used for speech authentication.

Voice recognition techniques can be divided into categories depending on the type of authentication domain.

II. LITERATURE REVIEW

Mohammod Abul Kashem et al. [1] Face recognition has received substantial attention from researches in biometrics, pattern recognition field and computervision communities. Face recognition can be applied in Security measure at Air ports, Passport verification, Criminals list verification in police department, Visa processing , Verification of Electoral identification and Card Security measure at

ATM's. In this paper, a face recognition system for personal identification and verification using Principal Component Analysis (PCA) with Back PropagationNeural Networks (BPNN) is proposed. The dimensionality of face image is reduced by the PCA and the recognition is done by the BPNN for efficient and robust face recognition.

David Monzo et al. [2] presented a algorithm based on EBGM which replaces Gabor features by HOG descriptors and the original EBGM. The experiments results show a better performance behavior using public available databases. This better performance is explained by the properties of HOG descriptors which are more robust to changes in illumination, rotation and small displacements, and to the higher accuracy of the face graphs obtained compared to classical Gabor-EBGM ones.

Rajinda Senaratne et al. [3] Elastic Bunch Graph Matching is one of the well known methods proposed for face recognition. In this work, we propose several extensions to Elastic Bunch Graph Matching and its recent variant Landmark Model Matching. We used data from the FERET database for experimentations and to compare the proposed methods. We apply Particle Swarm Optimization to improve the face graph matching procedure in Elastic Bunch Graph Matching method and demonstrate its usefulness. Landmark Model Matching depends solely on Gabor wavelets for feature extraction to locate the landmarks (facial feature points).

III. PROBLEM FORMULATION

Even though several algorithms have been proposed in the literature, there cannot be a direct comparison between them since their success depends on the database of faces and the testing protocol used. The degree of difficulty of a database of faces is dependent on numerous factors such as the variation of illumination, face occlusions and degree of aberration of pose and expression of the faces in the images. The purpose of this paper is the implementation of two different families of face recognition algorithms in Matlab and their evaluation under a specific protocol and compare them also on bases of various parameters. We prepare a model which consist of both EBGM and eigenvalue metods for face recognition. We present a system for recognizing human faces from single images out of a large database containing one image per person. The task is difficult because of image variation in terms of position, size, expression, and pose.

A. *Basic steps of EBGM algorithm*

STEP 1:-Jets are selected by hand to serve as examples of facial feature

STEP 2:-A bunch graph is created. Each node of the bunch graph corresponds to a facial land mark and contains a bunch of model jets extracted from the model imagery

STEP 3:-Landmark points are located for every image,First a novel jet is extracted from the novel image.The novel jet's displacement from the actual location is estimated by comparing it to the most similar model jet from the corresponding bunch.

STEP 4:-A face graph is created for each image by extracting a jet for each landmark.The graph contains the location of the landmarks and the value of jets.The original image can then be discarded.

STEP 5:-Face similarity is computed as a function of landmark location and jet values.

B. *Objective of the Research*

Step1: To recognize a sample face from a set of faces.

Step2: Implementation of the hybrid model i.e combination of EBGM and eigenvalues.

Step3: Use of hybrid model for face recognition by using both the algorithm of face recognition.

Step4: Comparison of EBGM and eigenvalues on the basis of accuracy.

IV. IMPLEMENTATION STEPS

The proposed face recognition system passes through three main phases during a face recognition process. Three major functional units are involved in these phases and they are depicted in Figure .The characteristics of these phases inconjunction with the three functional units are given below:

A. Face library formation phase

In this phase, the acquisition and the preprocessing of the face images that are going to be added to the face library are performed. Face images are stored in a face library in the system. We call this face database a "face library" because at the moment, it does not have the properties of a relational database. Every action such as training set or eigenface formation is performed on this face library.

B. Training phase

After adding face images to the initially empty face library, the system is ready to perform training set and eigenface formations. Those face images that are going to be in the training set are chosen from the entire face library. After choosing the training set, eigenfaces are formed and stored for later use. Eigenfaces are calculated from the training set, keeping only the M images that correspond to the highest eigenvalues. These M eigenfaces define the M-dimensional "face space". As new faces are experienced, the eigenfaces can be updated or recalculated. The corresponding distribution in the M-dimensional weight space is calculated for each face library member, by projecting its face image onto the "face space" spanned by the eigenfaces. Now the corresponding weight vector of each face library member has been updated which were initially empty. The system is now ready for the recognition process. Once a training set has been chosen, it is not possible to add new members to the face library with the conventional method that is presented in "phase 1" because, the system does not know whether this item already exists in the face library or not. A library search must be performed.

C. Recognition and learning phase

In this, the User initiates the recognition process by choosing a face image. Based on the user request and the acquired image size, pre-processing steps are applied to normalize this acquired image to face library specifications (if necessary). Once the image is normalized, its weight vector is constructed with the help of the eigenfaces that were already stored during the training phase. After obtaining the weight vector, it is compared with the weight vector of every face library member within a user defined "threshold". If there exists at least one face library member that is similar to the acquired image within that threshold then, the face image is classified as "known". Otherwise, a miss has occurred and the face image is classified as "unknown". After being classified as unknown, this new face image can be added to the face library with its corresponding weight vector for later use (learning to recognize).

V. RESULT AND DISCUSSION

Our system has an important core of structure which rejects the fact that the images of coherent objects tend to translate, scale, rotate, and deform in the image plane. Our basic object representation is the labeled graph; edges are labeled with distance information and nodes are labeled with wavelet responses locally bundled in jets. Stored model graphs can be matched to new images to generate image graphs, which can then be incorporated into a gallery and become model graphs.

In our proposed system, we have created an Input Ouput Console from which the EBGM create a mean image of all the images stored in the database during the training session. The next step is to select an image in any posture for the recognition. Any finally our proposed system executes and deliver the results on the basis of both the algorithms Eigen Values and EBGM in terms of Time, Distance and the Face number(it will provide and prove the accuracy to the results of our system). Results which we achieved are more accurate and efficient.

A. Initial Image

During initialization, the Console window appears in which the mean image will be created and the image that has to be recognized will be selected.

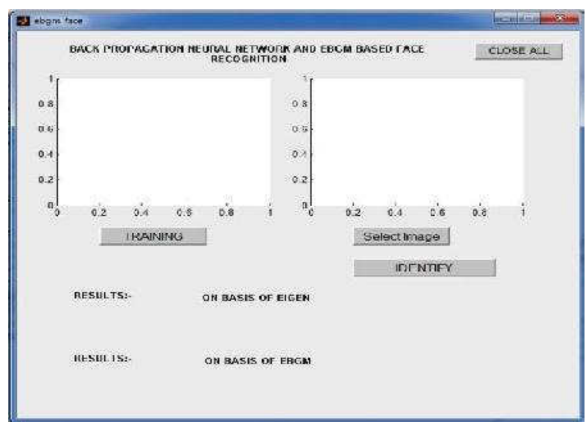


Fig.1. Initial Image

B. Training Step

In this step, the objective of the system is to create a mean image of all the pictures stored in the database. It will go to the database as shown in fig 2 and will create a mean image as shown in fig 3 shown below.

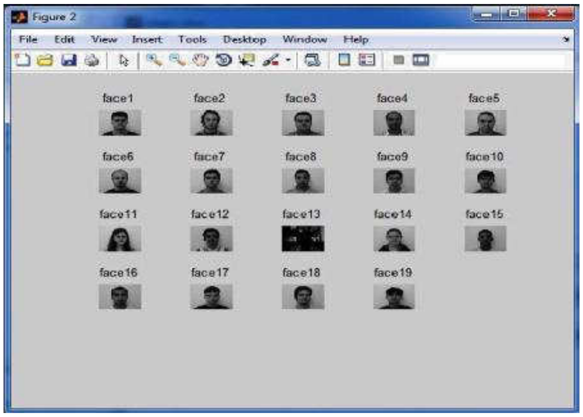


Fig.2. Training phase

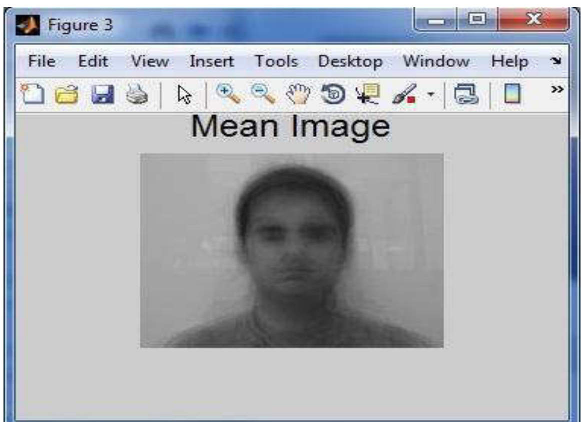


Fig.3. Mean image

C. *Select Image Phase*

During this step, the system will ask for the image (which is to be recognized) from the user. For this user need to click on Select Image Button in the below shown picture and it will open a browsing window.

Fig.4. Selection of image

D. *Choosing image from the browsing window*

Fig.5. Image from the browsing window

E. *Final image from the browsing window*

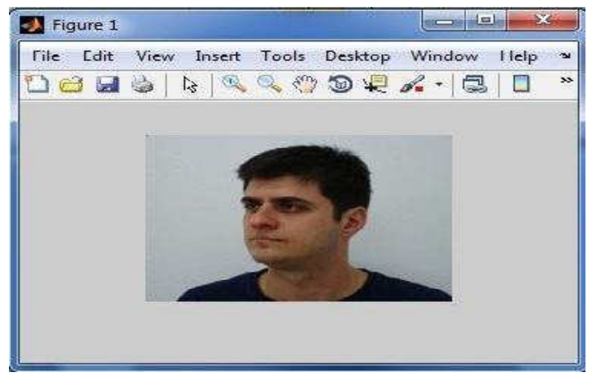


Fig.6. Final image from the browsing window

F. *Identification Phase*

During this step the system will start its execution, for which we just need to click on Identify Button.

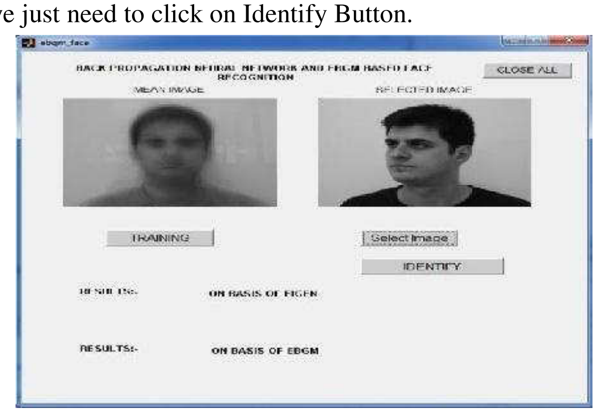


Fig.7. Identification of mean image with the selected image

G. *Neural Network Training Module & Final Result*

In Fig 8(a) the Neural Network Training Module will show the progress of the face recognition process showing the number of iterations done, time taken to recognize the face and calculating the gradient, MSE and the performance of the system.

In Fig 8 (b) Final result window displays the results of our proposed work, in which it clearly shows that Eigenvalues method failed to find the accurate face match and our proposed system(EBGM) recognize the image with full accuracy, as the accurate matched face is 1, which has been found by our proposed system whereas the EBGM provided us the false results showing the matched face 10.

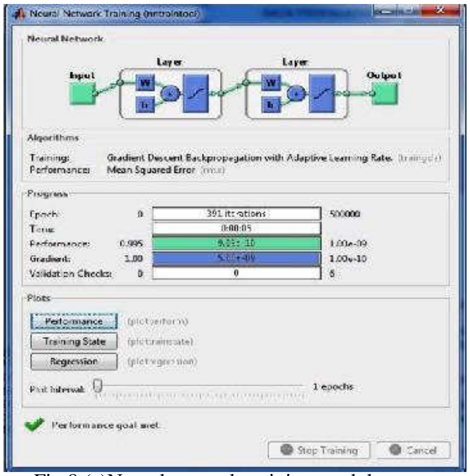


Fig.8 (a)Neural network training module

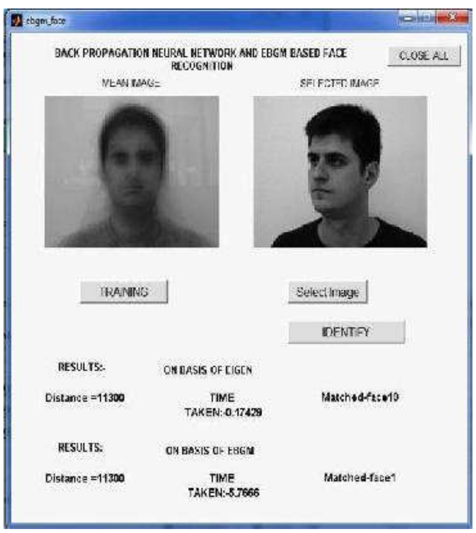


Fig.8 (b)Final result

VI. RESULT ANALYSIS

In our system we are doing face recognition using EBGM (Elastic Bunch Graph Method) and its Comparison with EIGEN values. In the initial phase we have created a database in MS-Access containing 18 images each having the variations (Left pose, Right Pose & Low Variation). We have applied our proposed algorithm along with EIGEN values on all the 18 images with 300 plus iterations on a single image, the conclusion derived by doing the analysis in the initial phase is that our proposed system is more accurate in recognizing the faces as compared to the EIGEN values. We have shown the collective results obtained in form of the graphs.

VII. CONCLUSION

In this paper, we have used Elastic Bunch Graph Method and Eigenvalues and obtained graphs in the Matlab environment. The results depicted in graphs state that the EBGM algorithm helps in recognizing the faces stored in the database with those selected by the user.Accuracy is the main issue in face recognition .We can recognize the face more accurately only when we have to detect the edges of the face from all the angles and all the possible postures in order to enhance the recognition of the face. Elastic Bunch Graph Method which is simple and easy to implement in order to optimize the cost for processing Client Job. EBGM has been successfully applied in many areas of image processing. This algorithm is very effective in giving quality solutions. It is motivated by the concept of elasticity which helps us to recognize the images more accurately. In this paper, we have used EBGM to accurately recognize the faces by creating a mean image and matching the selected image with it whether the selected face be any poster provided the main features like eyes, nose, mouth, etc. be clearly visible in order to recognize the face. This would help in enhancing the performance of the face recognition and other image processing strategies by provided the more accurate results.

REFERENCES

[1] Mohammod Abul Kashem, Md.Nasim Akhter, Shamim Ahmed and Md.Mahbub Alam,*). "Face recognition system based on principal component analysis with back propagation neural networks,"*in International journal of Scientific and Engineering research, *2011*

[2] David Monzo, Alberto Albiol, Jorge Sastre, Antonio Albio*, "HOG-EBGM VS. GABOR-EBGM,"* in Science Direct", *2008.*

[3] Rajinda Senaratne *"Face Recognition by extending Elastic Bunch Graph Matching with Partical Swarm Optimization"* in Journal of Multimedia, Vol. 4, No. 4, August 2009.

Health Monitoring System Using Fuzzy Logic

Parul Mohindru
Dept. of CSE
CT Institutions
Jalandhar, India
mohindruparul30@gmail.com

Vikshant Khanna
Dept. of Computer Engineering
Ramgahria Institutions
Phagwara, India
vikshantkhanna7@gmail.com

***Abstract*—Wireless Sensor Networks (WSN) offers proficient communication solution to the present healthcare systems. The miniaturized sensors together with advance micro-electro-mechanical systems (MEMS) technology create a body sensor network (BSN) that continuously monitors the health condition of patients. Some of the other applications where WSN can be used are military, interactive entertainment, and portable audio/video systems. The proposed system for measuring health parameters of patient consists of temperature and pulse sensor connected to base station through a microcontroller and allowing device to be controlled and monitored by remote computer. Wireless sensor network system continuously monitors pulse and temperature of patients at remote or in hospital. A wearable wireless sensor system is designed to continuously capture and transmit the bio-signals to the Doctor / Patient mobile phone. In case of emergency alert will be issued to doctors, relatives, and ambulance in the form of Short Message Services (SMS). Doctor can also provide remote prescription for the patients. Data stored at database is passed to fuzzy logic controller to improve accuracy and amount of data to be sent to the remote user. As the data read from sensor are imprecise / crisp here we design a Fuzzy logic controller (FLC), which is one of component in our computing prototype health status. The FLC system receives context information from sensor as input (which is in crisp form) and the fuzzification module converts input into fuzzy linguistic variable and output is sent to Patient / Doctor which can be easily understood by common person.**

Keywords*—*Fuzzy logic controller, Fuzzifier, Fuzzy rules, Inference Engine, Wireless Body Area Network (WBAN)

I. INTRODUCTION

Over few years WSN have proved their potentiality in various applications. A WSN is a collection of typically small, battery-powered, autonomous, wireless devices (also known as nodes). These devices have on-board processing, communication and sensing capabilities that can monitor physical or environmental conditions, such as temperature, sound, pressure, etc and pass data through a unidirectional or bi-directional network. The nodes are composed of low-power processor that has limited processing, memory device with limited storage capacity, a radio transceiver with a low-power internal/external antenna, low-data rate & limited range, sensors (scalar, cameras, microphones), and power source (batteries and solar cells). Each device is normally battery operated. Examining each such single device individually might appear to have small utility. WSNs find their place strong in the military applications. This extensive usage of WSNs drives into sensor networks research However, WSNs offer varied applications to most diverse fields like environmental monitoring, habitat monitoring, classroom/home, structural monitoring and health monitoring. Based on their characteristics, each of the applications has its own design concept and implementation to meet its specific requirements [1]. The advances of technology along with WSN's characteristics offer the maximum benefits to health care. A sensor network that is designed to sense the health parameters of a human being is a body sensor network (BSN). The nodes of BSNs are directly attached to human body and this requires utmost care. Some of the health care applications require the BSN to continuously operate gathering patient data for several days without user intervention. Such applications need to take care of the energy constraints of the sensor networks [3].

Challenges in Wireless Body Sensor Networks

In this modern life health-care is a must for quality life of every individual. The growing population of developed countries and government's budget are directly proportional. This represents challenges for health-care systems. One of the important challenges is to facilitate health-care for the senior citizens living independently. Generally, health monitoring is performed on a periodic check basis, where the patient must remember its symptoms; the doctor performs some tests and plans a diagnostic, then monitors patient progress along the treatment. Healthcare applications of wireless sensor networks allow in-home assistance, smart nursing homes, clinical trial and research augmentation. Before describing medical applications of BSNs, the following section focuses on several challenges and general aspects that describe this kind of technology. Challenges in healthcare application includes: low power, limited computation, security and interference, material constraints, robustness, continuous operation, and regulatory requirements with elderly people [4].

II. LITERATURE REVIEW

In [3], authors present the system architecture based on a three-tier Body Sensor Network. The BSN is responsible for health parameter's sensing, the sink is responsible for data aggregation from all sensors, fault tolerance, and outside communication, while the mobile device is responsible for data processing, calibration and presentation. As a result, if the mobile device malfunctions or is not present, data is not compromised.

In Phone-Centered Body Sensor Network [9] Platform consists of one mobile phone, multiple sensor nodes, and a Cache-watch Sensor nodes and the Cache-watch communicate directly with the phone via Bluetooth. With limitation of battery life, user acceptance, storage, with Bluetooth imposing power overhead along with minimum range.

A Wireless Body Sensor Network (WBSN) for healthcare monitoring is presented in [10], where a sink and a central system are used. The sink acts as the "middleman" between the WBSN and the central system. The central system is a standard personal computer and is intended for hospital facilities rather than for personal use.

Another solution to the growing problem of older adults care comes from [11]. The authors present the telecare concept and provide some information about a pilot study. The need for thresholds to achieve flexibility and support a broad range of individuals is also mentioned.

III. System Architecture

Several approaches exist with different design goals, but the same purpose: to provide health information to the patient, to the medical staff or for both. In this work we aim to provide a convenient mean to sense bio-signals, process it and show health information. Since nowadays almost everyone carries around a mobile phone, the choice seems obvious. The proposed wireless body sensor network for health monitoring integrated into a three tier telemedicine system. The lowest level consist set of intelligent sensors or nodes. These are the reduced function devise. The second level is the personal server (Internet enabled PDA, cell-phone, or home computer). These are full function devices. The third level encompasses a network of remote server which is the remote application to which data or information is transferred.

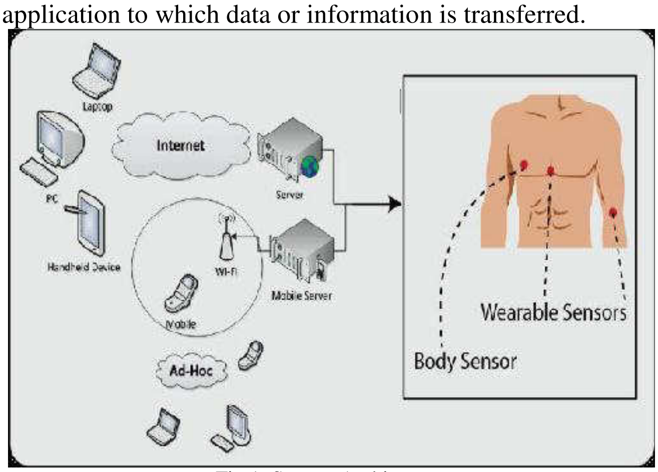


Fig 1: System Architecture

A. Operating Environment

Micro-controllers are useful to the extent that they communicate with other devices, such as sensors, motors, switches, keypads, displays, memory and even other micro-controllers. The microcontroller board chosen for this project is AT89S52 is a low-power, high-performance CMOS 8-bit microcontroller with 8K bytes of in-system programmable Flash memory. The AT89S52 provides the following standard features: 8K bytes of Flash, 256 bytes of RAM, 32 I/O lines, Watchdog timer, two data pointers, three 16-bit timer/counters, a six-vector two-level interrupt architecture, a full duplex serial port, on-chip oscillator, and clock circuitry.

The microcontroller [1] is interfaced to the laptop (server) through serial port communication. In this system, we used a 9600 baud connection with 8 data bits, no parity bit, and one stop bit .With this configuration, data can be sent over RF wireless communication.

Keil [8] is an IDE (Integrated Development Environment) which is used to develop an application program, compile and run it even the code can be debugged .It is a simulator where we can check the application code even in the absence of the hardware board.

Keil is also a cross compiler. The process of development of the soft code on a processor for a particular application and which can be implemented on the target processor is known as Cross Development. In our design the main heart of the hardware module is the micro controller which is the programmable IC .The programming language used for developing the software to the micro controller is Embedded C. The KEIL cross compiler is used to edit, compile and debug this program. Programmer is used for burning the developed code on Keil into the micro controller Chip.

B. Sensors

Wearable sensors and systems have evolved to the point that they can be considered ready for clinical application. The use of wearable monitoring devices that allow continuous or intermittent monitoring of physiological signals is critical for the advancement of both the diagnosis as well as treatment of diseases. Wearable systems are totally non-obtrusive devices that allow physicians to overcome the limitations of ambulatory technology and provide a response to the need for monitoring individual's over weeks or months.

C. Heart Beat Sensor

For heart beat sensor [7] each time the heart muscle contracts. Blood is ejected from the ventricles and a pulse of pressure is transmitted through a circulated system. This pressure pulse when traveling through the vessels causes vessel wall displacement which is measurable at various points. In order to detect pulsatile blood volume changes by photo electric method, photo conductors are used. Normally photo resistors are used, for amplification purpose photo transistors are used.

Light is emitted by LED and transmitted through the artery and the resistance of photo resistor is determined by the amount of light reaching it. With each contraction of heart, blood is forced to the extremities and the volume of blood in the finger increases, it alters the optical density with the result that the light transmission through the finger reduces and the resistance of the photo resistor increases accordingly. The photo resistor is connected as a part of voltage divider circuit and produces a voltage that varies with the amount of blood in the finger.

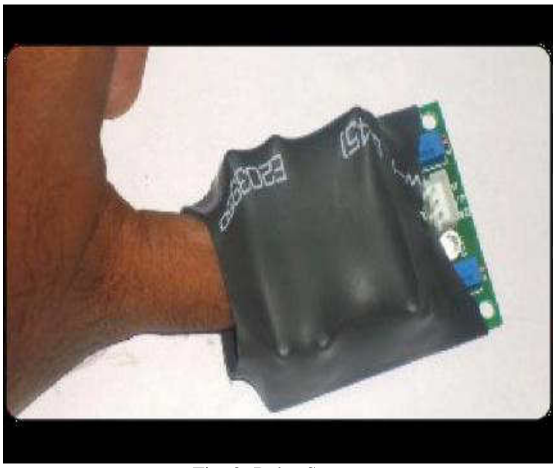


Fig. 2: Pulse Sensor

D. Temperature Sensor

Due to unavailability of body temperature monitoring sensor during our experimentation we have utilized environmental temperature monitoring sensor [LM35]. The LM35 [6] is a precision integrated-circuit temperature sensor, whose output voltage is linearly proportional to the Celsius (Centigrade) temperature. The LM385 is a voltage regulator diode. Analog to digital converter (ADC) is used for converting analog value got as input. ADC08031 is used as ADC, which gives a digital value with a serial output with digital temperature transmitter circuit. Its full-scale is up to +128 deg C. The output can be taken out by the following pins: CLK, ENA, respectively enable client-side and the clock.

These sensors use a solid-state technique to determine the temperature. They use the fact as temperature increases, the voltage across a diode increases at a known rate. (Technically, this is actually the voltage drop between the base and emitter - the Vbe - of a transistor. By precisely amplifying the voltage change, it is easy to generate an analog signal that is directly proportional to temperature.

IV. SIMULATION TOOLS

A. MAX232

The MAX232 [8] is an integrated circuit that converts signals from an RS-232 serial port to signals suitable for use in TTL compatible digital logic circuits. The MAX232 is a dual driver/receiver and typically converts the RX, TX, CTS and RTS signals.

The drivers provide RS-232 voltage level outputs (approx. ± 7.5 V) from a single + 5 V supply via on-chip charge pumps and external capacitors. This makes it useful for implementing RS-232 in devices that otherwise do not need any voltages outside the 0 V to + 5 V range, as power supply design does not need to be made more complicated just for driving the RS-232 in this case.

The receivers reduce RS-232 inputs (which may be as high as ± 25 V), to standard 5 V TTL levels. These receivers have a typical threshold of 1.3 V, and a typical hysteresis of 0.5 V.

B. Microsoft Visual Studio

Microsoft Visual Studio [8] is an integrated development environment (IDE) from Microsoft. It can be used to develop graphical user interface [GUI] applications supported by Microsoft Windows, Windows Mobile, Windows CE, .NET Framework, .NET Compact Framework.

The GUI is created using Visual C# studio 2005, to display the measured temperature and pulse of the patient. The code written also determines the abnormalities of temperature and heart beat of a patient.

C. MSSQL 2005

MSSQL 2005 is used for storing biosignals sensed by the sensors. After a body sensor generates a measurement, it transmits sensed data to the base station. The base station (used here is personal computer) can be located within radio range of sensors. After data cleaning and summarizing, it is stored in the database (MSSQL) of the base station. The database will store all kinds of data such as sensed body/environment/ location/time data.

The doctor can check the patient condition by accessing the data from the server via the interface provided to him. Here the base station itself acts as server for receiving data from sensor and sending to caretaker or for remote monitoring by doctors.

V. SIMULATION RESULTS

A. Serial Port Communication

features for serial communication. The framework provides System IO. Ports namespace. The new framework provides classes with the ability to access the serial ports on computer, and to communicate with serial I/O devices. We are using RS 232 C standard for communication between PCs.

Initially port number and other properties are fetched from application configuration file to allow the user to set the appropriate properties (i.e., Instantiate the communications port with data stored in application configuration file settings).

Properties are set as follows:

“Port" value="COM4","BaudRate" value="9600"
“Parity" value="None","DataBits" value="8"

The read timeout is set to 1000 milliseconds and null bytes are ignored when transmitted between the port and the receive buffer. String using the *SerialPort.ReadExisting()* method is used to read data into a byte buffer and generate an event when you detect a valid packet in the data. And store data into string variable.

The proposed system is used for monitoring the heart rate of patient and temperature; output will be two parameters (Heart rate and Temperature). The data that has been received from the serial port is then analyzed and configured to the desired format to display in the User Interface. The incoming data is in the Temperature, pulse rate format, e.g.: 035072. Where data is split and displayed in desired format by editor. The # is used as delimiter to separate the two reading of body temperature and pulse rate.

B. Fuzzy Logic

Fuzzy-logic theory [4] has been mainly applied to industrial problems including production systems. There has been significant attention given to modeling scheduling problems within a fuzzy framework. Several fuzzy logic based scheduling systems have been developed, although direct comparisons between them are difficult due to their different implementations and objectives. In general, a Fuzzy Logic System (FLS) is a nonlinear mapping of an input data vector into a scalar output. Figure 3 depicts a FLS that is widely used in fuzzy logic controllers. A FLS maps crisp inputs into crisp outputs, and this mapping can be expressed quantitatively as y = f(x). It contains four components: fuzzifier, fuzzy rules, inference engine, and defuzzifier.

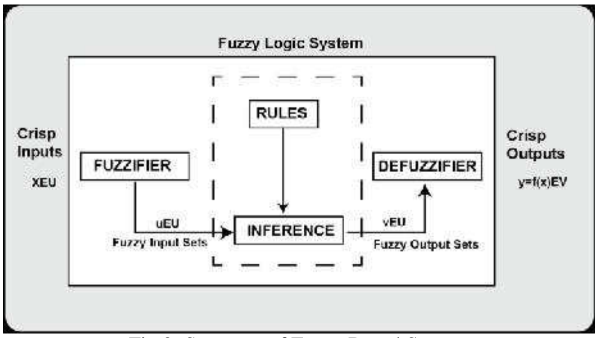


Fig 3: Structure of Fuzzy Based System

Here, we have designed a FLC system for health monitoring services, which is one of component in our pervasive computing prototype health status. The FLC system receives context information from sensor (sensor data stored in data base) equipments as the inputs of the FLC and the fuzzification module converts inputs into fuzzy linguistic variable inputs.

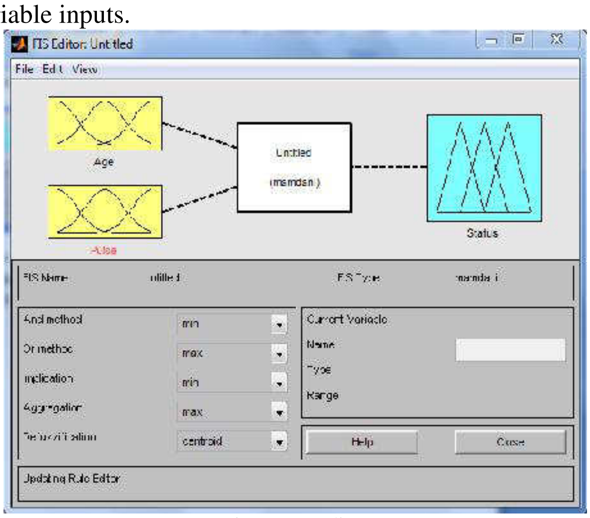


Fig 4: FIS Editor

On analyzing the data requirements of health services, four linguistic variables were defined, representing the physical sign of patient. The membership functions of these input parameters of the fuzzy logic are illustrated in Fig.4. The labels in the fuzzy variables are presented as follows.

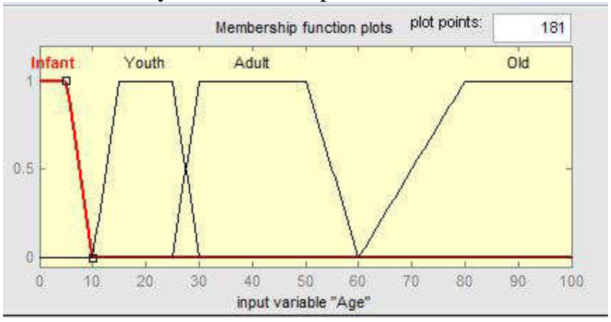


Fig 5: Membership Functions of Input Age

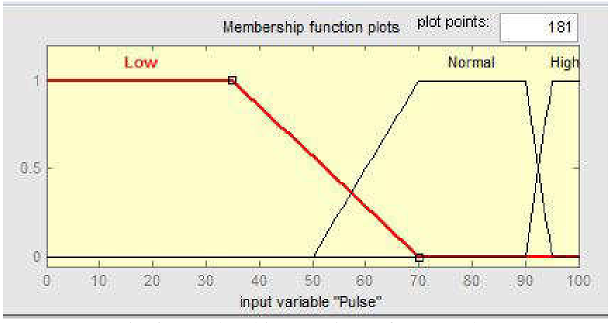


Fig 6: Membership Functions of Input Pulse

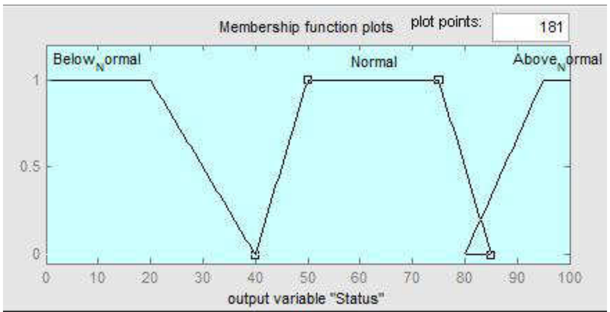


Fig 7: Membership Functions of Output Status

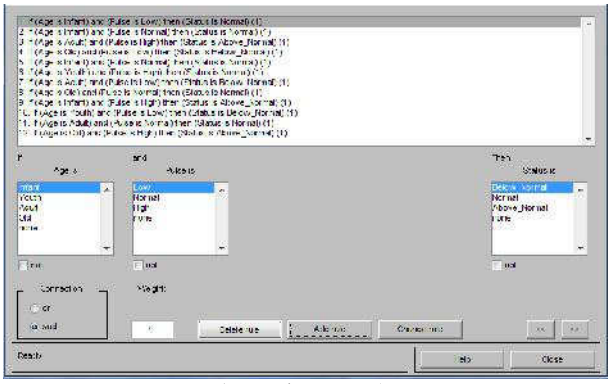


Fig 8: Inference Rules

Status values are evaluated based on rules given below:

If (Age is Infant) and (Pulse is Normal) then Status is Normal

If (Age is Infant) and (Pulse is High) then Status is Normal

If (Age is Youth) and (Pulse is High) then Status is Above Normal

If (Age is Youth) and (Pulse is Low) then Status is Below Normal.

A sample fuzzy calculation at a value of context information point is described in Figure 9.

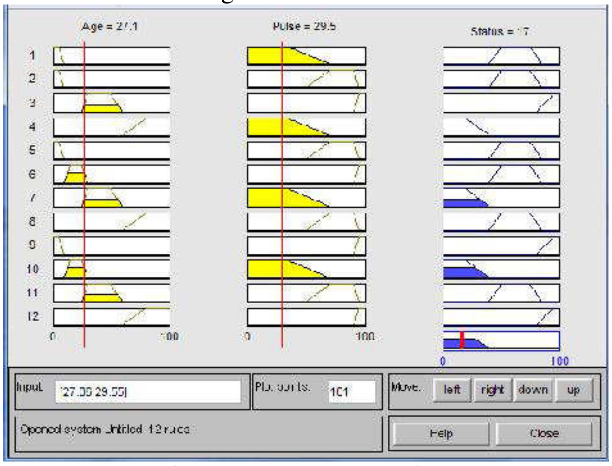


Fig 9: Sample Fuzzy Calculations

To calculate status values for all the health data entries from database, first database connectivity is done using function *"conn()"* here default mssql server connection is made where the health data is stored using serial port communication. After fetching the data from database using *"fetch()"* function for fuzzfication it is evaluated using *evalfis()* and displayed. The final output is stored into text file with first column input pulse, second column as defuzzified output and third column as status.

conn = database('.','','');

curs=exec(conn,'SELECT DISTINCT HBeatRate, Age FROM Demo1.dbo.StatusData');

Out =evalfis(t,b);

Status values are evaluated for all the health data entries stored in database. The status value is to be evaluated based on both age and pulse sensed. Here, status is evaluated by sensing the pulse value for different age groups. Keeping age as constant and sensed pulse sensor is attached to one person at a time. In the figure 9 shows defuzzified value for patient with age 32, where variation is shown in status/output if only pulse is considered as input, this is shown in figure 14 where the first graph is taken by considering pulse as only input to evaluate status of patient whereas graph second considers both age and pulse as input parameter as a result there is variation in the output (i.e. graph). As in case of infant, if the pulse of new born is 95 then it is considered as normal with respect to age, if only pulse is taken as output without age then status may be wrong in some cases.

VI. CONCLUSION AND FUTURE SCOPE

This paper demonstrates the use of wearable and implantable Wireless Body Sensor Network as a key infrastructure enabling unobtrusive, constant, and ambulatory health monitoring. This new technology has potential to tender a wide range of assistance to patients, medical personnel, and society through continuous monitoring in the ambulatory environment, early detection of abnormal conditions, supervised restoration, and potential knowledge discovery through data mining of all gathered information.

In this work, we have analyzed the on-line health monitoring system of temperature and pulse of patient using SMS and feedback/prescription from doctor given through mail. Any abnormalities in health conditions are informed via SMS to the doctor/emergency ambulance. Our study in this paper also demonstrates that fuzzy logic can be applied to health monitoring system using wireless sensor network for deciding final observation of patient health status based on bio-signal reading.

This paper proves that wireless sensor networks can be widely used in healthcare applications. We believe that the role of Body sensor networks in medicine can be further enlarged and we are expecting to have a feasible and proactive prototype for wearable / implantable WBSN system, which could improve the quality of life.

REFERENCES

[1] McCollum, Ph.D. "Using the iPhone and iPod Touch for Remote Sensor Control and Data Acquisition", IEEE 2010.

[2] [2] Josephine Selvarani, "ONLINE HEALTH MONITORING SYSTEM USING ", presented publication in International Journal on Computer Science and Engineering" Karunya University, Coimbatore, 2011.

[3] [3] Orlando R. E. Pereira, Paulo A. C. S. Neves, Joel J. P. C. Rodrigues, "Mobile Solution for Three-tier Biofeedback

[4] Data Acquisition and Processing ", presented publication in the IEEE "GLOBECOM" proceedings, Portugal,2008.

[5] [4] Begonya Otal, Luis Alonso and Christos Verikoukis, "A new MAC Approach in Wireless Body Sensor Networks for Health Care ", presented publication in the IEEE "GLOBECOM" proceedings, Portugal,2008.

[6] [5] Begonya Otal, Luis Alonso and Christos Verikoukis, A new MAC Approach in Wireless Body Sensor Networks for Health Care, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.

[7] [6] http://www.nskelectronics.com/files/lm35.pdf

[8] [7] Heartbeat sensor , Available: http:// http://www.nskelectronics.com/heart_beat.html

[9] [8] Sana Ullah, Pervez Khan et.al , A Review of Wireless Body Area Networks

[10] [9] A Review of Wireless Body Area Networks for Medical Applications, arXiv:1001.0831v3 [cs.NI] 3 Aug 2010

[11] [10] L. Zhong, M. Sinclair, and R. Bittner, "A Phone-centered Body Sensor Network Platform Cost, Energy Efficiency & User Interface", presented at International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2006), Cambridge, MA, USA, 2006.

[12] [11] S.-L. Chen, H.-Y. Lee, C.-A. Chen, C.-C. Lin, and C.-H. Luo, "A Wireless Body Sensor Network System for Healthcare Monitoring Application", presented at IEEE Biomedical Circuits and Systems Conference (BIOCAS 2007), Montréal, QC, Canada, 2007.

[13] [12] A. A. Reeves, J. W. P. Ng, S. J. Brown, and N. M. Barnes, "Remotely Supporting Care Provision for Older Adults", presented at International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2006), Cambridge, MA, USA, 2006.

Prediction through Classification and Clustering TECHNIQUES Using "R"

Ramandeep Kaur Grewal
Information System Security
Concordia University, Montreal
ramandeep121212@gmail.com

Pushpinder Singh Chattha
Information System Security
Concordia University, Montreal
pushpinderchattha@gmail.com

***Abstract*— In this paper, we have investigated various methods of data mining and introduced classifiers to predict the values in a dataset which is essential for data mining algorithms. Decision Tree and SVM are used as classification method for prediction. DBSCAN clustering method is also implemented for prediction and then finds the best method among these supervised and unsupervised methods. Different classifiers show different results on the dataset. In this paper, we also performed comparison between implemented classifiers used to extract knowledge and comparison of knowledge extracted from both methods. All these algorithms are compared based on the accuracy of data values.**

Keywords—Classification and Clustering, Data Mining,K-Means Clustering,Decision Tree Induction,Support Vector Machines (SVM),Density-Based Clustering (DBSCAN)

I. INTRODUCTION

Data mining is a technique which is used to extract the knowledge from the huge amount of data**.** To make the data useful, we need to apply various data mining functions like clustering, association, classification, outlier detection, prediction and others. Usually many organizations involve extracting knowledge through surveys or from various databases to collect appropriate information for their specific needs by collecting the views of community or users which are interested for their future research point of view. Then they apply techniques to generate rules and extract meaningful information from the data. To predict the important information, the various classification and clustering methods are used for prediction. Firstly, we implement classifiers with the training data and later it is used to predict the values [1].

A. *What is Classification?*

Classification is a data analyzing method which is basically used for categorization of data according to various class labels. Classification is the example of supervised learning which is totally based on known class labels of training data (for the initially provided data). The target variable and data values are initially provided to the data. Classification is used in many fields like decision making, prediction, knowledge mining and others [1][4].

i. *Goal of Classification*

The goal of performing classification on data is to generate a model based on known data values and check the accuracy rate of the results. Accuracy means satisfaction of results. If the results are true then rate of accuracy is 100%. Then we can implement our classifier or model for new data sets also. But before classification we need to prepare the data for it by performing data cleaning, selection of features by relevance analysis and data transformation also. It has number of applications like medical diagnosis, credit approval, fraud detection, product marketing and web page categorization [5] [6].

ii. *Difference between Classification and Prediction?*

Classification is the process of examining the training set from the historical data and class label is known. Prediction is the presumption or forecast for the values which are missing or unknown and it is particularly used for continuous values. Class attributes are numerical attributes in prediction for example: if we want to predict sales for a particular company in the future.

iii. *Various Types of Classification*

For the evaluation among various methods of classification we check the accuracy, speed, robustness, efficiency in databases and the knowledge provided by the particular model [7].

a) Decision Tree Induction

b) Support Vector Machines(SVM)

c) Rule Based Classification

a) Decision Tree Induction

It is one of the main methods of classification by which prediction of data values can be done easily and more efficiently. In decision tree induction, the output is shown in the format of nodes, leaves and root. There is one root node which is at the top of the tree and others are a leaf node which belongs to particular class. The main advantage is that no need to require domain knowledge while constructing a tree. The accuracy is good and easy to understand. It can handle multidimensional data and appropriate for knowledge discovery. There are further many methods which are used for the classification of nodes or attributes in decision tree induction like CART, ID3 and C4.5. Each method has its own significance and drawbacks [1][5].

b) Support Vector Machines (SVM)

This technique is used for both linear as well as non-linear data. In SVM, data is partitioned into two groups or classes through hyper plane which are also known as decision boundary. This boundary is searched by using support vectors. The strategy for this is to maximize the margin for minimizing the error. This method is good for both numeric prediction and classification in high accuracy [7].

c) Rule Based Classification

This type of classification is also an effective method to classify data using IF-THEN rules. A rule based classifier uses rules is in the form

IF condition THEN conclusion

For example: rule 1 is if an income of a person is high and credit_rating is fair then he/she must buy the computer. So basically rules are used by classifier for testing the new data with the previous training data. It is just like:

R1: IF income=high AND credit_rating=fair THEN buys_computer=yes [6].

There are some other popular classifications like associative classification, frequent pattern based classification and instance based methods. We used SVM and Decision Tree for our implementation.

B. Clustering

As classification is a supervised learning process clustering is an unsupervised learning because previously class labels are not known to us. In clustering, data within a class must have similar properties with each other and unrelated to the data which is in other group or class. It has typical applications like preprocessing step and stand-alone tool [7][8].

i. *Methods used for Clustering*: In our research paper, we use only two types of clustering:

1. K-Means
2. DBSCAN

a) *K-means Clustering*

K-Means clustering comes under the partitioning method of clustering. In this method, clusters are firstly divided into groups and then select the centers or mean points of the clusters in every group. The whole clusters are assigned according to their nearest centroids. This is an efficient method because of its time complexity is linear which is O(tkn), where t is number of iterations, k is number of clusters and n is the total number of objects.[7][11]

b) *DBSCAN (density-based spatial clustering of applications with noise)*

This method of clustering is totally based on the density of the clusters. There are various advantages of this clustering technique like it is used to implement clusters of any shape which means useful for discovery of arbitrary shape of clusters. The other thing is that it is more robust to noise or outliers. The whole process performs in only one scan. It is based on two parameters like and Minpts. The is the maximum radius of the neighbourhood and Minpts are the minimum number of points in neighbourhood. It is incremental in nature [6].

II. RELATED WORK

Data mining is in huge demand nowadays due to cause of abundant data. It is useful in many areas like science, bioinformatics, text mining, target marketing and many more. A literature survey showed that many researches and analysis already have been done in this field. Jiawei et al. proposed a classification method using two algorithms MedGen and MedGenAdjust which explained generalization and decision tree induction in large databases [14]. Huang et al. demonstrated a general criteria for the creation of decision trees in classification. In the paper, the main four factors are discussed that are based on their attributes, sub-attributes, samples and number of classes [15]. Jacob et al. in 2012 implemented the patterns of classification in clinical data through various data classification techniques like Quinlan's C4.5, CS-MC4, C-SVC, SVM and others. Then found the maximum accuracy with the random tree algorithm and Quinlan's C4.5 [16]. Then Chang et al. explained a new clustering method for large and high dimensional data which is cell-based clustering method. They proposed two features in the clustering method. First is for the creation of small number of cells and other is to implement the approximation technique for fast clustering [17]. Chandola et al. provided a detail knowledge for existing outlier detection techniques through survey [20].

III. PROBLEM STATEMENT

Prediction through both classification and clustering has been the most technologically challenging problem in the data mining processes in research area. We can make the data meaningful and can perform prediction about the new data belongs to which class, with the help of accurate prediction of values. Various supervised and unsupervised techniques are used for prediction in this work and every technique has its own method for predicting data values. We were performed with these different techniques and find the best method among them according to their quality. We found the appropriate method which is having high accuracy and the better results than the previous classifiers. Then detection of noise or outliers was done to improve the quality, consistency and robustness of data which is also an important issue in data mining [21].

IV. PROPOSED METHOD OR SOLUTION

We apply three methods for prediction that are SVM and

Decision Tree as classification and DBSCAN as clustering.

A. *Dataset Selection*

We used seed dataset for implementation which is having geometric parameters of three different varieties of wheat that are Kama, Rosa and Canadian. There are eight attributes in the dataset from them seven attributes are numerical and only one attribute is categorical that is Names. Area, Perimeter, Compactness, length of kernel, width of kernel, asymmetry coefficient and length of kernel groove and Names are the various attributes in our seed data set. It has 210 numbers of instances and multivariate in nature [23].

B. *Software used*

For implementation, we used R-Software because it provides environment for advances graphical data analysis and having flexible analysis toolkit for statistics. This software has huge amount of packages and having the quality of addition of code and new packages by anyone. We have learned R, R-Studio and RGUI (Rattle). But we choose R-Studio due to cause of its more advantages like better visualization of results, command line and can easily manage multiple working directories.

C. *Frame of Reference*

To make the comparison between various prediction techniques, we have to measure the Quality parameter. To analyze the quality, we measured accuracy of various methods in this work which is based on the correct predicted values to the total number of tuples.

Accuracy = (Predicted values/Total number of tuples)*100

D. *Methods used for Prediction*

Firstly, we convert the dataset format into CSV format so that our tool can understand and use the dataset. After that we load the data into the software with the help of this command.

```
ds<read.csv("/seeds_dataset.csv")
```

Then the whole data is divided into two phases that are training and testing data. 70% of data is used for training and

30% is for testing like

```
var1 <- sample(2, nrow(ds), replace=TRUE, prob =

c(0.7,0.3)) training <- ds[var1==1,]

testing <- ds[var1==2,]
```

i. *Prediction using Decision Tree Classification*

Then we used two classification techniques for prediction that are Decision Tree Induction and SVM. In Decision Tree, we use library "party" package for conditional tree. Then we apply ctree() function on the training data and generates the tree with this code:

```
library(party)

formula <- Name ~ Area + Perimeter + Compactness

+ Length_of_kernel + Width_of_kernel +
Asym_coefficient + Length_OF_groove
seeds_ctree <- ctree(formula, data=training)
```

ii. *Prediction using SVM Classification*

After the generation of tree on the training data, we made other classifier for prediction that is SVM classifier. For this, we apply "e1071" library function and implement our classifier on training data. After this, it perform prediction on the testing tuples and the code for this is:

```
library(e1071) index <-
1:nrow(ds) svm.model <-
svm(formula)

svm.pred <- predict(svm.model, testing)
```

iii. *Prediction using DBSCAN*

Then we apply DBSCAN clustering technique after performing prediction with classification methods. For this, "fpc" package is used and then seed function is written for getting the same results with the clusters. Then we set 0.7 eps and 4cm Minpts.

```
ds.dbscn<-dbscan(ds2, eps=0.7, MinPts=4)
```

```
myPred<- predict(ds.dbscn, ds2, ds3)

plot(ds.dbscn, ds2)

plot((ds2[c(1,7)]), col=1+ds.dbscn$cluster)

points((ds3[c(1,7)]), pch="*" , col= 1+myPred, cex=3)
```

V. EXPERIMENTAL RESULTS

With the help of R software, we evaluate results by using various classifications and clustering techniques. As we explained earlier the whole functioning and coding of all the algorithms. Now, the experimental results are shown below by using various methods:

A. *Prediction through Decision Tree Induction*

In this method, we implement decision tree using ID3 calculation. It classifies the attributes or nodes according to ID3 method which chose area attribute as a root node and has been shown the output of leave nodes in the form of bar graph in our implementation. Then the tree is implemented on the training data. The main advantage of this algorithm is the interpretability of results it explains the results in a good and efficient manner so that user can understand the results more easily. The tree is shown in the next figure 1[23].

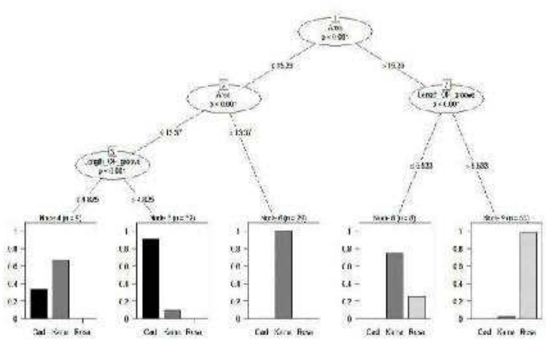


Fig. 1. Implementation of Decision Tree

And the quality of this decision tree is measured in terms of accuracy which is 98% by using this method as shown below:

```
> ##########Confusion Matrix Showing Counts########
>
> table(testing$Name,test1pred, dnn=c("Actual", "Predicted"))
      Predicted
Actual Cad Kama Rosa
  Cad   20    0    0
  Kama   1   22    0
  Rosa   0    0   16
>
>
>
> #########Confusion Matrix Showing Percentage#####
>
> round(100*table(testing$Name,test1pred,dnn=c("Actual","Predicted"))/length(test1pred))
      Predicted
Actual Cad Kama Rosa
  Cad   34    0    0
  Kama   2   37    0
  Rosa   0    0   27
```

Fig. 2. Accuracy in Decision Tree

B. *Prediction through SVM Classification*

This method is implemented for both linear and non-linear data. It shows 65 support vectors in the output as in figure 3 [23].

```
> print(svm.model)

Call:
svm(formula = Name ~ Area + Perimeter + Compactness + Length_of_kernel + Width_of_kernel +
    Asym_coefficient + Length_OF_groove, data = training)


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  1
      gamma:  0.1428571

Number of Support Vectors:  65
```

Figure 3 Model of SVM Classification

This method shows 97% accuracy with the testing data which is given in figure 4.

```
>
> table(testing$Name, svm.pred, dnn=c("Actual", "Predicted"))
      Predicted
Actual Cad Kama Rosa
  Cad   20    0    0
  Kama   0   21    2
  Rosa   0    0   16
>
> round(100*table(testing$Name, svm.pred, dnn=c("Actual", "Predicted"))/length(svm.pred))
      Predicted
Actual Cad Kama Rosa
  Cad   34    0    0
  Kama   0   36    3
  Rosa   0    0   27
>
```

Fig. 4. Accuracy in SVM

C. *Prediction through DBSCAN Clustering*

By using this unsupervised method, we perform prediction on the whole data and find out the number of clusters which are implemented on the whole numerical attributes. The output is explained by this figure 5. Here, we got five clusters after prediction.

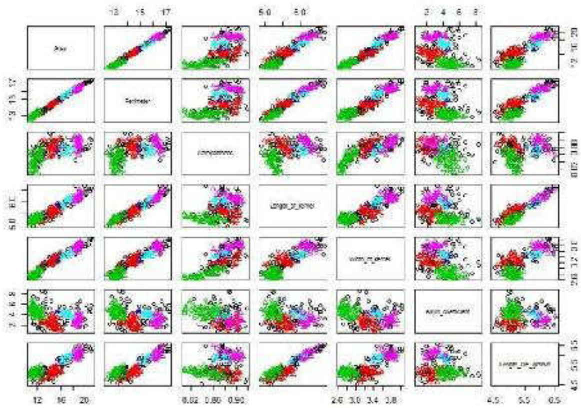


Fig. 5. View of DBSCAN clustering

DBSCAN predicts the clusters and shows which cluster belongs to which class and also explain the mapping of clusters between various classes. In our prediction with this method, it shows 6 columns as in the figure 6 from them 1 to 5 are clusters.

```
> table(ds$Name[id], myPred)
      myPred
        0  1  2  3  4  5
  Cad   4  1 26  0  0  0
  Kama  6 23  4  0  0  0
  Rosa  8  1  0  1 10 16
```

Fig. 6. Prediction of clusters using DBSCAN

We also evaluate performance of clustering with two dimensions of data as we explain in next figure 7. The clusters for width_of_kernal and length_of_kernal are shown only. The column is 5, 4 which is shown as fifth row for length_of_kernal and fourth column for width_of_kernal [23]

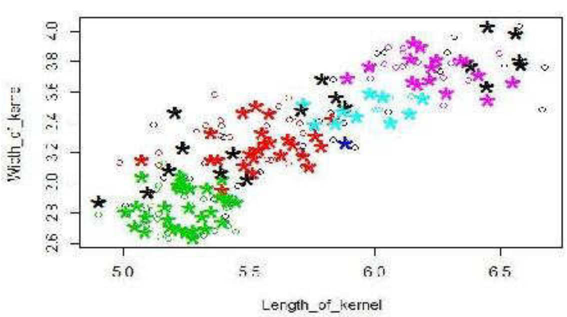


Fig. 7. DBSCAN clustering on two attributes

VI. CONCLUSION

In this paper, we compared various classification and clustering techniques for prediction. In supervised and unsupervised algorithms that are decision tree, SVM and DBSCAN, we found that the accuracy of decision Tree is better than other algorithms under experiment. In advance, DBSCAN does not require specifying number of clusters. But DBSCAN cannot well cluster the datasets which are having large variation in densities because it is difficult to choose appropriate min-points and epsilon combination for all clusters. In future, we will try new classifiers which could provide 100 percent accuracy and even can try another new technique for prediction of training data using classification and clustering.

VII. ACKNOWLEDGEMENTS

Our special thanks to DR. Benjamin Fung who provides technical support and his guidance during the project.

REFERENCES

[1] Nejad S.K., Seifi F., Ahmadi H., and Seifi N. Applying Data Mining in Prediction and Classification of Urban Traffic. World Congress on Computer Science and Information Engineering, 2009, 674-678.

[2] Heckerman, D. Bayesian Networks for Data Mining. Data Mining andKnowledge Discovery 1, 79–119, year 1997.

[3] Phyu T.N. Survey of Classification Techniques in Data Mining. Proceedings of the International Multi Conference of Engineers and Computer Scientists 2009, Hong Kong, Vol I, IMECS 2009.

[4] Ozekes S. and Camurcu A.Y. Classification and Prediction in a Data Mining Application. Journal of Marmara for Pure and Applied Sciences, Marmara University, Printed in Turkey, 18 (2002) 159-174.

[5] Chandra B., Mazumdar S., Arena V., and Parimi N. Elegant Decision Tree Algorithm for Classification in Data Mining. Proceedings of the Third International Conference on Web Information Systems Engineering, IEEE 2002.

[6] Fung, Benjamin, “Lecture Notes: INSE 6180 Security and Privacy Implications of Data Mining”, 2013.

[7] Advantages and Disadvantages of R. Web. 20 May 2013. <http://analyticstrainings.com/?p=101

[8] Chang J. and Jin D. A New Cell-based Clustering Method for Large, High-Dimentional Data in Data Mining Applications. ACM 2002, p. 503-507.

[9] Begum S.H. Data Mining Tools and Trends – An Overview. International Journal of Emerging Research in Management &Technology, February 2013.

[10] Zhang G.P. Neural Networks for Classification: A Survey. IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications And Reviews, November 2000, Vol. 30, No. 4.

[11] Wagsta K., Cardie C., Rogers S., and Schroedl S. Constrained K-means Clustering with Background Knowledge. Proceedings of the Eighteenth International Conference on Machine Learning, 2001, p. 577-584.

[12] Inside-R. A Community Site for R. Web. 20 may 2013. <http://www.inside-r.org/why-use-r>

[13] Zhao Y. R and Data Mining: Examples and Case Studies. Web. 20 may 2013. <http://www.rdatamining.com/docs/RDataMining.pdf> Published by Elsevier in December 2012.

[14] RDataMining.com: R and Data Mining Web 12 May 2013. <http://www.rdatamining.com/r>

[15] Velu C.M., and Kashwan K.R. Visual Data Mining Techniques for

Classification of Diabetic Patients. Third IEEE International Advance Computing Conference (IACC), 2013, p. 1070-1075.

[16] Huang Y. and Hoa V.T.T. General Criteria on Building Decision Trees for Data Classification. ICIS, November 2009.

[17] Jacob S.G. and Ramani R.G. Mining of Classification Patterns in Clinical Data through Data Mining Algorithms. International Conference on Advances in Computing, Communications and Informatics (ICACCI-2012), p. 997-1003.

[18] AI Computer Vision. 23 June 2010. Web. 02 June 2013. <http://blog.samibadawi.com/2010/06/orange-r-rapidminer-statistica- and-weka.html>

[19] Dalal M.A. and Harale N.D. A Survey on Clustering In Data Mining. International Conference and Workshop on Emerging Trends in Technology (ICWET), 2011, p. 559-562.

[20] Chandola V., Banerjee A. and Kumar V. Outlier Detection: A Survey.

[21] ACM Computing Surveys (2009). Web http://www.mendeley.com/catalog/outlier-detection-survey/

[22] Suguna N. and Thanushkodi K.G. Predicting Missing Attribute Values Using k-Means Clustering. Journal of Computer Science, 2011, p. 216-224. The R Project for Statistical Computing. Web. 15 May 2013. <http://www.r-project.org/index.html >

[23] UCI Machine Learning Repository. Seeds Dataset. Web. 24 may 2013. <http://archive.ics.uci.edu/ml/datasets/seeds>

[24] Han J., Kamber M., and Pei J. Data Mining: Concepts and Techniques, 3rd Edition. 2011.

Use of Fuzzy Inference System and ANFIS for Diagnosis of Various Diseases

Rupinder Kaur
Department Of Electronics and Communication
Punjabi University
Patiala, Punjab, India
rupinderkaur334@gmail.com

Amrit Kaur
Department Of Electronics and Communication
Punjabi University
Patiala, Punjab, India
amrit_aks@yahoo.ca

*Abstract—***Fuzzy inference systems (FIS) have been applied in recent years to many medical fields due to their ability to obtain good results. Knowledge represented as a set of explicit linguistic rules, can be more easily understood by people without technical expertise. Additionally, as fuzzy information deals with knowledge that is uncertain, ambiguous or imprecise, it is suitable to be used in medical contexts to represent certain elements as members of sets with some degree of membership. In the medicine area, many expert systems were designed to diagnose and treatment a disease. Fuzzy expert system has found applications in many areas. In the last decade, a major application area of fuzzy expert system includes diagnosis like local anaesthesia, appendicitis, brain tumour, breast cancer, cardiovascular disease, diabetes, ECG signal evaluation, hypertension diagnosis etc has emerged. This paper explains the use of fuzzy logic in diagnosis of various diseases.**

Keywords-FIS,ECG,FES

I. INTRODUCTION

An essential element of the medical profession is making numerous decisions for which doctors rely on gained knowledge and experience. However, it seems necessary for them to have the ability to think logically, to use reasoning, to precisely and clearly express their thoughts and justify the assertions made. Due to the lack of knowledge, there is a probability of error in making decisions. So an intelligent and accurate system is required for making decisions which can be used for diagnosis of various diseases.

The diagnosis of disease involves several levels of uncertainty and imprecision. The best and most precise description of disease entities uses linguistic terms that are also imprecise and vague. To deal with imprecision and uncertainty, fuzzy logic can be used. Fuzzy logic introduces partial truth values, between true and false.

FIS is based on the concepts of fuzzy logic, fuzzy set theory, fuzzy if-then rules, and fuzzy reasoning.

FIS can be defined as a process of mapping from a given input to an output, based on expert knowledge. The knowledge is encoded as a set of explicit linguistic rules, which can be easily understood by people without technical expertise. As fuzzy information deals with knowledge that is uncertain, ambiguous or imprecise, it is suitable to be used in medical contexts to represent certain elements as members of sets with some degree of membership.

A. Fuzzy Inference System

Fuzzy inference system (FIS) is a process of mapping from a given input to an output using the theory of fuzzy sets. Figure 1 shows the fuzzy inference system with five functional blocks.

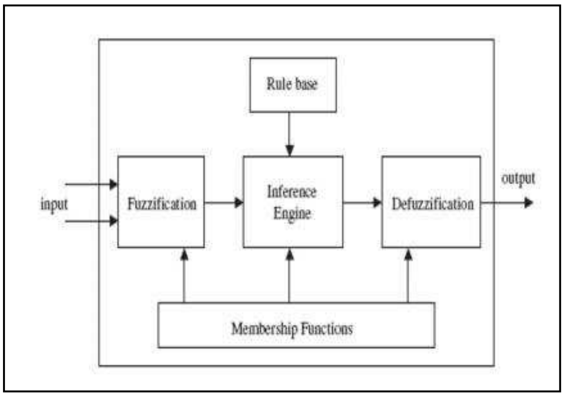


Fig.1. Fuzzy Inference System

In general, the main steps performed in the FIS are as follows:

- The fuzzification is the process in which each crisp input variable is transformed into a membership grade based on the membership functions defined.
- The inference engine then conducts the fuzzy reasoning process by applying the appropriate fuzzy operators in order to obtain the fuzzy set to be accumulated in the output variable.
- The defuzzifier transforms the fuzzy output into a crisp output by applying a specific defuzzification method.

The two types of fuzzy inference system most commonly used are the Mamdani method and the Sugeno method. The difference between these two fuzzy inference methodologies is the specification of the consequent part. In the Mamdani method, consequents are fuzzy sets, and the final crisp output of the Mamdani method is based on defuzzification of the overall fuzzy output using various types of defuzzification methods. In contrast, in the Sugeno method, consequents are real numbers, which can be either linear or constant. The final output is the weighted average of each rule's output.

The fuzzy rule of the first-order Sugeno-style has the form:

IF x is A AND y is B THEN z = ax + by +c

B. Adaptive Neuro Fuzzy Inference System (ANFIS)

ANFIS(Adaptive neuro fuzzy inference system)is used as a teaching method for sugeno type fuzzy systems. Adaptive neuro fuzzy inference system is a feedforward adaptive neural network which implies a fuzzy inference system through its structure and neurons. It is a hybrid neuro fuzzy technique that brings learning capabilities of neural networks to fuzzy inference systems. Using a given input-output data set, the toolbox function anfis constructs FIS whose membership function parameters are tuned using backpropagation algorithm alone or in combination with least square type of method.

The ANFIS architecture contains a five-layer feedforward neural network as shown in Figure 2.

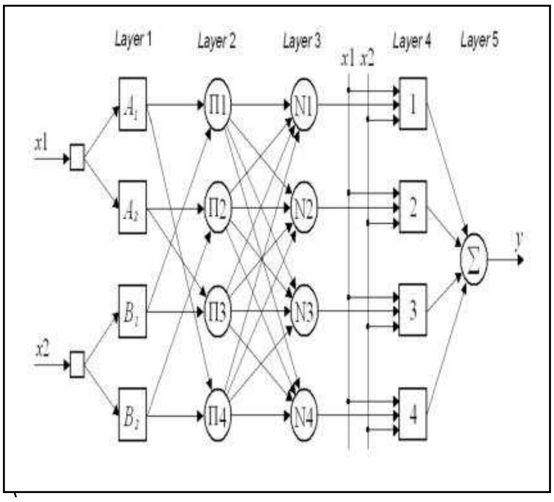


Fig. 2. ANFIS [3]

To update the parameters of the network, a combination of gradient descent and least squares estimator (LSE) is used. ANFIS uses two sets of parameters: a set of premise parameters and a set of consequent parameters. Two passes in the hybrid learning procedure are shown in table 1.

TABLE I.HYBRID LEARNING PROCEDURE

	Forward pass	**Backward pass**
Premise parameters (nonlinear)	Fixed	Gradient descent
Consequent parameters(linear)	Least square estimator	Fixed
Signals	Node outputs	Error signal

II. FUZZY LOGIC IN DIAGNOSIS OF DISEASES

In paper [1], Fuzzy Expert System (FES) is used to diagnose and treat this disease. Blood pressure, BMI, age, heart rate and life-style of the patient are used as input data. Fuzzy rules are developed to determine the risk factor of having high blood pressure. The developed system gave the user, the patient possibility ratio of the hypertension.

This paper introduces the role of fuzzy expert system in medical diagnosis of diseases. Then methodology for the data collection is given. Then fuzzy expert system is developed as shown in fig. 3. [1]

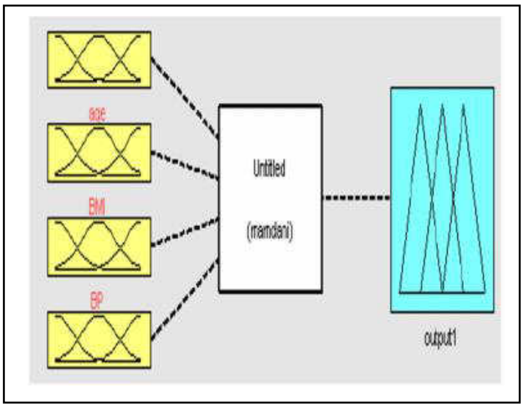


Fig. 3 Fuzzy expert system [1]

Then fuzzy rule base is designed which has 9 rules. The surface result of the output is shown in fig.4. [1]

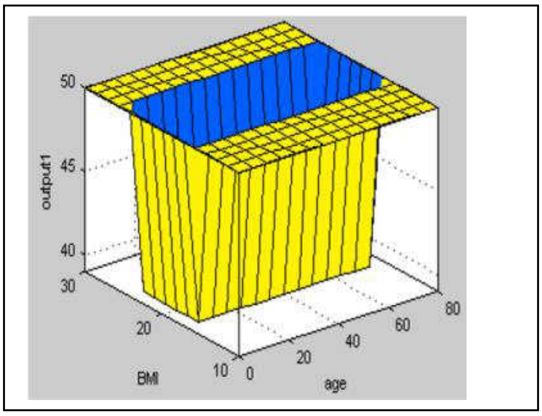


Fig. 4 surface view [1]

In paper [2], a fuzzy expert system is developed to diagnose the hypertension risk for different patients based on a set of symptoms and rules. The input parameters taken are age, body mass index (BMI), blood pressure (BP), and heart rate.

Then a neuro fuzzy system is also designed for the same set of symptoms and rules using three different types of learning algorithms which are Levenberg-Marquardt (LM), Gradient Descent (GD) and Bayesian Resolution(BR) based learning functions. Then this paper compares fuzzy expert system (FES) and feed forward back propagation based neuro fuzzy system (NFS) for hypertension diagnosis. NFS is more appropriate model to measure the risk factor than the fuzzy expert system.

This paper also presents a comparison among the learning functions (LM, GD and BR) where Levenberg-Marquardt based learning function shows its efficiency over the others.

In paper [3], fuzzy logic is used for medical diagnosis regarding the normality of a human function in human brain and the diagnosis of hemorrhage and brain tumor.

This system design the fuzzy logic medical diagnosis control system for the diagnosis of human disease. System uses five input variables Protein, Red blood cells, Lymphocytes, Neutrophils and Eosinophils, and takes three outputs: Normal, Hemorrhage and Brain Tumor.

Then MATLAB simulation is performed by seventeen rules. Then the simulated results are compared with calculated values which are very close values shown in table 2. [3]

This shows that the system is very accurate for the diagnosis of human disease.

TABLE II. COMPARISON BETWEEN CALCULATED AND SIMULATED RESULT [3]

Results	Normal	Haemorhage	Brain Tumor
MATLAB simulation	0.444	0.485	0.575
Calculated values	0.432	0.464	0.6

In paper [4], ANFIS (Adaptive Neuro-Fuzzy Inference System) is used as a diagnosis system on Breast Cancer Diagnosis. The performance of diagnosis system in computation is an important issue as well as the correctness of the output from the inference system. The paper is focussed on improving the computational performance by input data selection. For these purpose three methods, genetic algorithm, decision tree and correlation coefficient computation with ANFIS are used.

Experimental results indicate that ANFIS shows even better performance than other systems with input reduction methods.

In next paper [5], for the diagnosis of the four common types of Aphasia (language disability) more efficiently, an adaptive Neuro-Fuzzy inference system (ANFIS) is proposed. The proposed method in this paper is compared with a hierarchical fuzzy rule-based structure and a back propagating feed-forward neural network. Results show that proposed method achieves better accuracy compared to other methods.

In paper [6], ANFIS is used to determine an estimation method to predict glucose rate in blood which indicates diabetes risk. Initially continuous values in the dataset are converted into to fuzzy values.

After preprocessing dataset, ANFIS method and MLR method are applied. Results show that learning duration of ANFIS is much shorter than MLR's duration. When a more sophisticated system with a huge data is imagined, the use of ANFIS instead of MLR would be more useful to overcome faster the complexity of the problem.

In training of the data, ANFIS and MLR gave quite similar results with standard error. However, when the trained parameters were applied to checking data, standard error of ANFIS is smaller than that of MLR. This shows that ANFIS is a better and faster learning method than MLR. For a system which contains fuzzy inputs and output, ANFIS is better system than MLR for diabetes diagnosis.

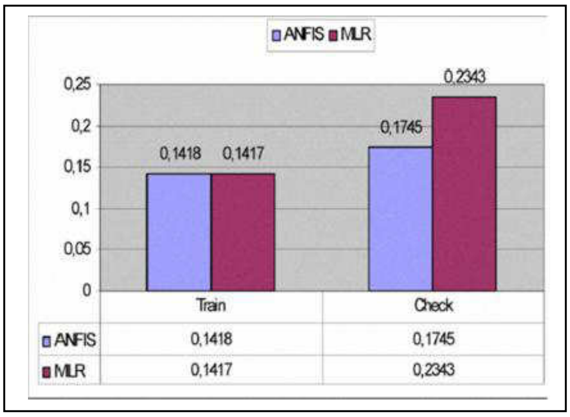


Fig.5.Root Mean Squared Error [6]

III. CONCLUSION

This paper explains the use of fuzzy logic, fuzzy inference system and adaptive neuro fuzzy inference system along with various other techniques for the diagnosis of various diseases. This is very efficient and accurate method for medical diagnosis.

REFERENCES

[1] A.A. Abdillah, Z.Zakaria And N.F. Mohammad, Design And Development Of Fuzzy Expert System For Diagnosis Of Hypertension, IEEE second international conference on intelligent systems, modelling and simulation,2011

[2] Sujit Das, Pijush Kanti Ghosh and Samarjit Kar, "Hypertension Diagnosis: A Comparative Study using Fuzzy Expert System and Neuro Fuzzy System", Proceedings of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2013), Hyderabad, July 7-10, 2013, pp. 1-7, DOI: 10.1109/FUZZIEEE.2013.6622434.

[3] Jyh-Shing Roger Jang, ANFIS (Adaptive-Network-Based Fuzzy Inference System), IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, VOL. 23, NO. 3, MAYIJUNE 1993

[4] Faran Baig, Dr.M. Saleem Khan,Yasir Noor,M. Imran, Design Model Of Fuzzy Logic Medical Diagnosis Control System, international journal on computer science and engineering, vol. 3 no.5 may 2011.

[5] Hee-Jun Song, Seon-Gu Lee, Gwi-Tae Park, A Methodology of Computer Aided Diagnostic System on Breast Cancer, Proceedings of the 2005 IEEE Conference on Control Applications Toronto, Canada, August 28-31, 2005

[6] Sepideh Fazeli, Maryam Naghibolhosseini And Fariba Bahrami,An Adaptive Neuro-Fuzzy Inference System For Diagnosis Of Aphasia,IEEE 2008.

[7] Riccardo P. and Stefano C. A Neural Network Expert System for Diagnosing and Treating Hypertension, University of Florence Riccardo Livi, pp.64-71. 1991

[8] Agboizebeta and O. J. Chukwuyeni. Cognitive neuro-fuzzy expert system for hypotension control, Computer Engineering and Intelligent Systems, vol. 3, no. 6, pp. 21-32, 2012.

[9] Aleksandra B., Gilles C., Jacques B., and Marion L., Bayesian Statistics as Applied to Hypertension Diagnosis, IEEE Transactions On Biomedical Engineering. Vol. 38. No.7.JULY 1991

[10] Sylvie C., Sylvie G., Gilles M., and Jean P. S., Statistical and Fuzzy Models of Ambulatory Systolic Blood Pressure forHypertension Diagnosis, IEEE Transactions On Instrumentation And Measurement, Vol. 49, No. 5, OCTOBER 2000, pp.998-1003

[11] M. Neshat, M. Yaghobi, M. B Naghibi, A.Esmaelzadeh,Fuzzy Expert System Design for Diagnosis of liver disorders, Department of Computer Engineering, Azad University of Mashhad Iran 2008 International Symposium on Knowledge Acquisition and Modeling, pp.252-256

[12] S. N.Sivanandam, S. Sumathi and S. N. Deepa, (2007).Introduction to Fuzzy Logic using MATLAB,pp. 304-309.

[13] Michael Negnevitsky, (2004). Artificial Intelligence, A guide to Inteligent Systems, Second Edition,Addison-Wesley.

[14] T. O Lim and Z Morad (1996). Prevalence, Awareness,Treatment and Control of Hypertension in the Malaysian Adult Population: Results from the National Health and Morbidity Survey 1996.

[15] P. Degoulet,C. Devries, and D. Sauquet, “Data and knowledge management integration in hospital information system,” in Towards News Information Systems, Amsterdem, North-Holland, 1989, pp. 149-156.

[16] Novruz ALLAHVERDI, Serhat and TORUN Ismail SARITAS, Fuzzy Expert System Design for Determination of Coronary Heart Disease Risk, International Conference on Computer Systems and Technologies (2007).

[17] National Heart Lung and Blood Institute, U.S Department of Health & Human Services. Retrieved Oct 10, 2010, from http ://www.nhlbisupport.com/bmi. 117

[18] Joseph Cheriyan, Carmel Mceniery, Ian B. Wilkinson, Hypertension, Oxford University Press, 03-Dec-2009

Adaptive Modulation of OFDM by using Back Propagation Neural Network (BPNN)

Parminder Kaur
Dept. of ECE
Guru Nanak Dev University
parminder2090@gmail.com

Kuldeep Singh
Dept. of ECE
Regional Campus GNDU
Sultanpur Lodhi,India

Hardeep Kaur
Dept. of ECE
Guru Nanak Dev University
Amritsar, India

***Abstract*: Adaptive communication is one of method used for high value transmission with a efficient spectrum efficiency and to improve the error performance for upcoming wireless communication systems. Fixed modulation systems uses only one type of modulation technique (order), so in that case we can choose either capacity or performance of the system. Adaptive modulated systems are better than the fixed modulated systems because they change their modulation order according to the present SNR. In this paper, we propose an adaptive modulated OFDM system based on BPNN and then their performance (MSE) and accuracy is evaluated according to the number of neurons in the back propagation neural network.**

***Keywords*: OFDM, Adaptive Modulation, Back Propagation Neural Network, Mean Square Error (MSE), Accuracy.**

I. INTRODUCTION

Orthogonal Frequency Division Multiplexing is one of the major technology used in 4^{th} Generation Systems. In OFDM systems high data stream is divided into several low data streams. After that these streams are modulated over different orthogonal subcarriers. The performance of wireless communication system improves according to the SNR, BER and by adjusting transmission parameter like modulation and code-rate. Transmission parameters of each subcarrier of OFDM are changed by applying the adaptive transmission to the OFDM system. Adaptive modulation is evaluated by using adaptive quadrature amplitude modulation schemes which provides better performance than fixed modulation systems for different channels [1]. In OFDM system by using the adaptive modulation with code rate (convolution code) which maximizes the total capacity and provides proper service for each user according to transmit power and channel conditions as maintaining the transmission's quality of service (QoS) at the receiver side [2].The throughput of an OFDM system is improved by adding adaptive modulation (AD) with turbo coding. Each OFDM block is independently modified according to channel state information (CSI) [3].Channel estimation and compensation based on back propagation neural network in OFDM systems is analyzed, which shows BP neural network is better than LS algorithm with respect to MSE . The MMSE algorithm gives the better performance than BPNN algorithms, But the neural network has lower complexity than the MMSE algorithm [4].The performance of OFDM system under different environment such as Ricean fading & Rayleigh fading with AWGN by using neural network which is better than the conventional channel equalizer structures [5].

II. OFDM SYSTEM MODEL

Orthogonal Frequency Division Multiplexing (OFDM) is one of the current modulation techniques used in order to conflict the frequency-selectivity of the transmission channels, by achieving high data rate without inter–symbol interference (ISI) [6]. OFDM is an advance technique that removes the necessity of very difficult equalization. In a conservative serial data system, the codes are transmitted sequentially with the frequency spectrum of each data allowed to take up the entire available bandwidth. There are more chances that the frequency selective channel response disturbs in a very distinguishing manner the different spectral constituents of the data symbol, hence presenting the inter symbol interference (ISI) [7].The performance of Bit error rate and spectral efficiency of various modulations like QPSK, MQAM is analyzed and then compared with the fixed modulation technique, result shows adaptive modulation is better than the fixed modulation [8].

A neural network is a mathematical tool that can simplify a relationship between input and output data of a system. In this paper we propose that the performance of adaptive modulation of OFDM system can be improved with the help of back propagation neural network (BPNN) to realize an adaptive modulation of the subcarriers. The neural network inputs are the channel gains of all the subcarriers. The structures have input layers, hidden layers and an output layer which is fully connected. Below is a block diagram of BPNN with OFDM communications system structure.

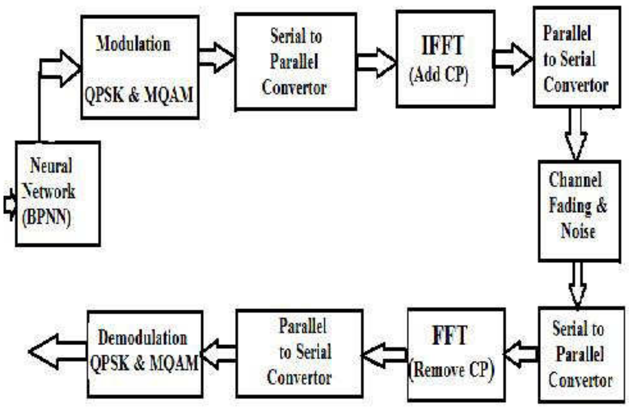


Fig.1. Block diagram of OFDM by using BPNN

III. BACK PROPAGATION NEURAL NETWORK (BPNN)

Neural Networks (NN) are significant data mining tool which is used for classification and clustering. It is an effort to construct machine that will copy brain activities and be able to learn. Neural network usually learns by examples. If neural network is supplied with a sufficient amount of examples, then it provides better classification and discover more new patterns in data Basic neural network is composed of three layers, input layer, hidden layer and output layer. Each layer can have number of nodes and all the nodes can be fully connected to each other node. These connections represent weights on the nodes. There are two types of neural network based on learning technique: supervised where output values are known (back propagation algorithm) and unsupervised where output values are not known (clustering).

Running and training of BPNN consist of a two passes: forward and a backward pass. In the forward pass outputs are measured and then compared with desired outputs.Now error from desired output and actual output are calculated. This error is used in the backward pass to change the weights in the network to reduce the error. Forward and backward pass are repeated until the error is very low. In this paper we have train the BPNN,while training BPNN, we are feeding the BPNN with set of samples that have inputs and desired outputs, out of these some of samples are used for testing the BPNN.Now by Choosing the learning rate and momentum which helps in weight adjustment. Now setting correct learning rate is difficult because if the learning rate is too small, algorithm might take long time to converge and if it is largethen it might diverge. Sometimes in BPNN every weight has its own learning rate.

IV. Back Propagation (BP) Algorithm

Back propagation algorithmis one of the most popular neural network's algorithms. Rojas [2005] claimed that BP algorithm have four main steps. After choosing the weights of the network, the back propagation algorithm is used to calculate the necessary corrections. These steps are as following:

i) Feed-forward computation

ii) Back propagation to the output layer

iii) Back propagation to the hidden layer

iv) Weight updates

This algorithm is stopped when the value of the error function has a very small. This is very uneven and basic formula for BP algorithm. There are some changes proposed by other scientist but this is quite accurate and easy to follow. The last step is weight updates of the network is happening throughout the algorithm.[9]

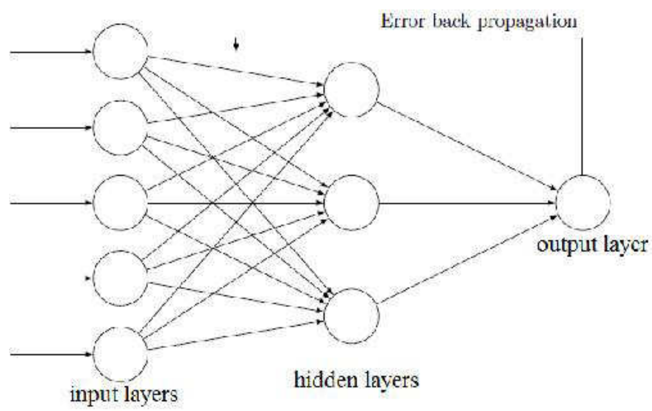


Fig.2. Block diagram of BPNN

V. PROPOSED APPROACH

In this research work ,we have analyzed adaptive modulation of orthogonal frequency modulation(OFDM) by using BPNN ,in this BPNN is trained by using manual data, which includes signal to noise ratio(SNR),Bit Error Rate(BER),Present Modulation(PM),Code Rate(CR) as inputs and Adaptive Modulation(AM) as output represents the best modulation. That manual data contains total 2100 samples out of which 2000 samples are used for training for the BPNN and remaining 100 are used for testing of the BPNN . We have used 1000 epochs for training and testing of BPNN.That all data has been evaluated by using BPNN tool of MATLAB. Below the table shows the values of the Time, MSE and Accuracy according to the variation of the no. of neurons in the BPNN.

Table I. Values of the Time, MSE, and Accuracy with the variation of Neurons

Sr.No.	No. of Neurons	Training Time (Minutes)	MSE (*10^-5)	Accuracy (Percentage)
1	40	1.50	2.97	35
2	80	2.40	4.74	36
3	120	2.57	5.75	42
4	160	4.50	9.7	44
5	200	4.80	13.3	49
6	240	5.20	15.8	50
7	300	7.36	25.2	51
8	360	7.48	26.2	52
9	400	8.12	29.2	58
10	480	8.59	30.6	59
11	500	9.13	31.8	60
12	540	12.51	35.8	65
13	600	15.20	59.3	68

From the above experimental implementation during the training and testing of BPNN, we can analyze:

- As the number of neurons in the network increases, the MSE decreases, thus it improves the performance of the overall system.
- As the no. of neurons increases, the time taken to train and test the BPNN is also increases.
- As the no. of neurons increases the overall accuracy of the system is also increases.
- As the number of training epochs increases, the error is minimum.

Training and testing determine the most suitable variables for this network, working on this particular dataset. The performance of the BPNN is analyzed in terms of its mean-squared error (MSE).

Fig.3.,This graph shows by increasing the no. of neurons, the MSE of the system decreases which further improves the performance of the system. So at no. of neurons.600, it provides minimum MSE.

Fig.4: This graph shows that as by increasing the no. of neurons, accuracy of the system increases.So, at no. of neurons 600, it provides maximum accuracy of 68%.

Fig.5.: This graph shows the variation of time with no. of neurons, as by increasing the no. of neurons training time is also increases. At no. of neurons 600 it takes maximum time of 15.20 minutes.

Fig.6.: This graph shows the MSE of the system at 1000 epochs, which are considerably decreases.

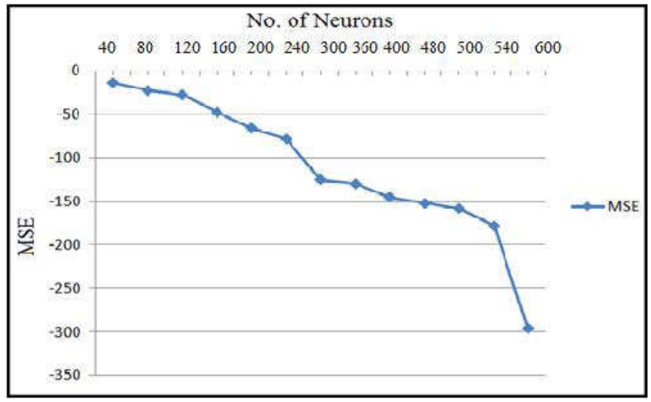


Fig.3. MSE vs No. of Neurons

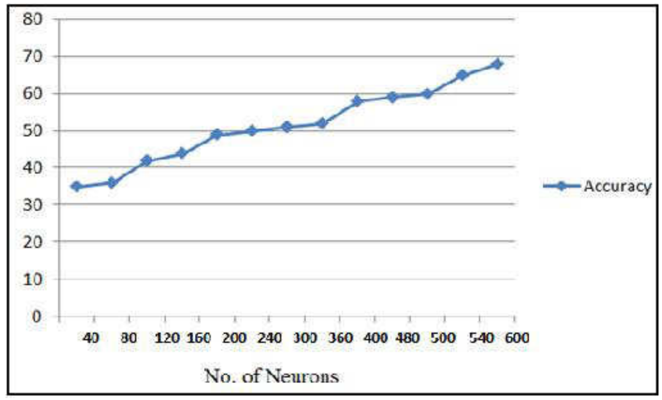


Fig.4. Accuracy vs No. of Neurons

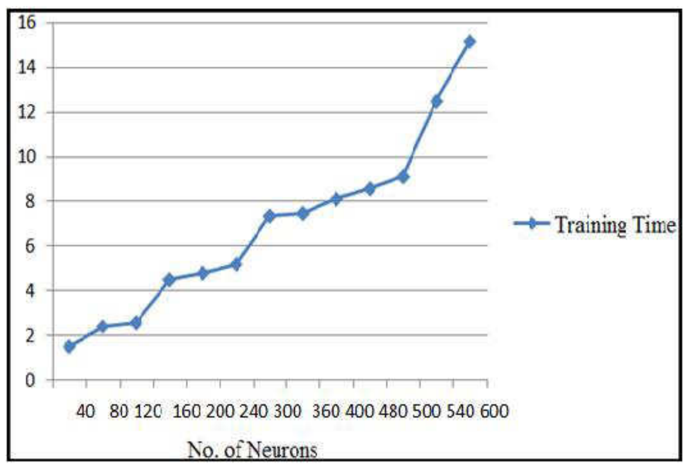


Fig.5. Training Time vs No. of Neurons

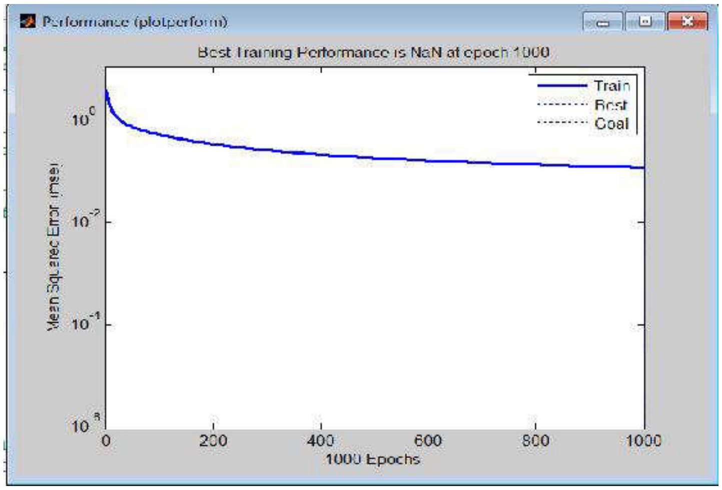


Fig.6. MSE at 1000 epochs

VI. CONCLUSION

This paper proposed the Adaptive Modulation of OFDM based on back propagation neural network ,From all the above result ,we can analyzed by increasing the no. of neurons,MSE decreases and accuracy increases which improves the overall performance of the system, as the system's accuracy increases it gives best modulation for OFDM at that particular time. But we can see at 600 no. of neurons it provides 68% accuracy with minimum MSE at a training time of 15.20 minutes, which is maximum, but here it also takes more time for training and testing.Now this performance can be further improved by using RBF Neural Network.

REFFERENCES

[1] Arne Svensson,"An Introduction to Adaptive QAM ModulationSchemes for Known and Predicted Channels", Proceedings ofthe IEEE on Adaptive modulation and transmission in wirelesssystem | Vol. 95, No. 12, December 2007.

[2] Jaek young Moon and Sangjin Hong "Adaptive Code-Rate and Modulation for Multi-User OFDM System in WirelessCommunications" - IEEE transaction on Vehicular Technologyvol.4, pp-1943-1947, 2002.

[3] Sami Ahmed Haider and Khalida Noori-"Adaptive Turbo Coded OFDM" Journal of Digital Information Management Volume 5, Number 6 December 2007.

[4] Cheng, Chia-Hsin, and Yung-Pei Cheng."Using Back Propagation Neural Network for Channel Estimation and Compensation in OFDM Systems."Complex, Intelligent, and Software Intensive Systems (CISIS), 2013 Seventh International Conference on.IEEE, 2013.

[5] Aldar, Dilip S., and B. B. Godbole. "Performance analysis of neural network based time domain adaptive equalization for OFDM." International Journal of Recent Trends in Engineering 2.4 (2009): 143-147.

[6] B. Sklar, *Rayleigh Fading Channels in Mobile Digital Communication Systems- Part I: Characterization*, IEEECommun. Mag., July 1997.

[7] J.G. Proakis, *Digital Communications*, McGraw-Hill, 1987.

[8] Harivikram, T. S., et al. "Adaptive Modulation and Coding Rate for OFDM Systems."

[9] Raul Rojas "*Neural Networks*" Springer-verlag, berlin, 1996.

[10] Parminder Kaur, Department of Electronics Technology, Guru Nanak Dev University, Amritsar/M.Tech. Student/parminder2090@gmail.com.

[11] Kuldeep Singh, Depart of Electronics and Communication Engineering, Guru Nanak Dev University Regional Campus, Sultanpur Lodhi, Kapurthala/Assistant Professor/ kuldeepsinghbrar87@gmail.com.

[12] Hardeep Kaur, Department of Electronics Technology, Guru Nanak Dev University, Amritsar/Assistant Professor/hardeepgndu@gmail.com

Back Propagation Neural Network (BPNN) Based Routing Optimization for Wireless Sensor Network

Navpreet Kaur
Deptt. of Electronics Technology
Guru Nanak Dev University
Amritsar, India
navpreetneha@gmail.com

Kuldeep Singh
Dept. of ECE
Sultanpur Lodhi
Regional Campus GNDU
Amritsar, India
kuldeepsinghbrar87@gmail.com

Hardeep Kaur
Dept. of Electronics Technology
Guru Nanak Dev University
Amritsar, India
hardeepgndu@gmail.com

***Abstract*—Wireless Sensor Network (WSN) is defined as the network which comprises of number of sensors nodes and these sensor nodes are ready to communicate with one to another node through wireless medium. WSN has many wide applications due to further improvement in technology in various fields. Power consumption, routing and quality of service are the main concerning issues of WSN applications. We have been investigating the problem of optimization of routing of WSN based on Back propagation neural networks BPNN to increase the lifetime of wireless sensor networks. This approach of routing optimization of WSN is based on different parameters like transmission energy, energy consumption, data transmission rate, distance to gateway, queue size and weight. BPNN is very simple, adaptable and efficient method to enhance the lifetime of networks.**

Keywords—Wireless sensor networks, BPNN, link cost, accuracy and mean square error (MSE)

I. INTRODUCTION

In this new World of rapid technology Wireless Sensor Networks (WSNs) can be recognized as the most important technology among all others. Wireless sensor networks have different concerning issues like storage capacity and power management give a lot of contribution like in VLSI. These networks have many good features so they have military and civil applications like target tracking, space investigation, environmental control, environmental monitoring and patient care [1], [2],[3], and [4].

Wireless sensor network made up of a large number of micro sensors which are small in size like in cubic centimeter and these sensors are spread out in specific area to carry out specific application. Micro sensor has many important features such as capability of sensing data from the environment, ability to solve complex computations and transmission of data over wireless medium from a node to gateway or vice versa.

In wireless sensor networks, routing protocols can be either proactive or reactive which is related to determination of route to destination by source. Before routes are required, routes are work out in proactive routing. On the other way, in reactive routing the routes are work out only when it is required.

Depending upon the application of WSN, the design component of the routing protocol varies because of the application's travel demand and pattern may very much. Power consumption, mobility, scalability and QoS are the other most important issues in designing routing protocols in WSN. Routing is the main concerning issue in WSN. We are using the back propagation neural network system that will optimize the routing path. This approach is depending on the different parameters: transmission energy, energy consumption, data transmission rate, and distance from gateway, queue size and weight [5].

II. WSN SYSTEM ARCHITECTURE

Wireless sensor network comprises a number of sensor nodes to generate high-quality information related to the environment. The different functions like sensing, processing, transmission, location finding, and power consumption are existing in each of the nodes. These main features are building discrete, local measurement, create a wireless network by communicating over a wireless medium and accumulate information and route information back to the user that is Base Station through sink. The sink (Base Station) is communicated with the user through internet or satellite communication. This medium is positioned near the sensor field or whether it is nearby a well equipped node of the sensor network [6].

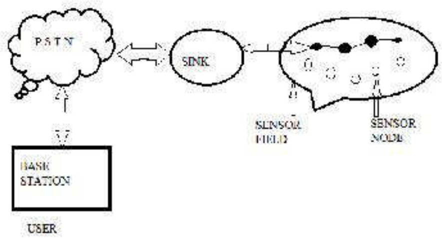


Fig.1.Architecture of a typical Wireless Sensor

III. WSN CONDITIONS AND CONFRONTS:

WSN should follow these requirements in developing various fields: scalability, reliability, responsiveness, mobility, and power efficiency. These can be described as:

A.Reliability- The capability of the network for reliable data transmission for continuous change of network system.

B.Scalability- It is the capability of the network to produce without extreme overhead.

C.Responsiveness - The capability of the network to adapt itself quickly as soon as possible to changes in topology.

D.Mobility- It is the capability of the network to hold mobility of nodes and unreliable data paths.The important characteristics of WSNs are:

- Reduction of power consumption
- Ability to manage with node breakdown
- Mobility feature of nodes
- Communication breakdown
- Heterogeneity of nodes
- Large scale use
- Withstand in critical environmental conditions and Easiness of use [7]

IV. BACKPROPAGATION NEURAL NETWORKS

Neural network is also known as an artificial neural. Network (ANN) which is a mathematical model or computational model. This model is derived by the configurationally and functional facts of biological neural networks. A neural network made up of a group of artificial neurons which is connected to each other and it is used to process information to have better computation. Modern neural networks are used to representation complex relationships between inputs and outputs which are further used to find patterns of information [8].

In feed-forward neural networks, the neurons of the first layer take output to the neurons of the second layer in a single direction which implies that the neurons are not acknowledged from the reverse direction. Multilayer Perceptron Neural Networks or Multilayer Feed-forward Neural Network is a method of feed-forward neural network system. This network comprises of three layers: input layer, output layer and intermediate layer. Different weights are provided to the sums to each other layer while the output from the first hidden layer is given to the input of the next hidden layer. Here we have more than one hidden layers present. A general structure of MLPNN have three layers is shown in Fig.

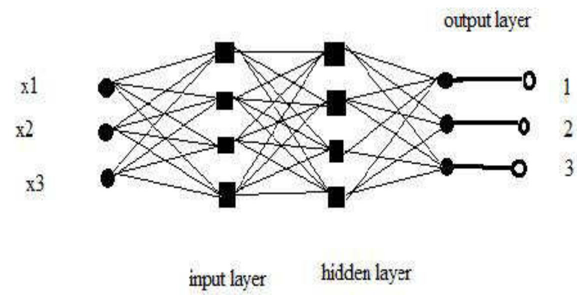


Fig.2.The multilayer perceptron neural network

The function of the neurons in the input layer is to a lot the input signal xi to neurons in the hidden layer. Every neuron j in the hidden layer calculates its input signals xi after weighting these values. The strengths of the particular links wji from the input layer and calculate its output yj as a function f of the calculation i.e. represented by: $Yj = f (WjiXi)$.

Where f function is a threshold function like a sigmoid or a hyperbolic tangent function. The output of neurons in the output layer is calculated in the same way. Learning algorithm is used to maintain the strengths of the links that is required to permit a network to attain a desired overall performance.

The inspiration for the development of neural network technology is to build up an artificial system that will be used to perform brainy functions that was actually similar to functions of the human brain. Neural networks are similar to the human brain in the following two traditions:

1. A neural network obtains knowledge through learning.

2. The knowledge of neural network is accumulated within in between neuron connection strengths which are known as synaptic weights.

The Back-Propagation Network Model:

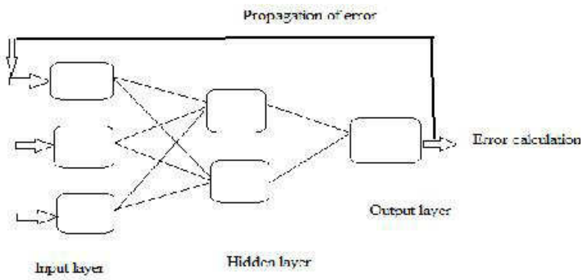


Fig.3.Back propagation neural network

In this paper we are using the most important Back-Propagation Network (BPN) algorithm to investigate the routing optimization of WSN. The BPN algorithm is a classic Supervised Learning Network which is meant by learning is done under observation of a teacher. These observations give the error between the desired output and actual output and also known about the synaptic weights of network neurons [9], [10].

The Back-Propagation Network model has three system layers and expressed as follows.

1. Input Layer is consisted of the inputs of the BPN and it will represent the initial standards of result.

2. Hidden Layer is consisted of the neurons which are required for regulating the synaptic weights of neuron connections and calculating the appropriate synaptic weights. For better accuracy effects, the hidden layer is created of several sub-layers to study the various observations between inputs and outputs.

3. Output Layer is consisted of the outputs of the BPN and will characterize the final results at this preparing process.

Table I.determination of link cost using BPNN

Sr. no	Number of neurons	MSE(10^-5)	Training Time(min)	Accuracy(%age)
1	5	-2.5	0.3	29
2	50	-19.1	0.54	30
3	100	-39.95	1.23	31
4	210	-70.5	2.15	31
5	270	-76	2.38	32
6	390	-96.5	3.36	36
7	450	-153.5	4.05	40
8	510	-160	4.37	41
9	550	-182	5.34	44
10	600	-189.5	6.07	45
11	700	-193.5	6.51	45
12	800	-194.5	7.33	46
13	900	-212.5	8.31	49
14	1000	-229.5	9.26	52

The control process of the BPN algorithm has the following processes:

Two stages of calculation:

• Forward Pass**:** this stage includes the processing of the neural network and calculates the error for each neuron of the output layer.

• Backward pass: This stage include that the process will be started at the output layer and get ahead of the errors backwards through the network, layer by layer. And then it will compute the local gradient of each neuron [11].

V. PROPOSED APPROACH

In this paper, we have created a manual data set of 2100 samples by using i5 core processor and Matlab R2009b based on different parameters used for investigating the routing optimization problem wireless sensor networks. These parameters are transmission energy, energy consumption, data transmission rate, distance to gateway, queue size and weight and the corresponding output parameter link cost. This manual data set has 2100 total samples corresponding to all above parameters, out of which 2001 are taken as training data and 99 samples are taken as testing data.

We are utilizing a back propagation neural network method for routing optimization in wireless sensor networks by using manual data set. With the help of Back propagation neural network method, we are able to determine the accuracy of the networks of different number of neurons at different values of time having different performance values in terms of its mean square error.

Although, BPNN is not best method to give best result of the network. We have a following table that shows different accuracy values.

Performance: performance of the network system is evaluated in terms of mean square error. With the number of neurons increases, accuracy of the system and also increases time to execute the process. The value of the mean square error decreases as increasing the number of neurons.

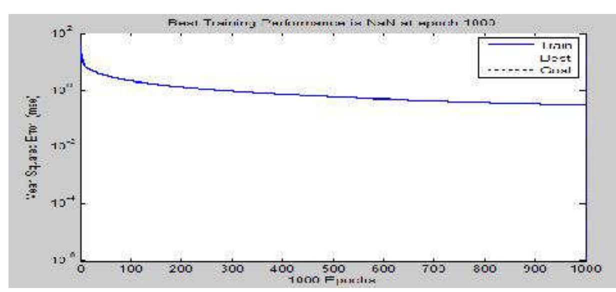


Fig. 4. performance in terms of mean square error of the network using BPNN

The following graph is shown a relationship between number of neurons and mean square error of the network system. Its result shows that with the increase in number of neurons in network system by using BPNN, the mean square error decreases. The value of MSE is very less approximately(-229.5) at 1000 neurons.For better performance of the system, this parameter should be decreases.

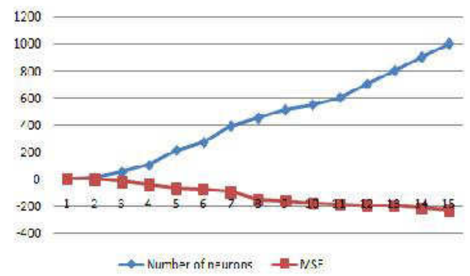


Fig.5.Graph between number of neurons and mean square error

Now we have an another graph which gives a relationship between number of neurons and training time taken by neurons.This graph shows that as increase in number of neurons,training time of neurons for processing also increases.These both terms are important to have better performance of network system.

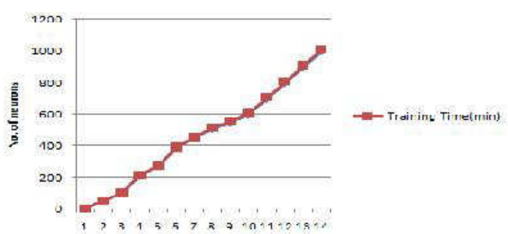


Fig. 6.Graph between numbers of neurons and training time.

Now, we have one more graph which shows a relationship between numbers of neurons and accuracy of the network system. With increase in number of neurons, the accuracy of network also increases. This is also a very important parameter to give better performance.

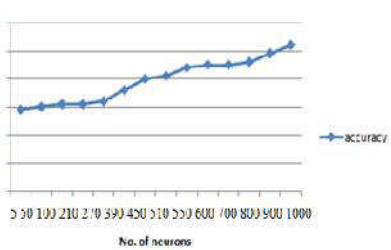


Fig. 7. Graph between accuracy of network and number of neurons.

The optimum value of accuracy of the system is 52% of 1000 number of neurons and its corresponding training time is 9.26 min.

VI. CONCLUSION

In this paper, performance and accuracy of the wireless sensor networks have been evaluated with the help of back propagation neural network BPNN. BPNN method gives some accurate values of performance which helps to determine best link cost for routing optimization of WSN. This ultimately enhances the lifetime of sensor network and reduces battery consumption. Indeed BPNN network and reduces battery consumption. Indeed BPNN method is very simple, best adaptable and efficient. But this method is not giving so accurate results of performance of system. Further improvement for the performance of WSN, Radial basis function RBF method is used.

REFERENCES

[1] Tarique Haider and Mariam Yusuf, a Fuzzy Approach to Energy Optimized Routing for Wireless Sensor Networks, the International Arab Journal of Information Technology, 2009

[2] Yee Ming Chen, Yi jenpeng, Energy efficient fuzzy routing protocol in wireless body area networks, International Journal of Engineering and Applied Sciences, 2012.

[3] Akyildiz F, A Survey on Sensor Networks, Computer Journal of IEEE Communications Magazine, 2002.

[4] Jain N, Madathil D., and Agrawal D, Energy Aware Multi Path Routing for Uniform Resource Utilization in Sensor Networks, International Workshop on Information Processing in Sensor Networks (IPSN'03),California,2003.

[5] Mohammad Abdul Azim and Abbas Jamalipour, Optimized Forwarding for Wireless Sensor Networks by Fuzzy Inference System, The University of Sydney, NSW 2006, Australia,2006.

[6] Anjali, Navpreet Kaur, a Review: Optimization of Energy in Wireless Sensor Networks, International Journal of Engineering Trends and Technology, 2013.

[7] Manoj Ahlawat, Wireless Sensor Network, A Theoretical Review International Journal of Wired and Wireless Communications, 2013.

[8] Er. Lavina, Er Pankaj Dev Chadha, [illegible] Prediction Using Back Propagation Neural, International Journal of Advanced Research in Computer Science and Software Engineering, Jul[illegible] 2013.

[9] [illegible], [illegible], [illegible], 2011 [illegible].

[10] Swati Singhal, Santosh Kr.Bharti, Localization in Wireless Sensor Network: An Optimistic Approach, (IJCSIT) International Journal of Computer Science and Information Technologies, 2013.

[11] Simon haykin, Neural Networks, An imprint of Pearson Publication.

Handover Based on Backpropagation Neural Networks

Payal Mahajan
Dept. of Electronics Technology
Guru Nanak Dev University
payalmahajan27@gmail.com

Kuldeep Singh
Dept. of ECE
Guru Nanak Dev University Regional Campus, Sultanpur Lodhi, Kapurthala
kuldeepsinghbrar87@gmail.com

Hardeep Kaur
Dept. Electronics Technology
Guru Nanak Dev University, Amritsar
hardeepgndu@gmail.com

Abstract- **In mobile cellular systems, the handover is a very important process, which refers to a process that allocates an ongoing call from one Base station to another. The performance of the handover technique is very important to maintain the QoS. Handover algorithms, based on Neural Networks, Fuzzy Logic etc can be used for the same purpose system to keep QoS as high as possible. In this paper, it is proposed that back propagation networks can be used for taking handover decisions in wireless communication networks.**

Keywords:UMTS,BS,WiMax.

I. INTRODUCTION

When a mobile user moves from coverage area of one base station or cell to another cell within a call's period, the call should be shifted to the new cell's BS and the process of shifting an ongoing call from one BS to another is called handover[1][2]. Otherwise, the call will be dropped because the link with the present BS becomes too feeble as the mobile moves away.

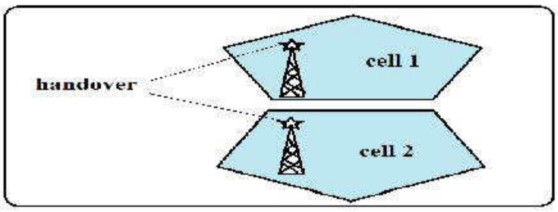


Fig. 1.Handover taking place in two cells

Handover among different wireless technologies is a very vital and burning topic today. Every user needs connectivity to the Internet everywhere and every time through any technology available at the time [8]. The connection of all these technologies is a problematic matter because every network has its characteristics. Handover in the previous generation systems was not tough to attain proficiently as the cell size was taken large enough, but in recent cellular systems, the cell size is kept minor to house a large number of users. In the case of the smaller cell size with amplified probability of the MS intersecting a cell border, the handover decision becomes more challenging. This problem becomes more complicated when there is an intersection of the signals from different BS in the surrounding area of the cell boundary. Therefore Soft Computing approaches based on Fuzzy Logic (FL), Artificial Neural Networks (ANN) Genetic Algorithm (GA), can prove to be efficient for next generation wireless networks[3][4][6]. It is necessary to ensure that handover should be performed constantly and without disturbance to any calls. Weakening which, can lead to dropped calls and customer disappointment. The performance of the handover method is very important to preserve the desired Quality of Service (QoS)[6][7].Thus it is necessary to recover the handover matter in cellular mobile systems. Traditional handover process, called horizontal handoff, provides an unbroken service when a customer passages between two next to cells [5]. A vertical handover happens in the diverse wireless network when a mobile user changes its connection between dissimilar networks i.e. the mobile users switched in different networks which have different technology (WiMax to UMTS). So in heterogeneous network vertical handover decision is mainly used for seamless service with the best network which provides the best Quality of service [5].

II. INTRODUCTION TO BACK PROPAGATION NEURAL NETWORK

In 1969, a method for learning in multi-layer network, Back propagation, was created by Bryson and Ho. It is best-known illustration of a training algorithm. Back propagation, an acronym for "backward propagation of errors", is a process of training ANN [9]. From a preferred output, the network acquires from inputs, similar to the way a child learns to identify a cow from examples of cows. It is a supervised learning method. It needs a dataset of the preferred output for a number of inputs, making the training set. It is most beneficial for feed-forward networks. Back propagation have need of the activation function used by the AN be differentiable. The objective of a supervised learning method is to discover a function that does best mapping of a set of inputs to its exact output. The back propagation algorithm needs the minimum error function in weight space using the method of gradient descent. The grouping of weights which decreases the error function is considered to be an answer of the learning problem. Since this technique needs calculation of the gradient of error function at every iteration step, we must warranty the steadiness and differentiability of the error function. The most popular activation functions for back propagation networks is the sigmoid

$$s_c(x) = \frac{1}{1 + e^{-cx}}. \quad (1)$$

The constant c can be carefully chosen randomly. The figure of the sigmoid changes according to the value of c, as can be seen in Figure

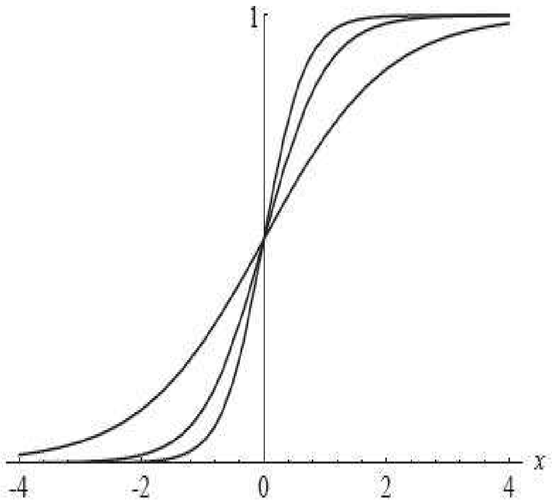


Fig. 2. Shows sigmoid function

A. Architecture of Backpropagation Network:

Multi-layer, FF networks have the following features:-They should have in any case one hidden layer. It must be non-linear units. All layers must be connected between units in two sequential layers. For a net with single hidden layer, each hidden unit accepts input from all input units and directs output to all output units. Number of output units doesn't require to be equal to number of input units. Number of hidden units per layer can be more or less than input or output units [10]

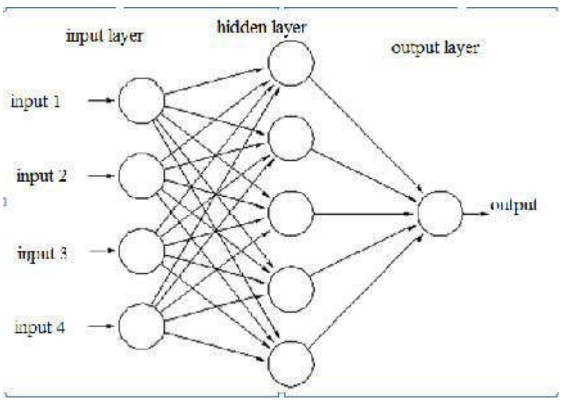


Fig. 3. Architecture of back propagation neural network

B. Learning in back propagation networks [12]:

1) Feed Forward Stage

Initially, weights take small and random values for each training pair (input/output).Each input unit transmits its value to all hidden units .Each hidden unit amounts its input signals & put on activation function to compute its output signal. Each hidden unit transmits its signal to the output units. Each output unit sums its input signals & put on its activation function to calculate its output signal

2) Back Propagation Stage

Each output calculates its error term, its own weight correction term and its correction term & sends it to layer below. Each hidden unit sums its delta inputs from above & multiplies by the derivative of its activation function; it also calculates its own weight correction term and its bias correction term

3) Adjusting the Weights

Each output unit updates its weights and bias and each hidden unit updates its weights and bias term. Each training cycle is called an epoch. The weights are reorganised in each cycle.

III. PROPOSED APPROACH

In this paper, a method to take handover decision is proposed. In this paper, a manual data set of 2500 inputs and outputs is created among which 2400 are used for training and testing of remaining 100 neurons is done. The input parameters used for creating the data set are: RSS, bandwidth, user preference, power consumption, velocity and network coverage This data set is then used in the MATLAB R2009b to develop back propagation neural network. Different number of neurons are taken and the performance of BPNN at those particular number of neurons is noted. The following table shows the output results which we get after developing the BPNN. Here the performance shows the value of mean square error.

Table I. Evaluated parameters using back propagation neural network

Number of neurons	time	Performance(*10^-5)	Accuracy (%)
20	0:26	0.699	75
25	0:27	1.51	77
40	0:33	2.11	78
100	0:50	7.07	79
150	0:59	7.27	80
175	1:07	18.3	82
200	1:18	20.6	85
450	2:15	29.8	86
600	3:05	36.5	87
750	3:12	47.2	88
1000	4:48	60.4	91

The following graph shows the mean square error of BPNN decreases with increase in time and number of neurons and one thousand epochs are taken

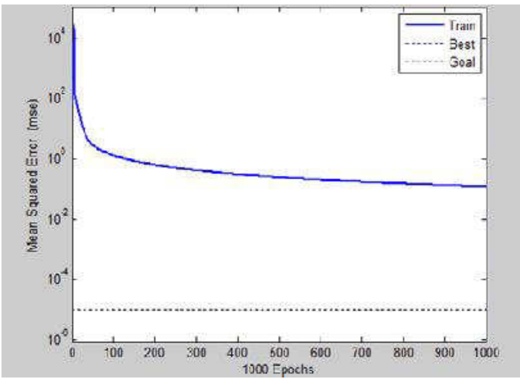


Fig. 4. Graph for performance

IV. RESULTS

The results in this paper are obtained by using the BPNN training tool and the desired graphs are obtained.

The following graph shows that as the number of neurons increase time taken increases.

Fig. 5.Graph between number of neurons and time

The following graph shows that as the number of neurons increases accuracy also increases and reaches 91% for 1000 neurons

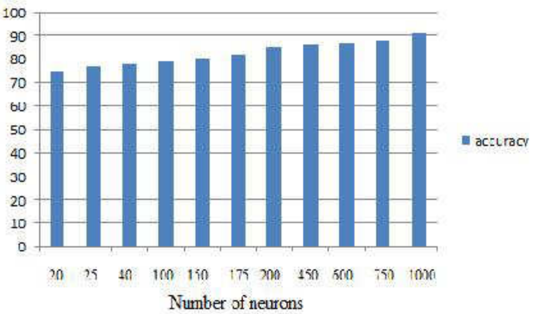


Fig. 6. Graph between number of neurons and accuracy

The following graph shows that the mean square error decrease as the number of neurons increases which is desirable for a system to work efficiently

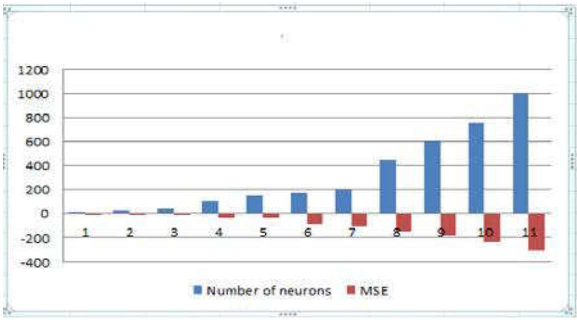


Fig. 7.Graph between number of neurons and mean square error

V. CONCLUSION

In this paper, we have first designed a back propagation neural network. The result shows that the accuracy of the system increases as the number of the neurons and time increases whereas the mean square error decreases. This shows that the system is simple, adaptable and accurate but when we consider the same for taking the handover decision, the fact cannot be neglected that the training time taken by the system is much more which is not feasible for practical handover decisions. Furthermore, it is proposed that the RBF system can be used to increase the practical utility of the neural network for handover decisions.

REFERENCES

[1] Rappaport, Theodore S. Wireless communications: principles and practice. Vol. 2. New Jersey: Prentice Hall PTR, 1996.

[2] Kassar, Meriem, Brigitte Kervella, and Guy Pujolle. "An overview of vertical handover decision strategies in heterogeneous wireless networks." Computer Communications 31.10 (2008): 2607-2620.

[3] Sangwan, Suman, Parvinder Singh, and R. B. Patel. "Adaptive Vertical Handoff in Heterogeneous Wireless Networks." INTERNATIONAL JOURNAL OF DATA & NETWORK SECURITY 1.3 (2012): 98-100.

[4] Sharma, Manoj, and R. K. Khola. "FUZZY LOGIC BASED HANDOVER DECISION SYSTEM." International Journal of Ad Hoc, Sensor & Ubiquitous Computing 3.4 (2012).

[5] Kausar, M., and DHANARAJ CHEELU. "Context Aware Fuzzy Rule Based Vertical Handoff Decision Strategies for Heterogeneous Wireless Networks." International Journal of Engineering and Science 3.7 (2013): 6-12.

[6] Sati, Dayal C., Pardeep Kumar, and Yogesh Misra. "FPGA implementation of a fuzzy logic based handoff controller for microcellular mobile networks." International Journal of Applied Engineering Research (0976-4259) 2.1 (2011).

[7] Thumthawatworn, Thanachai, Anjum Pervez, and Pratit Santiprabhob. "Adaptive Traffic Dependent Fuzzy-based Vertical Handover for Wireless Mobile Networks." ICN 2013, the Twelfth International Conference on Networks. 2013.

[8] Kumar, Rajiv, and Rajesh Khanna. "Quality of Service Approach in UMTS-WLAN Handover." Green Computing and Communications (GreenCom), 2012 IEEE International Conference on. IEEE, 2012.

[9] Hagan, Martin T., and Mohammad B. Menhaj. "Training feedforward networks with the Marquardt algorithm." Neural Networks, IEEE Transactions on 5.6 (1994): 989-993.

[10] Leung, Henry, and Simon Haykin. "The complex backpropagation algorithm." Signal Processing, IEEE Transactions on 39.9 (1991): 2101-2104.

[11] Satish Kumar, neural networks, MC Graw Hills companies.

[12] Narendra, K.G.; Sood, V.K.; Khorasani, K.; Patel, R.V. "Investigation into an artificial neural network based on-line current controller for an HVDC transmission link", Power Systems, IEEE Transactions on, On page(s): 1425 - 1431 Volume: 12, Issue: 4, Nov 1997

Decision Support System for Candidate's selection in Placement Drives using Fuzzy Logic

Karanpreet Kaur
Department of Electronics Technology
Guru Nanak Dev University
Amritsar, Punjab, India
KaranpreetKaur93@gmail.com

Gurpreet Kaur
Department of Electronics Technology
Guru Nanak Dev University
Amritsar, Punjab, India
Gurpreetghotra99@yahoo.in

Kuldeep Singh
Department of ECE
Guru Nanak Dev University Regional Campus, Sultanpur Lodhi,
Kapurthala, Punjab, India
Kuldeepsinghbrar87@gmail.com

***Abstract*—The paper deals with a decision support system meant to simplify the process of recruitment of candidates by the companies at the time of placements. The software uses fuzzy logic and is an attempt to help the companies in the personnel selection process i.e to select the best fit employee for a job vacancy. Based on five inputs such as Communication Skills, Technical Knowledge, I Q, Activeness and Academic Record each of which is allocated three membership functions i.e poor, good and excellent in the input space and the output was the selection/rejection of the candidate. The results were compared with the previous cumbersome process and found to be matching. This shows that the analysis system that we developed is effective in evaluating the candidates and can help the decision maker to make the best choice.**

Keywords—Decision support system, Fuzzy Logic, placement,recruitment.

I. INTRODUCTION

Campus Recruitment Drive means the colleges invite the IT or Non-IT companies to visit their colleges for the purposes of recruitment. The employer company would visit colleges to select the potential employees. The various companies that come for recruitment of candidates are Wipro, TCS, Cap Gemini, LG, Samsung etc. Each one of these companies would have its own selection criteria like the students should have consistently achieved 65% or above marks and they should have excellent communication skills (both oral and written). They will conduct first, written tests, then, group discussions and finally the one to one interviews. This is the process of recruitment and is quite cumbersome.[1], [2].

At the time of placements, students with tense faces, worried minds are seen everywhere on the campus eagerly waiting for being recruited. .

According to Pigorsand Myres, placement may be defined as "the determination of the job to which an accepted candidate is to be assigned, and his assignment to that job. It is a matching of what the supervisor has reason to think he can do with the job demands (job requirements); it is matching of what he imposes (in strain, working conditions) and what he offers in the form of pay roll, companionship with other promotional possibilities etc.[1] The Package which the company is expecting from the candidates consists of effective communication skills and strong willpower and determination to work etc.

Fuzzy logic is a problem solving control system methodology that tends itself to the implementation in systems ranging from simple ,small, embedded microcontrollers to large networked, multi-channel PC or workstation based data acquisition and control systems. It can be implemented in hardware, software or a combination of both. Fuzzy logic provides a simple way to arrive at a definite conclusion based on vague, ambiguous, imprecise, noisy or missing input information. Fuzzy logic approach to control problem mimics how a person would make decisions and how much faster [2].

This decision support system takes five inputs namely: Communication Skills, Technical Knowledge, IQ, Activeness and Academic Record and gives an output which determines the selection of a candidate. Using this repeatedly, a system is designed to select a candidate for placement. We choose this because there are several fuzzy logic inputs which the company should consider to recruit a candidate. Each input is allocated three membership functions which are poor, good and excellent. A membership function (MF) is a curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1.Thus, the various inputs are as follows [4]:

- Communication skills- It is the ability to affectively transfer the information from one place to another whether vocal or written. The ability to communicate information accurately, clearly and as intended, is a vital life skill and something that should not be overlooked. Success in communication is indicated by how closely the other person interprets what you say or do with what you have in mind.

- Technical Knowledge- Technical knowledge is the ability to grasp and fully understand key skills and concepts required in a certain field of work. In order for one to remain relevant in any field, he/she must possess, develop and uphold knowledge of key trends in technology so as to be able to experiment and innovate.

- I Q- Problems do arise even when the best of the work strategies are used .A person with good IQ can sense the problem, pinpoint and nail down the factors causing that problem very quickly and calmly before they can do any harm to the establishment. Thus, for a candidate with high level of IQ, tackling every critical situation in a company is a cakewalk.

- Activeness-Activeness is the state of being physically and mentally active and alert to the surrounding stimuli /happenings around. Moving or acting rapidly and energetically, showing quick responses and reflexes is a quality that every employer values and desires in an employee. We can say that it is one of the most impressive and essential traits required in any candidate to be placed.

- Academic Record-The official record of the achievements of the candidate is called an academic record. This record details the entire academic history. The Academic record is basically the transcript (no, it doesn't include your disciplinary record), so a "good academic record generally means that you maintained decent grades, no academic probations or expulsions.

Thus when all the above inputs are applied we get the output namely ‘selection’. This is the optimum decision making method that helps the companies to recruit the candidates.

II. LITERATURE REVIEW

The role of fuzzy logic in Personnel selection process effective HRM practices have been proved to relate to company’s performance by contributing to the organization’s success in terms of customer’s satisfaction, profitability, innovation and so on. The hiring procedure is the first contact of future employee with an organization .Recent Research has shown that the employee’s commitment to organization is dependent on the employee’s treatment in the hiring process [2].

Personal selection followed by human resources planning and recruitment, is the activity in which organization use various methods to judge whether the candidates are suited for positions needed to the occupied. Reviewing the contemporary recruitment literature, it seems that this challenging task of trying to distinguish “good” workers from “poor” workers with aid of tests and other devices coincides with early formulations of industrial psychology [3]. That is the principle of personnel selection have changed very little since Freyd [4].

A person without right skills will not perform the job effectively, and thus will affect the performance of the organization. The remedy of this situation is either repeating the selection process and hiring a new employee or training the hired employee .Both the cases strongly influence the effectiveness of HRM practices and require spending a great amount of resources especially time, effort and money.

A. The Personnel Selection Process

Personnel Selection process is hiring the right person to do the right job in order to increase the performance of company to meet the demands of customers. There are standard ways many organization’s go through to select its employees. Some companies interview straight way qualified applicants and select the appropriate personnel for the job.

Other organization also go through procedures that involve the applicants to write skill based examinations. The reason for this rigorous process is to reject inappropriate applicants.

According to (management study guide, 2012) the personnel selection process takes place in the following order [5].

- Preliminary Interviews-It is used to eliminate those candidates who do not meet the minimum eligibility criteria laid down by the organization. The skills, academic and family background, competencies and interest of the candidate are examined during preliminary interview.

- Application blanks-The candidates who are successfully in preliminary interview are required to fill application blank. It contains data records of the candidates such as details about age, qualifications, reasons for leaving previous job, experience etc.

- Written tests-Various written tests conducted during selection procedures are aptitude tests, intelligence tests, reasoning tests, personality tests, etc. These tests are used to objectively assess the potential candidates.

- Employment Interviews-It is a one to one interaction between interviewer and the potential candidate. It is used to find whether the candidate is best suited for the required job or not. But such interviews consumes time and money both. There should be honest communication between candidate and interviewer.

- Medical examination-Medical tests are conducted to ensure physical fitness of the potential employee. It will decrease the chance of employee’s absenteeism.

- Appointment letter-A reference check is made about the candidate selected and finally he is appointed by giving formal appointment letter.

B. A fuzzy approach to personnel selection

Attempts to help decision makers minimize subjective values judgments and make more effective selection from the pool of applicants, under the various evaluation criteria. Exact evaluation of job applicants is a difficult process because of the complexity and multi-variances of their skills.

The basic postulate behind this work is that many real world problems have more to do with fuzziness than randomness as the major source of imprecision (Misha Lovrich, Sonja Petrovic-Lazarevic, Bruce Brown, 1999. In such situations, it is more appropriate to handle uncertainty by fuzzy set theory than probability theory (Misha Lovrich, Sonja Petrovic-Lazarevic, Bruce Brown, 1999) [6], [7]. Tools employed the spreadsheet, single-skill task and multi-skills per task. The exact evaluation of job applicants is a difficult process because of complexity and multivalences of their skills.

However fuzzy theory developed by Zadeh can be applied to improve efficiency of assessments and reduce subjective judgment.

Decision makers allocate weights to each skill with its corresponding ranked.

fuzzy attributes, with the aid of a fuzzy dictionary in a fuzzy suitability table and a generalized fuzzy suitability table is constructed for single-skill and multi-skills respectively.

III. INTRODUCTION TO FUZZY LOGIC

Dr. Lotfi Zadeh, a professor of mathematics from U.C. Berkeley, proposed the fuzzy theory 1965 [8].Fuzzy logic is a valuable tool, which can be used to solve highly complex problems where a mathematical model is too difficult. It is also used to reduce the complexity of existing solutions as well as increase the accessibility of control theory. The development of software has always been characterized by parameters that possess certain level of fuzziness.

Fuzzy logic controller is rule based controller where a set of rules represents a control decision mechanism to correct the effect of certain cause used for many systems. The configuration of fuzzy logic based system is divided into three parts they are, Fuzzification , Interface Mechanism and Defuzzification. It is implemented in three phases as shown in the following figure [1].

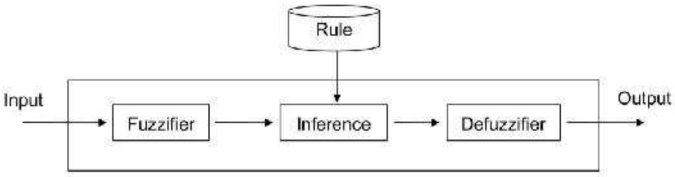


Fig. 1. Fuzzy Inference System: Different phases

Fuzzification is the process which converts classical data or crisp data into fuzzy data or Membership Functions (MFs). *Fuzzy Inference Process* combines membership functions with the control rules to derive the fuzzy output.

Defuzzification is the process which use different methods to calculate each associated output and put them into a table: the lookup table and after that, pick up the output from the lookup table based on the current input during an application.

In this research paper, Fuzzy Inference System has been implemented in MATLAB. FIS has five blocks named as FIS editor, membership function editor, rule editor, rule viewer, surface viewer. All the five blocks are shown in following figure[2].

Fig. 2. Different parts of Fuzzy Inference System

IV. SIMULATION & RESULTS

The system was tested a couple of times to ascertain the working of the system and to see if we were on right track with our membership functions by considering factors necessary for the selection of candidate in placement. A particular candidate was chosen, the values for our input and fuzzy logic system were estimated to determine the necessary factors required by the company. Some flaws were detected with values; which necessitated reconsideration & re-evaluation of rules. After the accuracy of the system was ensured we made the rules and inputted our data.

Following are the few sets of observations obtained during our experiment:-

TABLE I. Input Variables Along With Their Membership Functions

S.NO	INPUTS	MEMBERSHIP FUNCTION	RANGE
1.	**Communication Skills**	Poor, Good, Excellent	0-100
2.	**Technical Knowledge**	Poor, Good, Excellent	0-100
3.	**IQ**	Poor, Good, Excellent	0-100
4.	**Activeness**	Poor, Good, Excellent	0-100
5.	**Academic Record**	Poor, Good, Excellent	0-100

The Fuzzy inference system, Placement, has been designed by taking into consideration input variables mentioned in table I given above. The output of this system is ‘selection’ which considered as ‘Yes’ for range in between 0-5 and ‘No’ above 5. This FIS is shown in following figure.

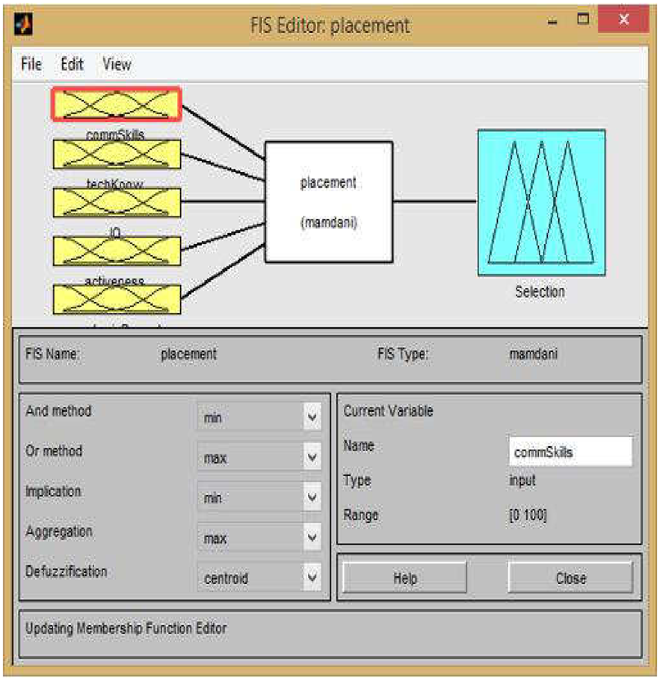


Fig. 3. Placement- Fuzzy Inference System

Input and output variables along with their membership functions are shown in figures 4 and 5. With the help of these five input variables, output 'selection' is decided by using different combinations of Fuzzy If-then rules. Fuzzy If-then rules are shown in figure 6.

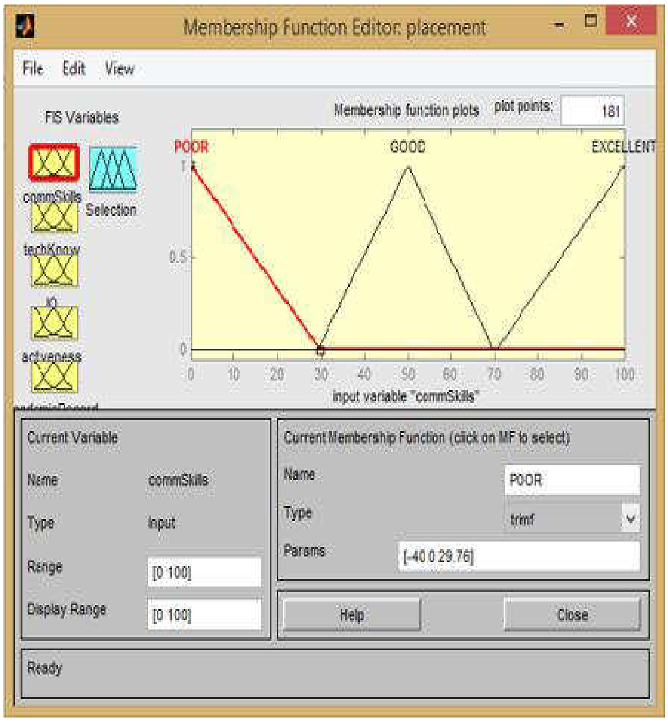


Fig. 4. Comm. Skills- Input of FIS showing its membership functions

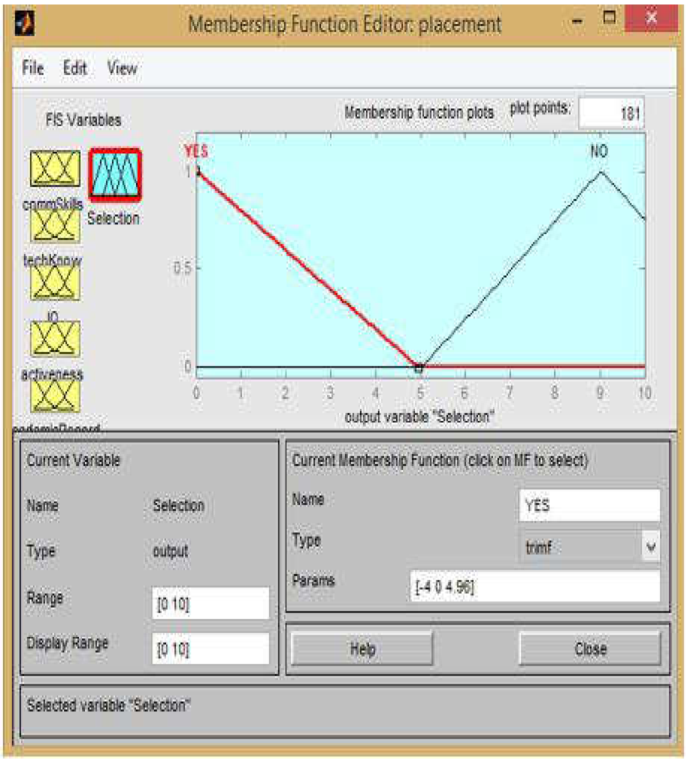


Fig. 5. Selection- output variable of FIS 'Placement'

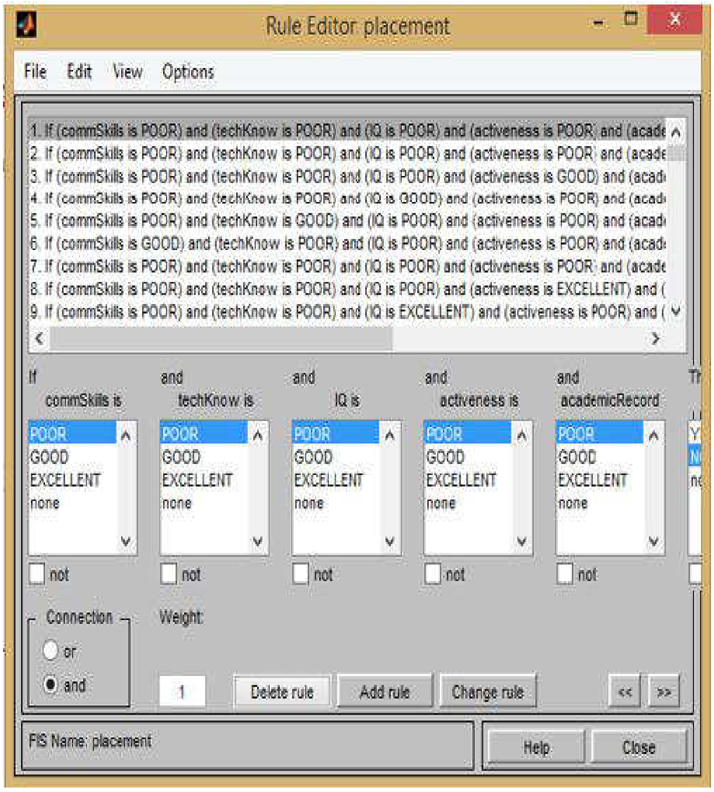


Fig. 6. Fuzzy IF-THEN rules of FIS

The following figure showns the surface viewer of this fuzzy inference system by taking into consideration of input variables comm. Skills and tech. know. This surface viewer depicts the degree of randomness in fuzzy If-then rules.

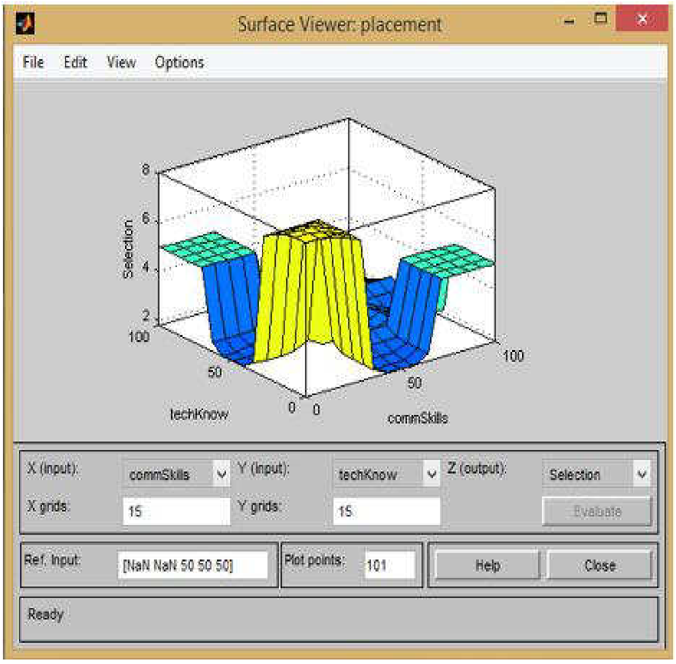


Fig. 7. Surface Viewer of FIS

After the complete design of Fuzzy Inference System 'Placment', this system has been verified for different combinations of input values and results obtained are shown in follwing table. After manual verification, it is quite clear that these results are very effective and satisfactory. So this Decision support system can be utilised to analyse the performance of candidates and to help the recuitment officials for proper selection.

TABLE II. RESULTS OF FIS FOR UNLABELLED INPUTS

S. NO	INPUTS TO 'x'	OUTPUT VALUE	SELECTION
1.	x= [60 50 40 70 10]	5.000	No
2.	x=[30 40 50 60 90]	2.4352	Yes
3.	x=[10 20 30 25 7]	7.6638	No
4.	x=[30 40 60 80 90]	2.4083	Yes
5.	x=[40 50 60 70 90]	2.1600	Yes

V. CONCLUSION

This is to conclude that the specific factors that are of concern in the analysis of results are : communication skills, technical knowledge, IQ, activeness and academic record. From this point of view, a novel framework is proposed to utilize personnel selection procedure, with the help of Fuzzy Logic. This decision support system is quite realistic and gives satisfactory results for selection of candidates in placement drives.

REFERENCES

[1] Bowen D, Gilliland, W. Folger R , "HRM and service fairness. How being fair with employee's spills over to customers", Organizational Dynamic, (winter), 1999.

[2] Boran Faith Emre, Serken Gene, Diyar Akay " Personnel Selection based on Intuitionistic fuzzy sets", Human factors and Ergonomics in Manufacturing and Servicing Industries, 494-496, 2011.

[3] Blum, M.L and J.C Naylor," Industrial psychology. Its theoretical and social foundations", Harper and Row, New York, 1968.

[4] Fred, "Measurement in personnel selection. An outline of research procedure", Journel of personnel research, 2,215-249, 1923.

[5] Management study guide, 2012 http://managementstudyguide.com/

[6] Misha Lovrich, Sonja Petrovic-Lazarevic, Bruce Brown, "A fuzzy approach to Personnel Selection" Monash University Faculty of Business and Economics, 1-12, 1999.

[7] Yaakob, S.B and S. Kawata, "Workers placement in industrial environment"; Fuzzy sets and systems, 106,289-297, 1999.

[8] Lotfi A. Zadeh , "Fuzzy Logic, Neural Networks, and Soft computing", Communications of the ACM, Vol. 37, Issue No. 3, March 1994.

Wildfire Detection System Techniques: A Survey

Balwinder Kaur
CT Institute of Engineering
Management & Technology,
Jalandhar, India

Manoj Kumar
CT Institute of Engineering
Management & Technology,
Jalandhar, India

Manju Bala
CT Institute of Engineering
Management & Technology,
Jalandhar, India

***Abstract*— Wireless Sensor Networks (WSNs) have been widely studied as one of the most important technologies for the twenty-first century which is enabled by recent advances in micro electro-mechanical systems (MEMS) and wireless communication technologies. In the past decade, a large amount of research activities have been carried out to explore and solve various design and application issues, and important advances have been made in the development and deployment of WSNs. WSNs provide bridges between the virtual world of information technology and the real physical world. As a result, WSNs also have the potential to produce new breakthrough scientific advances. Using fuzzy logic in wireless networks is a new advancement in this field. The fuzzy driven systems can be prepared so as to control the working of a particular procedure or function of a program. Fuzzy logic is implemented on the sensed data which is gathered at base station. Membership function is generalization of the indicator function in classical sets. In fuzzy logic, it represents the degree of truth as an extension of valuation. There are various membership functions such as Triangular, Trapezoidal, Bell shaped, Sigmoid, Gaussian etc. Various membership functions, rules can be modified according to parameters for more accurate results.**

Keywords*— *Fuzzy Logic, Wireless Sensor Networks, Member Functions, Inference Rules

I. INTRODUCTION

Wireless Sensor Networks consist of autonomous devices with inbuilt sensors for monitoring various environmental and physical conditions at various places and times. WSNs have unique characteristics, such as higher unreliability of sensor node, denser level of node deployment, and severe energy consumption, and storage constraints, which present many new challenges in the development and application of WSNs. It is visualized that in the near future WSNs will be widely used in various civilian and military fields, and modifies the way we live, work, and interact with the physical world [1]. Wireless sensors networks have significant advantages over traditional wired sensors networks. WSNs can not only reduce the cost and delay in deployment, but also be embedded in any environment, especially those in which traditional wired sensor networks are impossible to be embedded. WSNs were firstly motivated by military applications, which range from large - scale sensing surveillance systems for ocean surveillance to small networks of unattended ground sensors for ground target detection. However, the availability of wireless communication and low cost has promised the development of a wide range of applications in both civilian and military fields [2].

A. Wireless Sensor Network Standards

There is a need for building a large low cost market for sensor products in this field to facilitate the worldwide development and application of wireless sensor networks. For this purpose, the specification of relevant standards is important so that sensor products from different manufacturers may inter-operate. To a certain extent, success of the WSNs as a technology will largely depend on the success of these standardization efforts which are:

- The IEEE 802.15.4 Standard.
- The ZigBee Standard.
- The IEEE 1451 Standard.

B. Sensor Node Structure

Wireless sensor network refers to self configuring network of small sensor nodes connected with each other through wireless medium that have not only sensing capability but also processing and communicating. These sensor nodes can sense or detect a variety of conditions and parameters that may be temperature, smoke, light, humidity, pressure, position, speed, direction, air etc. Each Sensor node consists of four basic functional units that are Power unit, Sensing unit, processing unit and Communication unit which are shown in Figure 1.

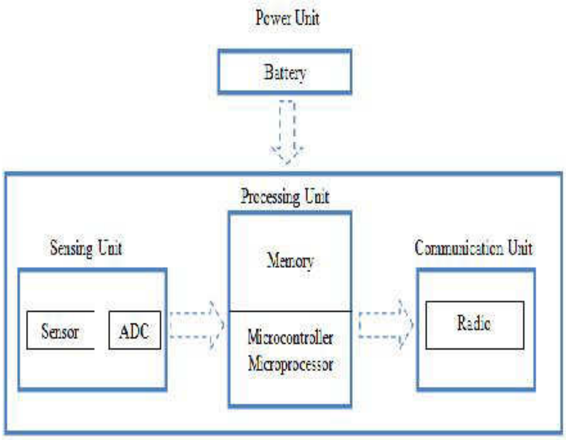


Fig. 1. Structure of Sensor Node [1]

II. REVIEW OF LITERATURE

During the last decade, a considerable number of studies have been carried out regarding the involvement of Wireless Sensor Networks in early detection problem of wildfires. Most of the research work in wireless sensor network is done towards enhancing and increasing the lifetime of sensor networks by proposing new energy efficient algorithms [4]. Some researchers proposed models for fire detection with high accuracy and low false alarm rate.

Hamdy Sloiman analyzed the potential of combining wireless sensor networks with artificial neural networks (ANNs) to build a "smart forest fire early detection sensory system" (SFFEDSS) in which temperature, smoke and light data from low cost sensor nodes spread out on the forest bed is aggregated into information [3]. Forest fire early detection system is implemented using small and cheap sensor nodes which can be left unattended. Large expensive centralized control equipment is used which adversely affects system robustness in such vulnerable environment. Sensor network represents a small scale cell that can be replicated in much larger topology to cover the entire forest. The data is collected via sensor nodes and transmitted to an already trained Artificial Neural Network. This system is not only able to detect fire but also accurately report the direction of fire progress which is deduced from the wind direction.

Al-Abbass Y. Al-Habashneh proposed three MAC protocols for forest fire detection. Two of these protocols are based on Carrier Sense Multiple Access with some modification to suit the forest fire detection application [5]. The protocols are called Persistent CSMA (P-CSMA) and Per Hope Synchronization CSMA (PHS-CSMA). The third one is a Time Division Multiple Access (TDMA) based protocol, called Sensor TDMA (S-TDMA). The simulation of three proposed protocols is done to compare their performance adaptively with the forest fire detection application. The three protocols are investigated and analyzed by simulation. Results showed that there is no superior protocol which outperforms others in terms of power consumption, delay and complexity, however trade-off does exist.

Federico Alimenti proposed a model with the development of microwave noise adding radiometer which is purposely designed for the fire detection in forest environments [6]. First, a simple system model is presented to estimate the radiometric contrast due to presence of fire at a certain distance from receiving antenna. Then the design of sensor is addressed underlying the key technologies that allow the required performance to be attained at low industrial costs. In this system the low cost building blocks are demonstrated, intended for the SAT-TV market which is well suited for the realization of 12-GHz microwave radiometers. These radiometers have enough sensitivity to detect fires with filling factor below 1 % and exploiting a continuous noise-adding calibration, good gain stability could be easily obtained. Actual fire experiments have been carried out both in the laboratory and open space environments. From the measured data, the fact emerges that the major contribution to the microwave emission of wooden fire is due to the embers. The agreement between experimental results and filling-factor simulation suggested that such a model can effectively used to predict the performance of microwave radiometer systems as forest fire sensors.

Giovanni Laneve implemented a new approach which exploits the increase in both spatial and temporal resolution with respect to previous metostat systems [7]. The proposed idea is based on the use of a change detection technique to maximize the detection capabilities of the system in spite of its limited spatial resolution. This approach consists of comparing two or more images acquired at 15 min intervals, for that any temperature change can be attributed to fast dynamical phenomena such as fires, when the natural changes are modelled and removed. An assessment of the performance of this algorithm was carried out comparing its result with the report made by Italian fire fighting agencies and with fire products based on higher resolution sensors such as the Moderate Resolution Imaging Spectroradiometer.

A.K. Singh proposed system for fire detection using wireless sensor networks and fuzzy type-2 logic [8]. Fuzzy gives the best result in such cases because there is uncertainty about how much extent of a factor like temperature, CO density, humidity and light intensity is responsible to cause a fire. This system was implemented in MATLAB in which the sensor nodes were deployed and clustering is done in such a manner that each cluster has its own cluster head. The sensor nodes can only communicate with cluster head and cluster head will further communicate with base station. The gathered data acts as input for the system on which Fuzzy logic was implemented to calculate the probability of the forest fire. Various rules and membership functions are used in this approach that can be altered according to the parameters available in particular environment. Fuzzy type-2 logic is used to make the results accurate and error free with very low false alarm rate so that there would be no uncertainty in decision making.

M.P.Sivaram Kumar, proposed a quadrant based clustering and deterministic deployment of nodes for detecting and extinguishing forest fires using Wireless Sensor and Actor Networks (WSANs) [9]. In WSANs the role of sensors is to collect information from the environment, while the actors collect and process data and perform appropriate actions. Beacon nodes and cluster head nodes are used, all beacon nodes were clustered based on quadrant and one actor. The actor moved to target area on the assumption that there are no obstacles.

Simon Y. Foo proposed a fuzzy logic approach to detect hydrocarbon fires in aircraft dry bays and engine compartments [10]. The inputs given to the fuzzy system consist of a set of statistical measures derived from the histogram and image subtraction analyzes of successive image frames. Specifically, fuzzy rules are based on the median, normalized first-order moment and standard deviation statistical measures of the histogram data and the mean statistical measure of image subtraction data of successive frames are used to compute the probability of a fire event. This approach is also tested for false alarms such as those due to flashlights and high-power halogen lights. Results show that image subtraction analysis can be used to accurately distinguish fires from false alarms.

Federico Alimenti proposed model that deals with the development of a microwave noise-adding radiometer, which is designed for the fire detection in forest environments [11]. Radiometers have enough sensitivity to detect fires with filling factor below 1%, and, exploiting a continuous noise adding adjust, good gain stability can be obtained easily. As a result, the instrument can operate with reliability at ambient temperature thus saving the electrical power needed by a thermal control unit to stabilize the receiver temperature. The components cost of the realized prototype was about $1800 which can be reduced with a proper engineering of the system. Actual fire experiments have been carried out both in the laboratory and open space environments. From the measured data, fact emerges that the major contribution to the microwave emission of a wooden fire is due to the embers. Such a model can be used to predict the performance of microwave radiometers systems as forest fire sensors.

III. PROPOSED WORK USING FUZZY LOGIC TOOLBOX IN MATLAB

We have proposed a system that can be used to monitor the forest for detection of fire with the help of wireless sensors that give best results and very low false alarm rate for detection. Wildfires are one of the main causes of environmental degradation nowadays. Current monitoring systems for wildfire lack in supporting real-time monitoring of every point of region at all times and early detection of fire threats. We can gather sensory data such as temperature and humidity, from all points of field continuously, day and night using wireless sensor networks which provide fresh and accurate data to fire fighting centre quickly. Sensor Networks face serious obstacles like limited energy resources that have to be considered carefully. Early detection of wildfire is very important in fighting against fires. Spread features of wildfire shows that, in order to put out a fire without making any permanent damage in the forest, the fire fighter centre should be aware of the threat in at most 6 min after the start of the fire (National Fire Danger Rating System, 2011) [2]. However, the accuracy of this system is highly affected by weather conditions such as clouds, light reflections and smoke from industrial or social and innocent activities. Moreover, since most of the forests are located on mountains, the sight view of devices will not be clear enough to control the whole forest. Also, considering the cost of the system and technical in capabilities of the devices whose view areas are not enough to cover a forest, it is observed that automatic video surveillance systems cannot always be applied effectively to large forest fields. The proposed system will accurately detect the probability of fire more accurately and it can be deployed in any environment. Various technologies are aims at making the fire fighters aware of forest fires as early as possible. There exists important technologies and systems that are currently used towards this goal are: systems employing infrared (IR) detectors, charge-coupled device (CCD) cameras and satellite systems and image processing and wireless sensor networks. All technologies except wireless sensor networks require satellites and satellite images. Usually, satellites capture a complete image of the earth every 1-2 days. This long scan period, however, is not acceptable for detecting wildfires quickly and accurately. As a promising alternative, wireless sensor networks (WSNs) are an emerging technology that can be used for wildfire detection and related activities. Fuzzy logic can be used for detecting fire as it gives best results and takes less memory, less processing and less energy. AK Singh and Harshit Singh proposed system in which fuzzy type-2 logic for fire detection is used. Fuzzy gives best results in such cases because there is an uncertainty about how much extent of a factor like temperature, CO density, humidity and light intensity should be involved to cause a fire. The model was implemented using four parameters having some set of inference rules for calculating probability of fire with the help of membership functions. These membership functions and rules can be altered according to input parameters for obtaining more accurate and error free results.

The main purpose for which we are using this system is to monitor the forest for detection of fire with the help of wireless sensors that give best results and very low false alarm rate for detection. For this we have to monitor or detect certain parameters on which fire depends that may be physical or environmental conditions such as temperature, smoke, light intensity, humidity etc. As fire detection is always been a crucial challenge for human being, moreover detecting fire using automated sensors definitely requires efficient and accurate ways. The objective of this study is to obtain beneficial information using sensor network and transform information to the Forest or Fire Management Systems (FMS). So we will alter membership functions and rules of the parameters for improving accuracy which will give us more accurate and error free model with low false alarming rate.

IV. METHODOLOGY

The Fuzzy Logic Toolbox in MATLAB (Matrix Laboratory) is a collection of functions built on the numeric computing environment. It provides tools to create and edit fuzzy inference systems within the framework of MATLAB, or integrate fuzzy systems into simulations with Simulink. Stand-alone C programs can be built that call on fuzzy systems you build with MATLAB. This toolbox relies heavily on graphical user interface (GUI) tools that help to accomplish the work, although work can be done entirely from the command line. The proposed system is simulated in MATLAB in which nodes are deployed in area of interest. The deployed Wireless Sensor Network is divided into clusters. Each cluster will have its own cluster head which will communicate with base station. Firstly we will deploy the nodes in the forest and divide it into clusters. Now each of these clusters will have their own cluster head to which they can communicate and transmit the sensed information. The cluster head will further communicate with the base station.

We have divided the deployed nodes into clusters and there is one cluster head for each cluster. This will save energy of rest nodes that can be utilized in other processing of system.

Type 2 Fuzzy logic is implemented on the sensed data which is gathered at base station. Membership function is generalization of the indicator function in classical sets. In fuzzy logic, it represents the degree of truth as an extension of valuation. There are various membership functions such as Triangular, Trapezoidal, Bell shaped, Sigmoid, Gaussian etc. Various membership functions, rules can be modified according to parameters for more accurate results. In Figure 3, Upper Membership Function (UMF) and Lower Membership Function (LMF) are shown. The area between UMF and LMF is known as Footprint of Uncertainty (FOU).

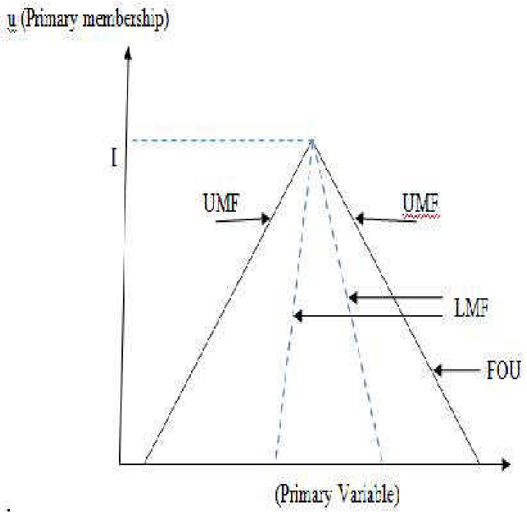


Fig. 2. Membership Functions of variable [15]

The fuzzification of input and output parameters is simulated in MATLAB. The membership functions are applied on these parameters with defined set of rules. These Inference Rules are in the form of IF THEN statements. Inference Rules are defined for these parameters according to dependency of input and output parameters as shown in Figure 3.

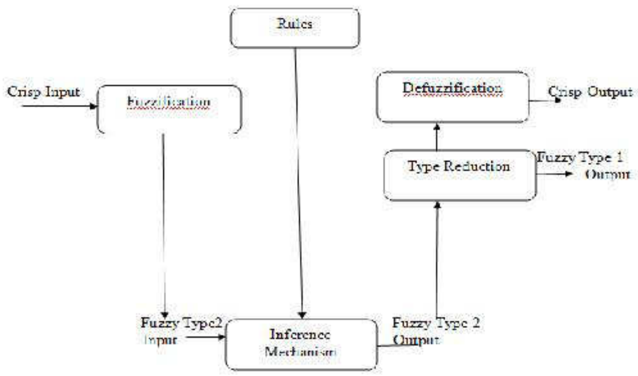


Fig. 3. Inference Engine [10]

After Defuzzification, Probability of fire can be calculated for different sets of input parameters which can be shown in Figure 4.

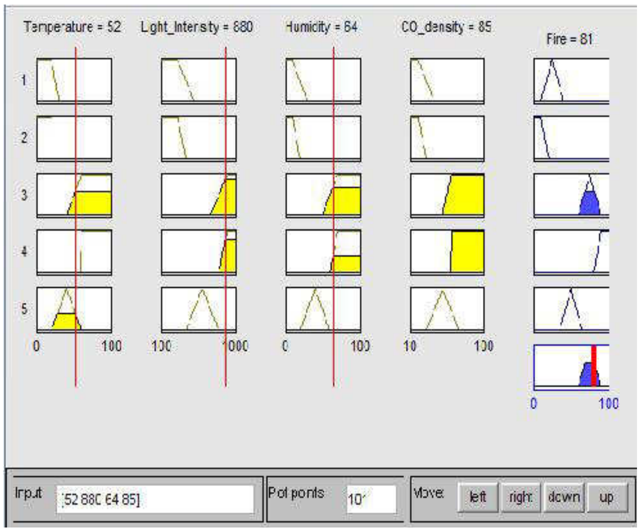


Fig. 4.Fire Probability based on input variables [8]

REFERENCES

[1] Wireless Sensor Networks-A Networking Perspective, A John Wiley & Sons, Inc., Publication, United States of America, pp. 1-27,2009.

[2] National interagency coordination center, "Wildland fire season 2004: statistics and summaries", 2004.

[3] Hamdy Soliman, Komal Sudan and Ashish Mishra, "A Smart Forest-Fire Early Detection Sensory System: Another Approach of Utilizing Wireless Sensor and Neural Networks", IEEE SENSORS Conference, pp. 1901-1904, 2010.

[4] A. S. Tanenbaum, C. Gamage and B. Crispo, "Taking Sensor Networks from the Lab to the Jungle", IEEE Computer, vol.39, Issue 8, IEEE Computer Society, 2006.

[5] Al-Abbass Y. Al-Habashneh, Mohamed H. Ahmed and Taher Husain, "Adaptive MAC Protocols for Forest Fire Detection Using Wireless Sensor Networks", IEEE, pp. 329-333, 2009.

[6] Federico Alimenti, Patrizia Basili and Klaus Solbach, "A Low-Cost Microwave Radiometer for the Detection of Fire in Forest Environments", IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 46, NO. 9, 2008.

[7] Giovanni Laneve, Marco M. Castronuovo and Enrico G. Cadau, "Continuous Monitoring of Forest Fires in the Mediterranean Area Using MSG", IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 44, NO. 10, 2006.

[8] A.K. Singh and Harshit Singh, "Forest Fire Detection through Wireless Sensor Network using Type-2 Fuzzy System", International Journal of Computer Applications (0975 – 8887) Volume 52– No.9, pp.19-23, 2012.

[9] M.P.Sivaram Kumar and S.Rajasekaran, "Detection and Extinguishing Forest Fires using Wireless Sensor and Actor Networks", International Journal of Computer Applications, Volume 24 No.1, 2011.

[10] Simon Y. Foo, "A Fuzzy Logic Approach to Fire Detection in Aircraft Dry Bays and Engine Compartments", IEEE TRANSACTION ON INDUSTRIAL ELECTRONICS, VOL. 47, NO. 5,2000.

[11] Federico Alimenti, Salvatore Leone, Gabriele Tasselli, Patrizia Basili, Luca Roselli, and Klaus Solbach, "A Low-Cost Microwave Radiometer for the Detection of Fire in Forest Environments", IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 46, NO. 9, 2008.

Future of Edge Detection Techniques: Multi-thresholding with ANN

Hartaranjit Singh
Dept. of ECE
CT Institute of Engineering Management & Technology
Jalandhar, Punjab, India
hartaransingh@yahoo.com

Gurpreet Kaur
Dept. of ECE
CT Institute of Engineering Management & Technology
Jalandhar, Punjab, India
gpreet.09@gmail.com

***Abstract*—Edge detection is one of the most commonly used preprocessing stage in image analysis. With the advancements in the field of machine and computer vision, the techniques used for detecting edges have also got improvised. The selection of an appropriate edge detection technique depends on numerous aspects like robustness, resources and adaptability. In order to get optimal response in these aspects, classical edge detection techniques have been replaced by ANN based detection in recent advancements. This paper compares the performance of various classical, fuzzy rules based and ANN based edge detection techniques. Also, a new robust approach of detecting edges by combining multi-thresholding with ANN is proposed as future scope.**

Keywords— Artificial Neural Networks (ANN), Fuzzy level, Edge Detection, Image Processing, Multi-Thresholding

I. INTRODUCTION

An Edge can be defined as a boundary which separates two regions of an image due to distinct gray levels. Several major algorithms of image processing use *Edge detection* as a prominent stage for image analysis. It is a technique which is basically used to identify the feature points in an image (sharp/abrupt changes in pixel intensity). Edge detection is extensively used as a pre-processing phase of image segmentation [18] which is being used in many vision-based applications [6] like, automatic detection of surface defects in agricultural produce, extraction/count number of objects from videos, selection of feature points in face recognition, automatic target recognition, x-ray analysis of patients etc.

The selection of any Edge Detection technique faces major challenges like false detection of edges [4], presence of noise, poor contrast [1],[20], and inappropriate thresholding. The use of image enhancement and filtering may solve the issues of noise and contrast, but the use of single level thresholding, (as shown in figure 1), of an image leads to the reduction of lot of important feature points.

A multi-level thresholding, (as shown in figure 2), minimize the loss of detail while binarizing the input image by setting more than one levels for thresholding [21]. Thus, the adoption of multilevel thresholding may help in yielding better results.

This paper discusses classical edge detection techniques in section II. Comparison of various classical techniques is described in section III. Section IV discusses the role of Fuzzy logic and ANN for edge detection. Section V describes the selection criteria of techniques and role of intelligent networks to get optimal results in edge detection and finally, the paper concludes by proposing a future advancement in the use of ANN for edge detection by replacing single-level thresholding with multi-thresholding.

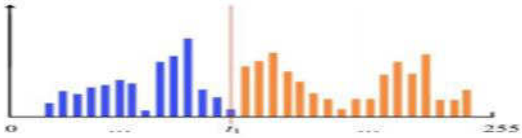


Fig.1. Single-level thresholding

Fig.2. Multi-level thresholding

II. CLASSICAL METHODS

These are the methods that are being used for the detection of edges since the major applications of image processing came into existence. For example,

A. *Sobel Edge Detector:*

It is a discrete differentiation operator, computing an approximation of the gradient work on image intensity function [16]. The operator makes use of 3×3 kernels (refer figure 3) which is convolved with the original image and calculate approximations of the derivatives.

-1	0	1
-2	0	2
-1	0	1

G_x

1	2	1
0	0	0
-1	-2	-1

G_y

Fig.3. Sobel Operator

B. *Laplacian of Gaussian Edge detector (LOG):*

Laplacian filters are derivative filters which is combination of laplacian and Gaussian filter [5], [13]. In this approach the image is smoothened by Gaussian filter before applying laplacian filter (which is shown in figure 4), as laplacian filter is alone very sensitive to noise.

0	-1	0
-1	4	-1
0	-1	0

G_x

-1	-1	-1
-1	8	-1
-1	-1	-1

G_y

Fig.4. LOG Operator

C. *Canny edge detector:*

The Canny edge detector was proposed by John F. Canny in 1986 [2]. It is the most commonly used Edge detector operator that uses a multi-stage procedure for edge detection as shown in figure 5 below.

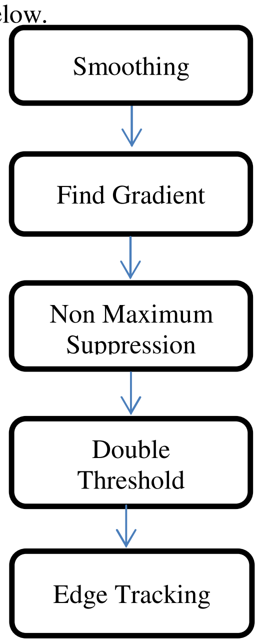


Fig.5. Canny Algorithm

III. COMPARISON OF CLASSICAL METHODS

These three classical methods have their own area of application and have different advantages and disadvantages in these application areas. Table 1 below summarizes their advantages and disadvantages [1], [13], [16], [18].

TABLE I. COMPARISON OF SOBEL, LOG AND CANNY

	Sobel	**LOG**	**Canny**
Advantage	Simple Detection of edges and their Orientation	Less False Edges	Better SNR Less noise sensitive
Dis-advantage	Noise Sensitive	Malfunctioning at corners and curves. Edge orientation not found	False zero crossing

IV. INTELLIGENT NETWORKS BASED METHODS

In recent years, new computational methodologies are developed that are able to take decisions and perform reasoning similar to human brains. These techniques uses computer to perform various intelligent tasks.

A. *Fuzzy logic:*

Use of fuzzy logic for edge detection was first proposed by Lu et al [8]. The author proposed an edge detection algorithm to recover missing edges and eliminate false edges with the help of fuzzy levels. Becerikli and Karan [10], proposed Fuzzy rules based assertion algorithm which was able to handle thick edges. Chaira and Ray [12] proposed an edge detection algorithm which was able to remove unwanted edges. This algorithm performed better than the other classical algorithms.

B. *Artificial Neural network (ANN):*

Artificial neural network is the Computational model of Biological neural network. This network performs similar to biological neurons of human brain. The basic structure of neural network consist of three layers; Input Layer, Hidden Layer and Output layer. (refer figure 6).

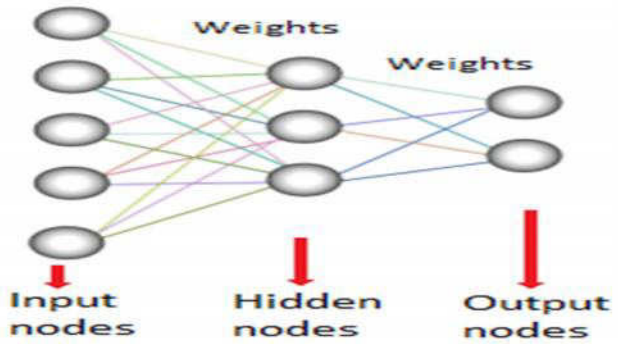


Fig.6 ANN (Artificial Neural Network)

Paik and Katsaggelos [3] proposed an ANN based rotation invariant edge detection algorithm to suppress noise in input image. Majority of the ANN based edge detection techniques use Back-propagation [7], [14], (Figure 7) as a supervised learning method. Back-propagation (BP) uses delta rule to update weight and bias .Once neural network is trained, BP network [17] is applied to the input image for detecting edges.

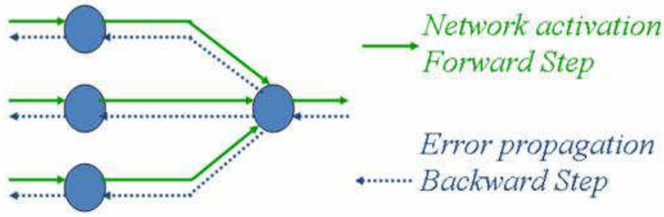


Fig.7 Back propagation

V. SELECTION CRITERIAS & GRADING

The selection of appropriate edge detection technique depends upon three major aspects; Utilization of Resources [4], Robustness and Choice of Threshold. These aspects can further be divided into other categories as shown in figure 8 below:

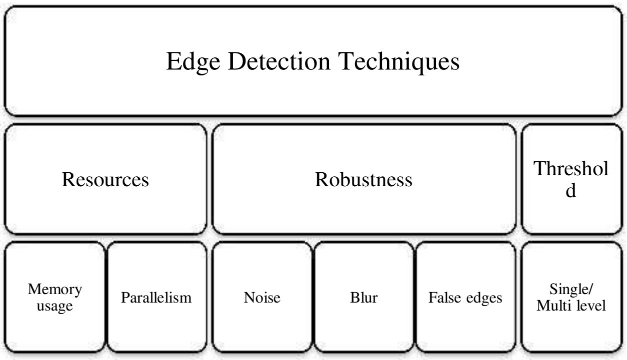


Fig.8. Selection criteria for edge detection technique

Sidra Naeem [18] proposed the pros and cons of all the previously discussed edge detection techniques on the basis of these selection criterions. Table 2 [18] below summarizes the performance comparison of classical methods with intelligent network based methods. The tabulated entries are using following abbreviations:

P: Poor

A: Average

G: Good

E: Excellent

S: Single-level

M: Both Single & Multi-level

TABLE II. INTELLIGENT NETWORKS V/S CLASSICAL METHODS

	Resource		**Robustness**						**Others**	
			Noise			**Blur**		**Prone**		
Method	Memory	Parallelism	Gaussian	Uniform	Salt & Pepper	Focal	Motion	False edges	Adaptability	Thresholding
Sobel	E	E	P	P	P	P	P	G	P	M
LOG	E	E	P	P	P	P	P	G	P	M
Canny	E	E	E	E	A	E	A	E	A	M
Fuzzy	A	G	E	E	G	E	A	E	G	M
ANN	P	A	E	E	G	G	E	E	E	S

It may be observed from Table 2 that, the classical methods (Sobel, LOG and Canny) make efficient use of available memory resources but their response to noisy & blurred images is very poor. Also, classical methods lack good adaptability even if multi-level thresholding is used.

On the other hand, Fuzzy and ANN based algorithms shows excellent performance to detect edges in case of noisy and blurred images. They also respond well in terms of adaptability issues. But, extensive use of memory resources is an issue with these networks.

VI. CONCLUSION

As discussed above, the selection of an edge detection technique depends on many criterions. It is also discussed that if one technique responds well on one aspect, then other responds good on another. In order to make a good choice, the criterions are prioritized as:

[Robustness, Adaptability, Resources]

Although both ANN and Fuzzy based networks outperformed classical methods when compared on the basis of robustness and adaptability, but adaptability of an ANN is even better than a Fuzzy level based algorithm. Hence, we may conclude that ANN is an optimum choice to develop an edge detection algorithm which performs well in all aspects.

FUTURE SCOPE

It can be seen from table 2 that, all the classical methods and fuzzy level based edge detection algorithms proposed so far by various authors are performing on both single-level and multi-level thresholding. But, ANN based edge detection algorithms proposed so far are using single-level thresholding.

It is also observed that a single-level ANN algorithm doesn't respond excellently to salt & pepper noise and focal blur. Also, the advantages of multi-level thresholding over single-level thresholding have already been discussed previously in paper. Thus, it is proposed that an ANN based edge detection algorithm with multi-level thresholding shall be designed to improve the performance of edge detection when the input image is either blurred or corrupted by salt & pepper noise.

REFERENCES

[1]. L. S. Davis, "A survey of edge detection techniques," Computer Graphics and Image Processing, vol. 4, pp. 248–270, 1975.

[2]. J.F. Canny, "A computational approach to edge detection", IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI), Vol. 8(6), pp. 769-798, 1986

[3]. J. Paik and A. Katsaggelos, "Edge detection using a neural network," in International Conference on Acoustics, Speech, and Signal Processing,ICASSP-90, vol. 4, pp. 2145–2148,1990

[4]. S. Venkatesh and L. J. Kitchen, "Edge evaluation using necessary components," CVGIP: Graphical Models and Image Processing, vol. 54,pp. 23–30, 1992

[5]. M. Basu, "A Gaussian derivative model for edge enhancement.", Patt. Recog., 27:1451-1461, 1994

[6]. Y.-C. Chung, J.-M. Wang, and S.-W. Chen, "A vision-based traffic light detection system at intersections," Journal of Taiwan Normal University:Mathematics, Science and Technology, vol. 47, no. 1, pp. 67–86, 2002

[7]. Kenji Suzuki, Isao Horiba, Noboru Sugie " Neural Edge Enhancer for Supervised Edge Enhancement from Noisy Images", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 25, No. 12, December 2003.

[8]. S. Lu, Z. Wang, and J. Shen, "Neuro-fuzzy synergism to the intelligent system for edge detection and enhancement," Pattern Recognition,vol. 36, no. 10, pp. 2395–2409, 2003.

[9]. M. Portes de Albuquerque , I.A. Esquef , A.R. Gesualdi Mello , M. Portes de Albuquerque "Image thresholding using Tsallis entropy," International Symposium on Robotics and Intelligent Sensors (ELSEVIER), pp.938 – 944, 2004

[10]. Y. Becerikli and T. M. Karan, "A new fuzzy approach for edge detectionss," in Computational Intelligence and Bio inspired Systems, vol. 3512, pp. 943–951, 2005

[11]. B. Singh and A. P. Singh, " Edge Detection in Gray Level Images Based on the Shannon Entropy", J. Comput. Sci., 4 , 3, 186 -191, 2008

[12]. T. Chaira and A.Ray,"A new measure using intuitionistic fuzzy set theory and its application to edge detection," Applied Soft Computing,vol. 8, no. 2, pp. 919–927, 2008.

[13]. Ehsan Nadernejad, Sara Sharifzadeh and Hamid hassanpour, "Edge detection techniques – evaluation and comparisons", Applied mathematical Sciences, Vol 2, pp 1507-1520, 2008

[14]. Hamed Mehrara , Mohammad Zahedinejad, "Quad-pixel edge detectionusing neuralnetwork," Y oung Resear cher s club,Science a nd Resear ch Bra nch, Islamic Azad University ,Tehran,Ira n pp 47-54, 2010

[15]. Mohamed A. El-Sayed, "A New Algorithm Based Entropic Threshold for Edge Detection in Images," IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 1, September 2011

[16]. G.T Shrivakshan and Dr C Chandrasekar, "A comparison of various edge detection techniques used in image processing", International Journal of Computer Science Issues, Vol. 9, Issue 5, No 1, pp 269-276, September 2012

[17]. Jesal Vasavada, Shamik Tiwari, "An Edge Detection Method for Grayscale Images based on BP Feedforward Neural Network," International Journal of Computer Applications (0975 – 8887) Volume 67– No.2, April 2013

[18]. Sidra Naeem,Sajid Siraj "A Framework to Select Edge Detection Method using Multi-creteria Decision Making," IEEE International Conference on Systems,Man,and Cyberntics 2013

[19]. Firas A.Jassim "Semi-Optimal Edge Detector based on Simple Standard Deviation with Adjusted Thresholding," International Journal of Computer Applications(0975-8887) Volume 68-No.2,April 2013

[20]. C. Lopez-Molina, B. D. Baets, and H. Bustince, "Quantitative errormeasures for edge detection," Pattern Recognition, vol. 46, no. 4, pp.1125–1139, 2013.

[21]. Amira S. Ashour, Mohamed A. El-Sayed ,Shimaa E. Waheed, and S. Abdel-Khalek, "New Method Based on Multi-Threshold of Edge Detection in Digital Images," (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 5, No. 2, 2014

Credit Risk Rating of Bank Customers by Fuzzy Expert System

Channpreet Singh
Department of Electronics Technology
Guru Nanak Dev University
Amritsar (Punjab) India
channpreet_695@hotmail.com

Kuldeep Singh
Department of Electronics & Communication Engineering
Guru Nanak Dev University (Regional Campus)
Sultanpur Lodhi, Kapurthala (Punjab) India.
kuldeepsinghbrar87@gmail.com

***Abstract*— Corporate insolvencies are rising and interest-rate (and thus profit) margins associated with increasing loan volumes are declining. Credit risk is defined as the potential that a borrower or counterparty will fail to meet its obligation in accordance with agreed terms. Mostly all banks today practice credit risk management. They understand the importance of credit risk management and think of it as a ladder to growth by reducing their NPA's. Moreover they are now using it as a tool to succeed over their competitors because credit risk management practices reduce risk and improve return capital. Banking institutions need instruments that can evaluate the credit standing and general financial situation of a company quickly and accurately. In this paper, we introduce and focus on a fuzzy rule-based expert system that replicates mathematically the expert knowledge of a Credit Manager. Our procedure uses Fuzzy Logic Toobox in MATLAB to rapidly evaluate company's credit standing.**

***Keywords*— Credit Risk Rating, Fuzzy Expert System (Rules) , Membership Functions, GUIDE (MATLAB), Banking Risk.**

I. INTRODUCTION

With the advancing liberalization and globalization, credit risk management is gaining a lot of importance. It is very important for banks today to understand and manage credit risk. Banks today put in a lot of efforts in managing, modeling and structuring credit risk. Assigning a credit grade involves analyzing information of the borrower. If the borrower is an individual, information of interest could be the individual's income, outstanding debt (mortgage, credit cards), household size, residential status, etc. For corporate borrowers, one may consider certain financial ratios (e.g., sales divided by total assets), industry type, etc. There are rating agencies that keep track of the credit worthiness of companies. Yet, most banks develop an internal methodology to assign credit grades for their customers. The scope of the current paper, however, is limited to the evaluation of a bank client's financial performance. In this paper, a Fuzzy Expert system has been used to evaluate the credit risk of Bank Legal customers. The aim is to analyze customers' credit risk based on the experts' analysis obtained from the financial ratios .The ratios have been extracted from financial Balance sheets [1]. So the knowledge of bank experts to link credit risk with financial situation of the customers can be taken in the form of rules for our fuzzy expert system in Fuzzy Logic Toolbox in MATLAB. In this study, a Fuzzy Expert system has been designed in which customer's financials will be considered as inputs and a predicted credit risk associated with it as the output. Since all the financials are not important in credit risk decision making, so we consider only the required financial inputs in our decision making process of Credit Risk Rating Analysis. The rest of the paper is structured as follows : In the next section, literature review is presented. Section 3 will focus on the proposed model and its practical use in financial institutions and banks. In Section 4, we present compiling up of our work into a user end interface ie. standalone executable (.exe) file using GUIDE in MATLAB. In the final section, some conclusions are drawn from the study.

II. LITERATURE REVIEW

Credit scoring is a technique that helps some organizations such as commercial banks and other financial institutions to determine whether it is viable to grant credit to the customers on the basis of predefined criteria. Internationally rating agencies like Standard & Poor's(S&P) and Moody's are well known. In India, the four authorized rating agencies are CRISIL, ICRA, CARE, and FITCH. They undertake rating exercises generally when an organization wants to issue debt instruments like commercial paper, bonds, etc. their ratings facilitate investor decisions, although normally they do not have any statutory/regulatory liability in respect of a rated instrument. This is because rating is only an opinion on the financial ability of an organization to honor payments of principal and/or interest on a debt instrument. However, since the rating agencies have the expertise in their field and are not tainted with any bias, their ratings are handy for market participants and regulatory authorities to form judgments on an instrument and/ or the issuer and take decisions on them.

Banks do undertake structured rating exercises with quantitative and qualitative inputs to support a credit decision, whether it is sanction or rejection [2]. However they may be influenced by the rating whether available of an instrument of a particular party by an external agency even though the purpose of a bank's credit rating may be an omnibus one that is to check a borrower's capacity and competence while that of an external agency may be limited to a particular debt instrument. Credit scoring tasks can primarily be classified into two distinct Models [3]. The first Model consists of financial information and demographic information about the loan applicant. Where as the second Model deals with existing customers and their payment history information. This is distinguished from the first type because this takes into account the customer's payment pattern on the loan and the model is called behavioral scoring. Also, as per the BASEL II committee recommendations and instructions, it has almost become a regulatory requirement for the banks to use sophisticated credit scoring models based on major factors shown in (Fig 1) for enhancing the efficiency of capital Base.

Fig 1. Use Balanced Score Card System for proper loan risk management.

By assessing credit portfolio, banks will be able to protect themselves against undesirable problems with the borrowers and potential partners who may turn insolvent [4]. Finance managers and credit experts recommend using Scorecard System as in (Fig 2) to measure loan and credit risks.

Fig 2. Use reliable tools to measure credit risks

Capital adequacy: Also known as Capital to Risk (Weighted) Assets Ratio (CRAR) is a measure of the amount of a bank's core capital expressed as a percentage of its risk-weighted asset.

Gross Debt Service Ratio: It is property taxes plus annual loan payment divided by Gross Customer Income and multiplied by 100.

Customer credit quality: This figure is based on credit history and reports.

Lenders can also use credit scores to determine who qualifies for what amount loan and at what interest rate. Research on credit risk management has shown that models based on Fuzzy Logic can yield more accurate predictions of credit risk than the other traditional models.

III. FUZZY EXPERT SYSTEM

Fuzzy Expert System is simply an expert system that uses a collection of fuzzy membership functions and rules, instead of Boolean logic, to reason about data. Fuzzy Inference System (FIS) incorporates fuzzy inference and rule-based expert systems. There are different types of fuzzy systems . Mamdani fuzzy systems and TSK fuzzy systems are two types of fuzzy systems commonly used and have different ways of knowledge representation. TSK (Takagi-Sugeno-Kang) fuzzy system was proposed in an effort to develop a systematic approach to generate fuzzy rules from a given input–output data set [5]. Rules in this fuzzy system are like:

If x1 is A1 AND/OR x2 is A2 Then y = f(x1, x2) (1)

Where A1 and A2 are fuzzy sets and y is a (usually linear) function of crisp variables. In order to perform inference operations, the output of each rule have to be weighted. For example regarding the jth rule

Wj=AND *method* ((x_1), (X_2))
Rj: If x1 is Aj AND x2 is Bj Then yj = fj(x1, x2) (2)

The weight wj is computed as the membership functions of Aj and Bj, respectively, and the AND method is the operation defined by the AND operator which is usually the ''min'' operation. Then the final output of the system will be obtained by final output.

Final output = j Wj Yj j Wj (3)

Mamdani fuzzy system was proposed as the first attempt to control a steam engine and boiler combination by a set of linguistic control rules obtained from experienced human operators. Rules in Mamdani fuzzy systems are like these:

If x1 is A1 AND/OR x2 is A2 Then y is B1 (4)

Where A1, A2 and B1 are fuzzy sets. The fuzzy set acquired from aggregation of rules' results will be defuzzified using defuzzification methods like centroid (center of gravity), max membership, mean-max, and weighted average. The centroid method is very popular, in which the ''center of mass'' of the result provides the crisp value. In this method, the defuzzified value of fuzzy set A, d(A), is calculated by the formula:

$$d(A) = {}_A\, x.\, \mu_A(x)\, dx \qquad {}_A\, \mu_A(x)\, dx \quad (5)$$

Where μ_A is the membership function of fuzzy set A. Various possible conditions of parameters are stated in form of fuzzy sets, the Mamdani fuzzy systems will be utilized due to the fact that the fuzzy rules representing the expert knowledge in Mamdani fuzzy systems, take advantage of fuzzy sets in their consequences, while in TSK fuzzy systems, the consequences are expressed in form of a crisp function.

A. Fuzzy Inference System

A Fuzzy Inference System (FIS) is a way of mapping an input space to an output space using fuzzy logic [5]. A FIS tries to formalize the reasoning process of human language by means of fuzzy logic (that is, by building fuzzy IF-THEN rules) as shown in (Fig 3). FIS are used to solve decision problems, i.e. to make a decision and act accordingly.

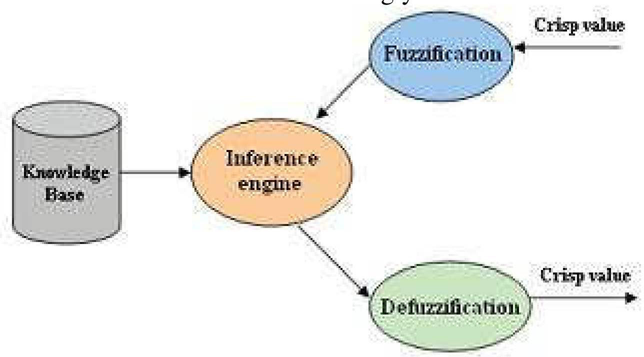


Fig 3. Structure of Fuzzy Inference System

Fuzzification module: transforms the system inputs, which are crisp numbers, into fuzzy sets. This is done by applying a fuzzification function.

Knowledge base: stores IF-THEN rules provided by experts.

Inference engine: simulates the human reasoning process by making fuzzy inference on the inputs and IF-THEN rules.

Defuzzification module: transforms the fuzzy set obtained by the inference engine into a crisp value.

B. Flow Chart for Credit Risk Rating Assessment [6]

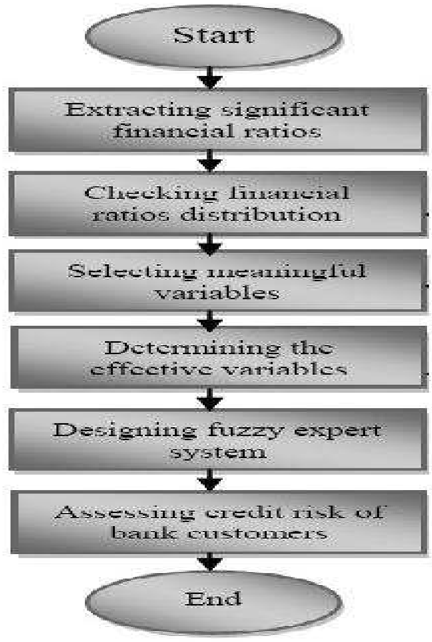


Fig 4. Flow chart for credit risk rating assessment

C. Selecting Inputs to the Credit Rating Model

The first step in developing the model is to decide on the input parameters. To get an accurate credit rating, both quantitative financial data and qualitative data are required. Due to lack of availability of qualitative data, only quantitative data is taken into consideration [6,7] . The following five ratios are selected as input parameters.

- Current Ratio
- Debt Ratio
- Return on Sales
- Average Collection Period
- Quick Ratio

Current Ratio: The current ratio is a financial ratio that measures whether or not a firm has enough resources to pay its debts over the next 12 months. It compares a firm's current assets to its current liabilities.

$$\text{Current ratio} = \frac{\text{Current Assets}}{\text{Current Liabilities}}$$

The current ratio is an indication of a firm's market liquidity and ability to meet creditor's demands. Acceptable current ratios vary from industry to industry and are generally between 1.5 and 3 for healthy businesses. If a company's current ratio is in this range, then it generally indicates good short-term financial strength. If current liabilities exceed current assets (the current ratio is below 1), then the company may have problems meeting its short-term obligations.

Debt Ratio: is a financial ratio that indicates the percentage of a company's assets that are provided via debt. It is the ratio of total debt (the sum of current liabilities and long-term liabilities) and total assets (the sum of current assets, fixed assets, and other assets such as 'goodwill'). Debt Ratio can be termed as total liabilities divided by total assets. The debt/asset ratio shows the proportion of a company's assets which are financed through debt. The higher the ratio, the greater risk will be associated with the firm's operation.

Return on Sales: Return on sales (ROS) is a ratio widely used to evaluate an entity's operating performance. It is also known as "operating profit margin" or "operating margin". ROS indicates how much profit an entity makes after paying for variable costs of production such as wages, raw materials, etc. (but before interest and tax). It is the return achieved from standard operations and does not include unique or one off transactions. ROS is usually expressed as a percentage of sales (revenue).

Average Collection Period: The approximate amount of time that it takes for a business to receive payments owed, in terms of receivables, from its customers and clients. possessing a lower average collection period is seen as optimal, because this means that it does not take a company very long to turn its receivables.

Quick Ratio: An indicator of a company's short-term liquidity. The quick ratio measures a company's ability to meet its short-term obligations with its most liquid assets. The quick ratio is more conservative than the current ratio because it excludes inventories from current assets.

D. Fuzzy Logic Toolbox in MATLAB

In modeling the inputs for the fuzzy inference engine, since some attributes may be psychographic in nature, the input variables are divided into various categories. Here, all inputs are divided into three categories "low", "medium" and "high" based on the probability of default calculated from the existing data. This process is called Fuzzification. According to these financial ratios as inputs (Table 1) and Credit Risk Degree of Customer as output (Table 2), the experts opinions in the form of rules have been obtained (Table 3).Then, a Mamdani's Fuzzy Expert system has been designed with the financial ratios as the system Inputs and customers credit risk degree as the output [7-9].

TABLE I. THE INPUTS OF FUZZY EXPERT SYSTEM

SIGN	INPUTS	INTERVAL	MEMBERSHIP FUNCTION TYPE	LINGUISTIC TERMS
CR	CURRENT RATIO	[0,2]	GBELL	LOW(L), MEDIUM(M), HIGH(H)
DR	DEBT RATIO	[0,1]	GBELL	LOW(L), MEDIUM(M), HIGH(H)
ROS	RETURN ON SALES	[0,3]	GBELL	LOW(L), MEDIUM(M), HIGH(H)
ACP	AVG COLLECTION PERIOD	[0,300]	GBELL	LOW(L), MEDIUM(M), HIGH(H)
QR	QUICK RATIO	[0,1]	GBELL	LOW(L), MEDIUM(M), HIGH(H)

TABLE II. THE OUTPUT OF FUZZY EXPERT SYSTEM

SIGN	INPUTS	INTERVAL	MEMBERSHIP FUNCTION TYPE	LINGUISTIC TERMS
CRDC	CREDIT RISK DEGREE OF CUSTOMER	[0,10]	GAUSSIAN1	LOW(L), MEDIUM(M), HIGH(H)

TABLE 3. Rules For Designing Fuzzy Expert System (62 Rules in Total).

RULE NO.	CR	DR	ROS	ACP	QR	CRDC
1	M	H	H	L	M	L
2	M	H	H	M	M	H
3	L	L	H	H	L	L
4	M	L	H	L	H	L
5	M	H	M	H	L	M
6	H	M	M	H	L	M
7	H	L	M	H	H	M
8	L	H	M	L	M	H
9	H	M	L	M	M	M
10	H	H	L	M	M	M
11	H	H	L	M	H	M
12	H	L	L	H	M	H
13	L	L	H	L	M	M
14	L	H	H	M	L	M
15	L	L	L	L	L	H
16	H	H	H	H	H	L
17	H	L	H	M	H	L
18	L	M	H	H	L	M
19	M	H	L	L	H	H
20	M	L	H	H	L	M

Rules are implemented using AND Function in Fuzzy Logic Rule Base in MATLAB according to Bank experts' opinions for the input and output variables. The suitable Membership functions have been defined in (Fig 5-10)

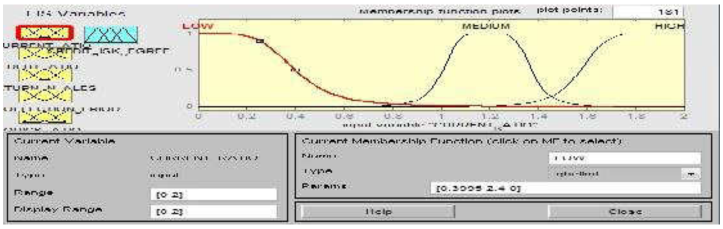


Fig 5. Three Gbell Membership functions for Current Ratio (L,M,H)

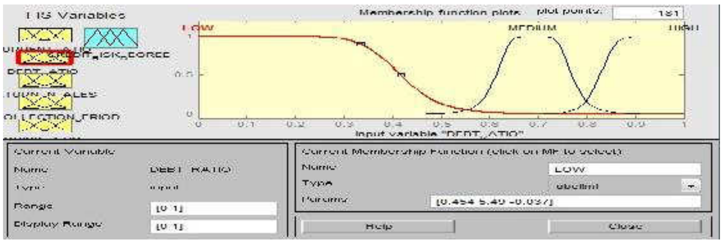


Fig 6. Three Gbell Membership functions for Debt Ratio (L,M,H)

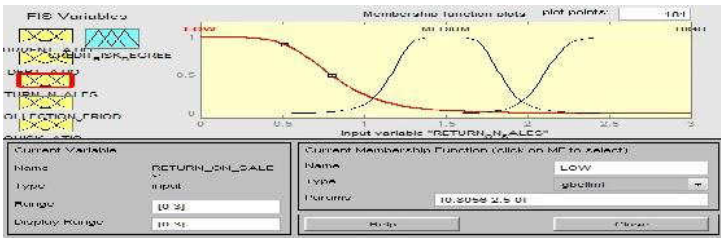


Fig 7. Three Gbell Membership functions for Return on Sales (L,M,H)

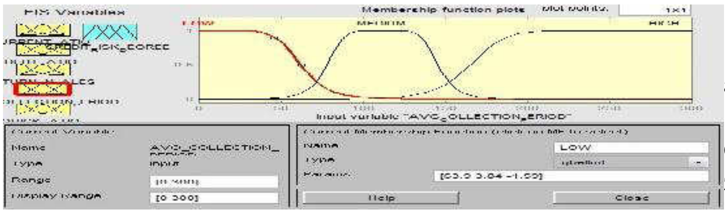


Fig 8. Three Gbell Membership functions for Avg Collection Period (L,M,H)

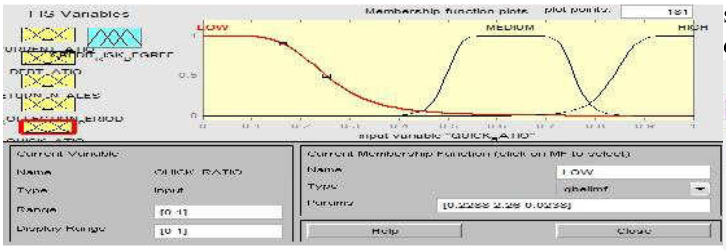


Fig 9. Three Gbell Membership functions for Quick Ratio (L,M,H)

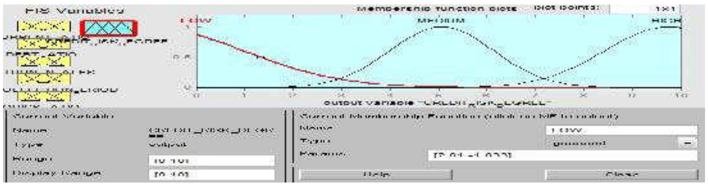


Fig 10. Three Gaussian Membership functions for Credit Risk Degree (L,M,H)

IV. COMPILING USING GUIDE IN MATLAB.

We can build the Graphical user interface using GUIDE (GUI BUILDER) in MATLAB. We use five slider buttons for Input and one slider button for output and adjust their range in their properties section. One push button is used as “CALCULATE” button and the other push button for “CLOSE”. We insert Text boxes to Label our inputs and output. GUI file is a ‘.fig’ file. As we save our ‘.fig’ file, a corresponding ‘.m’ file is generated by MATLAB which is a MATLAB file. This is a code file for GUI and here we can apply fuzzy logic codes to make our GUI work on Fuzzy Expert System’s Logic. We now do the additional coding in ‘.m’ file to implement Fuzzy Logic.

```
p = get(handles.slider1,'Value');
q = get(handles.slider2,'Value');
```

Where p,q are inputs. For ex: Current Ratio, Debt Ratio, etc.

```
a=newfis('fis_file_name');
a = addvar(a,'input','x',[0 2]);
a = addmf(a,'input',1,'A1','gbellmf',[0.39989  2.4  0]);
a = addmf(a,'input',1,'A2','gbellmf',[0.18301  2.5  1.19]);
a = addmf(a,'input',1,'A3','gbellmf',[0.36592  3.2  1.94]);
```

We can define the Membership functions type and their range in this format for every input function as well as the output.

Rules can be defined in ‘.m’ file in the following format :

```
rulelist=[ ...
  3 2 2 3 1 3 1 1
  2 3 3 1 2 1 1 1 ];

a=addrule(a,rulelist);
```

Command which evaluates our Fuzzy (fis) data is :

z = evalfis([p q r s t], a)

After saving ‘.m’ file, we can compile the ‘.m’ file into standalone executable ‘.exe’ as shown in (Fig 11) using deploytool command in MATLAB for easy end user interface.

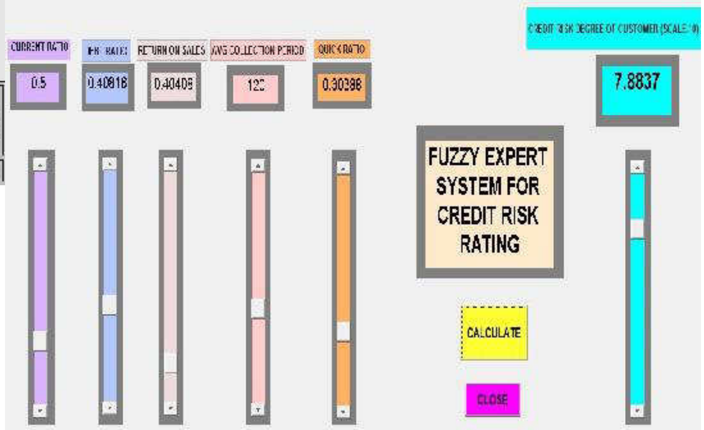


Fig 11. Compiling up of our work into a user end interface ie. standalone executable (.exe) file using GUIDE and Deploytool in MATLAB

The output can be analyzed for different sets of inputs by adjusting the slider position of inputs and using CALCULATE button (Fig 11). We have taken our Output (CREDIT RISK DEGREE OF CUSTOMER) on SCALE-10. Any company which scores greater than point 6 is taken as a defaulter company and any company with a score less than 4 is taken as a non-defaulter company. The accuracy of our system purely depends on expertness provided to system ie. Rule Base.

V. CONCLUSION

The credit risk rating model developed in this research to solve credit risk management problem is fully capable of self learning. Once trained to the expertness according to the Rule base, it is capable of discriminating good and bad accounts with better and consistent accuracy. In our model, proper financial ratios which are effective on the experts decision making process and the acquired rules via bank experts, a Mamdani's Fuzzy Expert system has been developed which predicts the bank Customers’ Credit Risk. This Fuzzy Expert system is used as the final predicting model for credit assessment of the bank's customers. The designed model helps the banks and other financial institutions to identify the level of credit risk involved for their customers based on their financials and hence prevent insolvency.

REFERENCES

[1] Tomasz Korol (2012). Fuzzy Logic in Financial Management, Fuzzy Logic - Emerging Technologies and Applications, Prof. Elmer Dadios (Ed.), ISBN: 978-953-51-0337-0.

[2] “How do banks measure Credit Risk”
(www.credit-risk-measurement.com/how-do-banks-measure-credit-risks.htm)

[3] Credit risk guide (www.credit-risk-measurement.com)

[4] R. Malhotra and D. K. Malhotra “Differentiating between good credits and bad credits using neuro-fuzzy systems", European Journal of Operational Research, 2002, 136(1), pp. 190–211.

[5] Mamdani’s method, eMathTeacher: Mamdani’s Fuzzy Inference Method.
(www.dma.fi.upm.es/java/fuzzy/fuzzyinf/mamdani_en.htm)

[6] M. C. Chen and S. H. Huang, "Credit scoring and rejected instances reassigning through evolutionary computation techniques". Expert Systems with Applications, 2003, 24 (4), pp. 433-441.

[7] T. S. Lee, C-C. Chiu, C-J Lu, and I-F. Chen, "Credit scoring using the hybrid neural discriminant technique", Expert Systems with Applications, 2002, 24 (4), pp. 433-441.

[8] Zahedi, F., Intelligent Systems for Business, Wadsworth Publishing, Belmont, California, 1993.

[9] ALTROCK, C. Fuzzy Logic & Neurofuzzy—Applications in Business & Finance, Prentice Hall, USA, 1996, 375 pp., ISBN 0-13-591512-0.

Eligibility of Prospective Student for Admission to Nursery Class Using Fuzzy Inference System

Manraj Singh
Department of Electronics Technology
Guru Nanak Dev University
Amritsar, India
manraj@gmail.com

Mukesh Kumar
Department of Electronics Technology
Guru Nanak Dev University
Amritsar, India
mukesh8833@yahoo.in

Abstract— **Education is a fundamental aspect of our lives. In this age of constantly changing technology and trends there is competition at in every sphere of life. Even at primary level there is a huge rush for admissions. Every school of every city/town has its own criteria for selecting prospective students which result in indulgent in unfair practices. The project presents a decision model using fuzz inference system to identify the eligibility of the concerned students. The eligibility criterion of various schools was studied in order to make the conditions. Variables for proximity to school, number of siblings, alumni and interstate students are considered and are treated as member functions. Each are further divided into low, medium and high or 2, 3 and 4. The FIS calculates the mathematical values according to which we can decide whether a child is eligible for admission. The usage of fuzzy inference system would offer possible justifications to set a new approach in identifying prospective students. It provides an impartial and balanced answer rather than take decisions from a biased point of view.**

Keywords—Nursery admissions, Fuzzy rule, Fuzzy logic, Eligibility

I. Introduction

Our current economy and society has changed. It has become competitive with competition entering into every stage of our lives. It affects every strata of society. Even in matters of education there is competition for the best schools and colleges. People realize that this is a winner takes it all situation. There is a mad rush to have the best things in life. People want their children t go the best schools, best universities and colleges and get the best jobs. There is a large pool of students for the organizations to choose from. In many cases there is partiality and nepotism regarding the selection of students to colleges. Hence there is a need for a fair system which s transparent and clear.

Nursery has become an important stepping stone for a child as he/she begins their long journey through childhood, grows up through various stages of education and age. People are very concerned about which schools their children should go to. An increased pressure on our schools to choose from huge number of students has led to opaqueness in the selection process. There are wide scale donations and money is exchanged under the desk, favouritism and nepotism. The problem underlines the need for transparent, impartial and robust system which is impervious to emotions of favouritism and economic partiality.

II. Problem

The state of Delhi has to face a number of problems regarding the nursery school admissions every year. Matters get out of hand sometimes which results in massive outrage by the parents towards the schools. The judiciary and the state government have to step in and take matters into their own hands in order to solve the problem. Not declaring the selection procedure, favouring a particular community of people, or people of certain professions such as doctors, engineers, asking for large amounts of money from the parents as donations for the school, not following the guidelines provided by the ministry of HRD, are some of the problems parents have to face at the hands of the school authorities.

III. FUZZY LOGIC MODEL

We have used MATLAB to solve this multitude of problems. MATLAB provides a wide range of facilities with which can accomplish numerous tasks. Decision making is one of them.

FIS toolbox in MATLAB is one such system which designed for the special purpose of decision making when large number of variables are concerned. We had initially collected data from various schools in Delhi regarding their procedure for admission into nursery classes. After collecting data from number of schools it was found that there are 4 major conditions depending upon which decisions are made. They are

- Distance from school i.e. proximity
- Number of siblings studying in the school
- Alumni
- Interstate cases

Each schools award points according to each category. The total no of points is 100. We have observed that preference is given to the students living close to the school followed by the number of siblings of the child studying in that school, any previous family member or the children of alumni of the school. Students whose parents have a transferable job also get points. Some of the schools are minority schools i.e. they have been specially established for the education of the minorities. They provided a special quota for the children from the concerned minority. The rules regarding selection of minority students have not been defined. Hence the schools can follow any criteria to select students from those communities. This project has been designed for non-minority schools i.e. normal schools with no quota. We use the FIS editor present in the Fuzzy logic tool box. We first define the input variables which form our membership functions (which are our criteria). Each variable has its own range. We then define the output variable and specify its range. After defining all variables we open the rules box and specify each possible combination of the variable and its corresponding result. This is a tedious task as the number of rules is large and a minor mistake can result in serious consequences. Our main program is complete.The whole package operations of fuzzy logic have been translated magnificently into fuzzy inference system (FIS). FIS based on fuzzy rules has been applied to numerous engineering applications such as control, signal processing, and pattern classification problems. Specifically, this paper aims to propose the eligibility of prospective students in nursery admissions depending upon which parents and school authorities can make a sound decision using FIS.

IV. Fuzzy Inference System

Fuzzy logic is a form of many-valued logic; it deals with reasoning that is approximate rather than fixed and exact. Compared to traditional binary sets (where variables may take on true or false values), fuzzy logic variables may have a truth value that ranges in degree between 0 and 1. Fuzzy logic has been extended to handle the concept of partial truth, where the truth value may range between completely true and completely false. Furthermore, when linguistic variables are used, these degrees may be managed by specific functions.

A Fuzzy Inference System (FIS) is a way of mapping an input space to an output space using fuzzy logic.

FIS uses a collection of fuzzy membership functions and rules, instead of Boolean logic, to reason about data.

The rules in FIS (sometimes may be called as fuzzy expert system) are fuzzy production rules of the form:

if p then q, where p and q are fuzzy statements.

For example, in a fuzzy rule

if x is low and y is high then z is medium.

Here x is low; y is high; z is medium are fuzzy statements;

x and y are input variables; z is an output variable, low, high, and medium are fuzzy sets.

Most tools for working with fuzzy expert systems allow more than one conclusion per rule. The set of rules in a fuzzy expert system is known as knowledge base. The functional operations in fuzzy expert system proceed in the following steps. (1) Fuzzification (2) Fuzzy Inferencing (apply implication method) (3) Aggregation of all outputs (4) Defuzzification

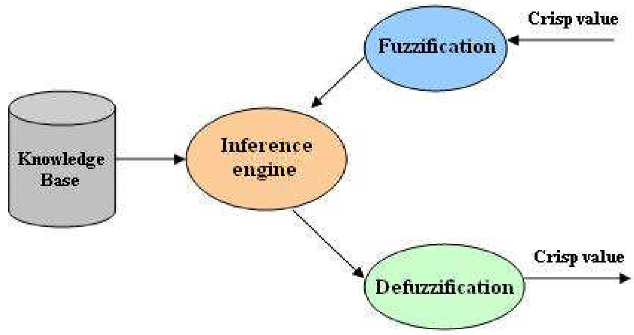


Fig.1.Fuzzy logic architecture

V. Using the Model

We have used FIS to determine whether a child is eligible for admission or not. We have earlier described the procedure in brief in the previous section. The details are as follows

STEP 1 : *Defining input and output*

We have taken 4 inputs which form the basis of our decision making model. They are

Proximity: This is the most important factor while choosing a particular student. It refers to the distance of the child's home from the school. The closer the house is to the school more the points he gets and more chances of him being selected.

Sibling: This variable stores the number of siblings i.e. brothers/sisters of the child studying in that school currently. It is advantageous to have a sibling studying in a school where the child is to be admitted as the school authorities are aware of the social and economic conditions of the family.

Alumni: This variable stores the number of relatives of the child who have passed out from the same school. The relatives should only be parents or closest uncle/aunt.

Interstate: This variable tells us if the child has been transferred from a different state i.e. if the parents have job which gets transferred from place to place.

And output variable as ELIGIBLE

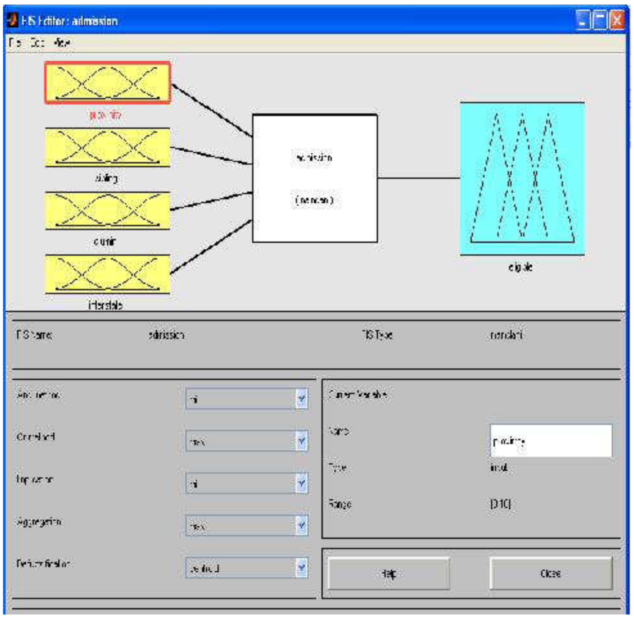


Fig.2.Simulations

STEP 2: Defining Fuzzy Sets for System Variables System variables need to fuzzify in order to obtain fuzzy membership. The system recognizes the input and output variables and defines its memberships.

STEP 3 : Defining Fuzzy Rules The rules are designed as to describe the importance of the factors responsible for eligibility.

Example: (a) If the student has he following

Proximity	Sibling	alumni	interstate
2	1	1	0

Then his categories will be

Proximity	Sibling	alumni	interstate
Closest	One	a1	none

Then according to our requirements the Student should be selected. Therefore we select the option of YES under ELIGIBLE

Proximity	Sibling	alumni	interstate	eligible
Closest	One	a1	none	yes

The student has the following

Proximity	Sibling	alumni	interstate
5	2	1	0

Then his categories will be

Proximity	Sibling	alumni	interstate
Away	Two	a1	none

Then according to our requirements the Student should not be selected. Therefore we select the option of NO under ELIGIBLE

Proximity	Sibling	alumni	interstate	eligible
Away	Two	a1	none	no

STEP 4 : Defuzzification Finally defuzification step is needed to convert all input data into linguistic terms that can be used to observe the likelihoods of selection. The defuzzification process transforms the fuzzy set into a crisp value that is meaningful to end-user.

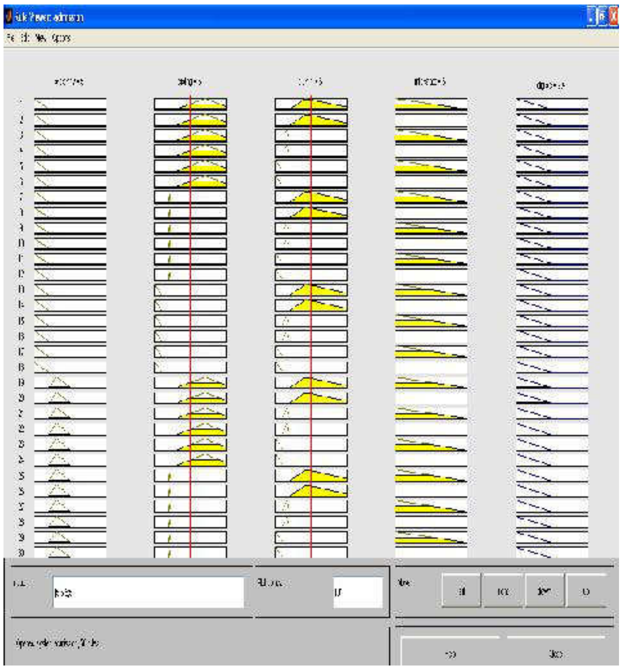


Fig.3.Rule viewer of the actual project on candidate selection.

Table I. Entries of inputs and corresponding outputs

Proximity	Sibling	Alumni	Interstate	Eligible
2	1	1	1	Y
7	1	1	1	N
5	1	2	1	N
3	2	2	1	Y
4	2	3	0	Y
1	2	0	0	Y
3	3	1	0	Y
2	3	0	0	Y
8	1	0	0	N
3	0	1	1	N
1	1	1	1	Y
7	0	0	0	N
2	0	0	1	N
9	1	0	1	N

7	2	0	1	N

VI. CONCLUSION

This project makes use of fuzzy logic to solve a real world problem. The project aims to solve the problems of the thousands of parents and children who have to face uncertainty and frustration every year. In this paper, the fuzzy rules based method to classify the likelihoods of eligibility/selection of students was utilized. Education is fundamental element in our lives and the beginning of school is the first stepping stone. We hope that this project makes matters easier and transparent for the authorities and the parents.

REFERENCES

[1] MATLAB guide

[2] Lazim Abdullah1 and Mohd Nordin Abd Rahman, "Employee likelihood of purchasing health insurance using fuzzy inference system" in IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 2, January 2012

[3] Zalinda Othman, Khairanum Subari and Norhashimah Morad, "Application of fuzzy inference systems and genetic algorithms in integrated process planning and scheduling", Academic Staff Training Scheme ,research grant NO:305/PTEKIND/622140

[4] Maedeh Rasoulzadeh, "Facial expression recognition using fuzzy inference system" ISSN:2277- 3754 International Journal of Engineering and Innovative Technology (IJEIT) Volume 1, Issue 4, April 2012.

[5] EKONG, Victor , EKONG, Uyinomen , UWADIAE, Enobakhare , ABASIUBONG, and ONIBERE, "A fuzzy inference system for predicting depression risk levels" Emmanuel African Journal of Mathematics and Computer Science Research Vol. 6(10), pp 197-204, November, 2013.

Track-1
Technical Session 6
ANALOG/DIGITAL/VLSI/ANTENNA

Comparative Analysis of Multistage Decimation Filters

Sajjan Godara
Department of Electronics
Punjabi University, Patiala
Punjab, India
sajjangodara@gmail.com

M.S. Pattar
Department of Electronics
Punjabi University Patiala,
Punjab, India

also be in hardware due to its symmetric

***Abstract*—This paper is concerned with multistage designs of decimation filters. Decimation filter has wide applications in both analog and digital systems for data rate conversion as well as filtering. CIC filters satisfy linear phase filter requirements, which are required for time sensitive features like video and speech. All the filter coefficients are unity in case of CIC filters, resulting in reduced hardware requirement. In the presented paper we have designed two decimation filters one consisting of CIC filter and three Half-band filters while the other filter makes use of Nyquist filter cascaded with CIC decimation filter. An IIR band pass filter is used at the final stage in both of the designs so as to get a signal of desired bandwidth. All the filters are designed in MATLAB. A comparison in terms of hardware used and corresponding calculations per input sample is made on the basis of results obtained.**

KeyWords*—CIC, Decimation, Half-band filters, Lth-band filters, Nyquist Design

I. INTRODUCTION

Fast sampling rates offer several benefits, including their ability to digitize wideband signals, reduced complexity of anti-alias filters, and lower noise power spectral density. The result is improved SNR in the system. Digital filtering is a computational process used for transforming a discrete sequence of numbers (the input) into another discrete sequence of numbers (the output) having a modified frequency domain spectrum. Digital filtering algorithms are most commonly implemented using general purpose digital signal processing chips for higher sampling rates. The decimation filter (decimator) is one of the basic building block of a sampling rate conversion system. The decimation filter performs two operations: low-pass filtering as well as down-sampling.. It has been widely used in such applications as speech processing, radar systems, antenna systems and communication systems. Considerable attention has been focused in the last few years on the design of high efficiency decimation filters.

In 1981, Eugene Hogenauer [1] invented a new class of economical digital filters for decimation and interpolation (converting the sampling rate from low to high) called a cascaded integrator comb (CIC) filter. This filter was composed of an integrator part and a comb part. No multipliers were required and the storage requirement was reduced when compared with other implementations of decimation filters. The CIC filter can

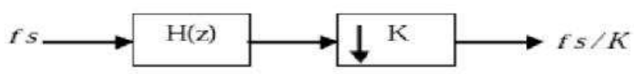


Fig. 1. Conceptual View of a Decimation Filter

This filter is a combination of digital integrator and digital differentiator stages, which can perform the operation of digital low pass filtering and decimation at the same time. Y. Djadi, et al. [2] designed a programmable decimation and interpolation digital filter based on the CIC structure. The circuit was configurable as either a decimation filter or an interpolation filter and the conversion ratio was programmable to any integer value from 10 to 256. The filter was designed with MOSIS 1.2 micron CMOSN standard cell libraries and data input rate could be as high as 50MHz.

Hyuk Jun Oh and Yong H. Lee [3] used a simple interpolated second-order polynomial filter (ISOP) cascaded with a CIC decimation filter to effectively reduce the pass band distortion caused by CIC filtering with little degradation in aliasing attenuation.

Hong-Kui Yang and W. Martin Snelgrove [4] decomposed the decimation ratio R into two factors, namely, R1*R2 (R1 and R2 are integers) to implementing a CIC decimation filter. Instead of using one CIC filter to decimate the high speed digital signal, they used two CIC decimation filters: one CIC filter with a decimation ratio of R1, and one with a decimation ratio of R2. The implementation of the first decimator was based on a polyphase decomposition.

Another way of implementation the CIC decimation filter was presented by Yonghong Gao et al. [5] in which the decimation factor N was constrained to an M-th power-of-two. The transfer function can be rewritten as the product of M identical low-order FIR filters. The sampling rate in this implementation decreased at every stage by a factor of two. Fred Harris et.al. [6] in 2008 proposed Simple multiplier free Sin based compensator with only two adders. The proposed method is computational efficient and less complex.

G.J.Dolecek et.al. [7], [8] in 2009, designed a multiplier less CIC compensation filter based on the 2M order filter and the sharpening technique. This technique attempts to improve the pass band and the stop band of a symmetric non recursive filter using the multiple copies of the same filter. G. Javanovic Dolecek et. al. [8] presented an efficient modification of the CIC cosine decimation filter in May

2010. The second order compensator filter is introduced at low rate in order to improve the pass-band performance.

Rajeev Ratan, Sanjay Sharma. al. [10] carried out performance evaluation of multirate filters for digital down converters. The proposed filter structures have wide applications in the designing of sample rate converters, analog to digital converter, decimators and interpolators. Li Jia-huiet. al. [11] in 2012 proposed a kind of comb filter with low power consumption. This comb filter can be applied to the first stage of digital decimation filter for delta-sigma AID converter, and realize the 32 multiples frequency reducing.

V. Jayaprakasan and M. Madheswaran al. [12] in 2013 proposed a cascade design of sharpened CIC filter and polyphase structure of FIR for efficient compensation in decimation, which has better pass band and stop band performance. In this paper we analyze the performance of CIC decimation filter with decimation factor 8 by cascading it to a compensation filter in two different ways to achieve a CIC decimation filter structure with an improved frequency response.

II. CASCADED INTEGRATOR COMB (CIC) FILTER

Cascaded integrator-comb (CIC), or Hogenauer filters, are multirate filters used for realizing large sample rate changes in digital systems. CIC filters are multiplierless structures, consisting of only adders and delay elements which is a great advantage when aiming at low power consumption. They are typically employed in applications that have a large excess sample rate. That is, the system sample rate is much larger than the bandwidth occupied by the signal. CIC filters are frequently used in digital down-converters (DDCs) and digital up-converters. The CIC filter is a class of hardware-efficient linear phase finite impulse response (FIR) digital filters, consists of an equal number of stages of ideal integrator filters and comb filters. Its frequency response may be tuned by selecting the appropriate number of cascaded integrator and comb filter pairs. The highly symmetric structure of a CIC filter allows efficient implementation in hardware. However, the disadvantage of a CIC filter is that its pass band is not flat, which is undesirable in many applications. Fortunately, this problem can be alleviated by a compensation filter. CIC filters achieve sampling rate decrease (decimation) and sampling rate increase (interpolation) without using multipliers. The CIC filter first performs the averaging operation then follows it with the decimation. The transfer function of the CIC filter in z-domain is given in equation (1).

$$H[z] = H[z]_I^p \cdot H[z]_C^p = \frac{\left(1-z^{-K}\right)^p}{\left(1-z^{-1}\right)^p} = \left(\frac{1-z^{-K}}{1-z^{-1}}\right)^p \qquad (1)$$

In equation (1), K is the oversampling ratio and p is the order of the filter. The numerator (1–z–K) ^ p represents the transfer function of a differentiator and the denominator 1/ (1– z –1) ^ p indicates the transfer function of an integrator. A simple block diagram of a first order CIC filter is shown in Figure 2. In a CIC filter, the integrators operate at high sampling frequency (fs), and the comb filters operate at low frequency (fs/N) The clock divider circuit divides the oversampling clock signal by the oversampling ratio, N after the integrator stage. By operating the differentiator at lower frequencies, a reduction in the power consumption is achieved.

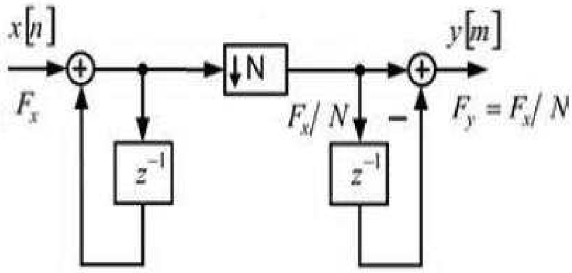


Fig. 2. Structure of First Order CIC Filter

III. DECIMATION TECHNIQUES

A .Using FIR Filters

The FIR Decimation block resamples the discrete-time input at a rate K times slower than the input sample rate. The block down samples the filtered data to a lower rate by discarding (K-1) consecutive samples following every sample retained. The design of FIR filter is based on the often added requirement that the phase response be linear. FIR filter is not based on any feedback path and can easily be designed to be linear phase by making the coefficient sequence symmetric i.e. equal delay at all frequencies. This property is sometimes desired for phase-sensitive applications. The structure of a FIR filter is given in Figure 3.

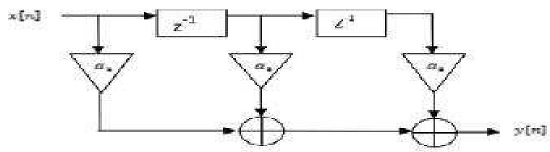


Fig 3: Structure of FIR Filter

The main disadvantage of FIR filters is that considerably more computation power in a general purpose processor is required compared to an IIR filter with similar sharpness or selectivity, especially when low frequency (relative to the sample rate) cutoffs are needed. However many digital signal processors provide specialized hardware features to make FIR filters approximately as efficient as IIR for many applications.

B. Using IIR Filters

In multirate applications, the computational requirements for FIR filters can be reduced by the sampling rate conversion factor. However, such a degree of computation savings cannot be achieved in multirate implementations of IIR filters. This is due to the fact that every sample value computed in the recursive loop is needed for evaluating an output sample. Based on the polyphase decomposition, several techniques have been developed which improve the efficiency of IIR decimators and interpolators. Infinite impulse response (IIR) filters are used in applications where the computational efficiency is the highest priority. It is well known that an IIR filter transfer function is of a considerably lower order than the transfer function of an FIR equivalent. The drawbacks of an IIR filter are the nonlinear phase characteristic and sensitivity to quantization errors.

C. Using CIC Filters

These filters require no multipliers and use limited storage thereby leading to more economical hardware implementations. They are designated Cascaded Integrator-Comb (CIC) filters because their structure consists of an integrator section operating at the high sampling rate and a comb section operating at the low sampling rate. Using CIC filters, the amount of passband aliasing or imaging error can be brought within prescribed bounds by increasing the number of stages in the filter. However, the width of the passband and the frequency characteristics outside the passband are severely limited.

IV. Lth-BAND FILTERS

Digital Lth-band FIR and IIR filters are the special classes of digital filters, which are of particular interest both in single-rate and multirate signal processing. The common characteristic of Lth-band lowpass filters is that the 6 dB (or 3 dB) cutoff angular frequency is located at /L, and the transition band is approximately symmetric around this frequency. In time domain, the impulse response of an Lth-band digital filter has zero valued samples at the multiples of L samples counted away from the central sample to the right and left directions. Actually, an Lth-band filter has the zero crossings at the regular distance of L samples thus satisfying the so-called zero intersymbol interference property. Sometimes the Lth band filters are called the Nyquist filters. Due to the zero intersymbol interference property, the Lth-band filters are very important for digital communication transmission systems. Another application is the construction of Hilbert transformers, which are used to generate the analytical signals. The Lth-band filters are also used as prototypes in constructing critically sampled multichannel filter banks. They are very popular in the sampling rate alteration systems as well, where they are used as decimation and interpolation filters in single-stage and multistage systems.

IIR filters can be designed with much smaller orders than FIR filters at the expense of the non-linear phase. Since it is difficult to design a linear phase IIR filter and a linear phase FIR filter needs high number of coefficients, we have designed a half-band FIR filter for the same application. In half-band filters, the number of taps is reduced considerably since the odd coefficients are zeros, which reduces the hardware and the power consumption.

Basically Halfband filter is a special case of Nyquist filter. Generally a Nyquist filter with L=2 is called a Half band Filter.

V. DECIMATION FILTER DESIGN

In the presented paper, we have carried out a decimation process by a factor of 128 in two different ways. One methods makes use of three half-band FIR filters and one CIC decimation filter while in the other method we have replaced the three half-band filter by a single Nyquist filter which results in decimation by a factor of 8. CIC filter in both the methods is used to decimate the input signal by a factor of 16. An IIR bandpass filter is used at the final stage in both of the designs to achieve a decimated signal of desirable bandwidth. MATLAB functions ellip and ellipord are used to design the IIR bandpass filter. Figure 5 and Figure 6

shows the block diagrams of the two decimation filter designs. All the specifications of the decimation filter are described in table I.

Table I.Filter Specifications

Filter Used	Transition Width(TW)	Stop band Attenuation	Decimation Factor(M)
CIC	354Khz	100 dB	16
First HB	105 KHz	130 dB	2
Second HB	52.5 KHz	130 dB	2
Third HB	30 KHz	130 dB	2
Nyquist	47KHz	130 dB	8

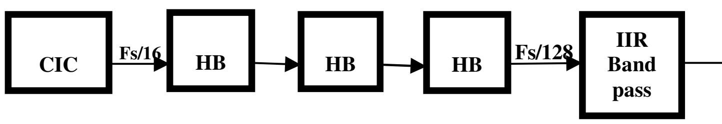


Fig. 5 Decimation Filter Consisting of Half-band Filters

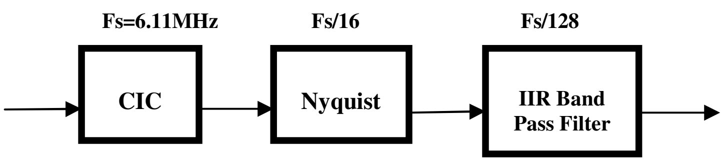


Fig. 6 Decimation Filter making use of Nyquist Filter

VI. ANALYSIS AND RESULTS

When an input signal of 6.11 MHz is decimated by a CIC decimation filter by a factor of 16, The magnitude response achieved is as that of in Figure 6. Figure 7, 8, and 9 shows the magnitude response of the three half band filters at second. third and fourth stage respectively.

Figure 10 shows the magnitude response of Nyquist filter which is used in place of three half-band filters to carry out a single stage decimation by a factor of 8 as compared to previous design in which a decimation by a factor of 2 was carried out by each half-band filter. Magnitude response of IIR band pass filter at the final stage is shown in Figure 11. Table II *shows the comparison of hardware utilization in both of the decimation filter designs.

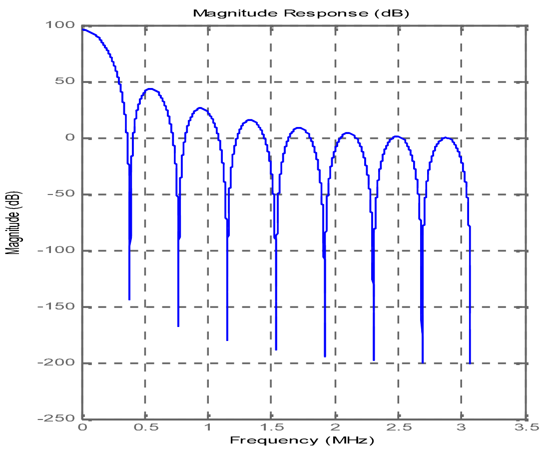


Fig. 7 Magnitude Response of CIC Filter

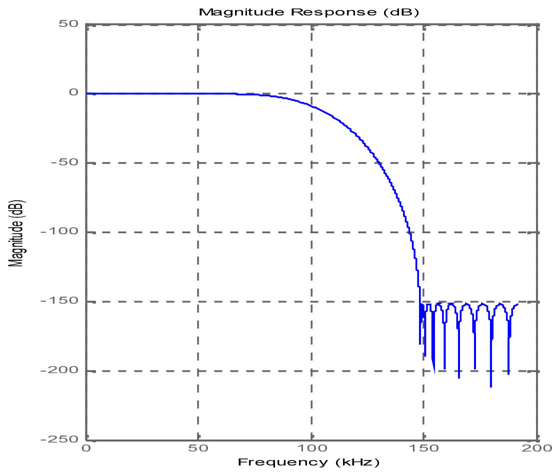


Fig.8 Magnitude Response of First Half-band Filter

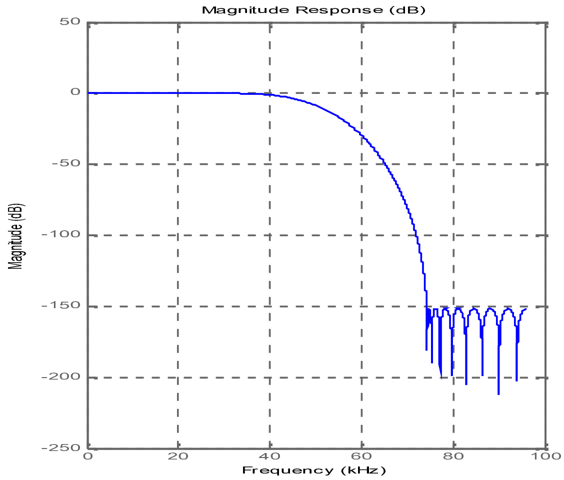


Fig. 9 Magnitude Response of second Half- band Filter

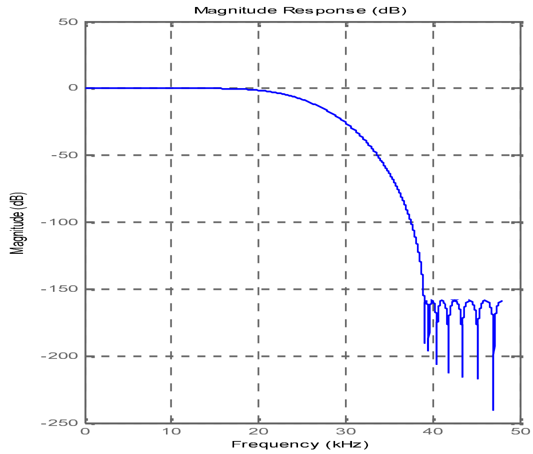


Fig. 10 Magnitude Response of third Half- band Filter

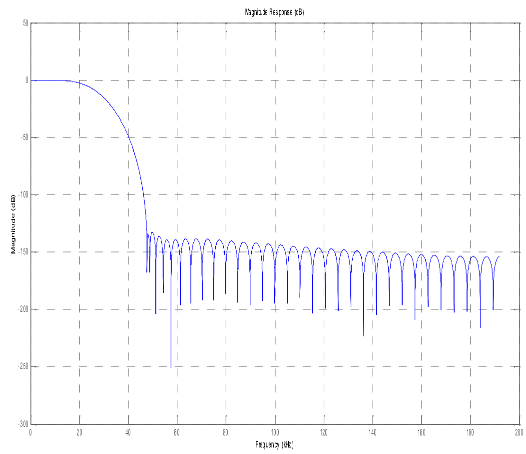


Fig. 11 Magnitude Response of Nyquist Filter

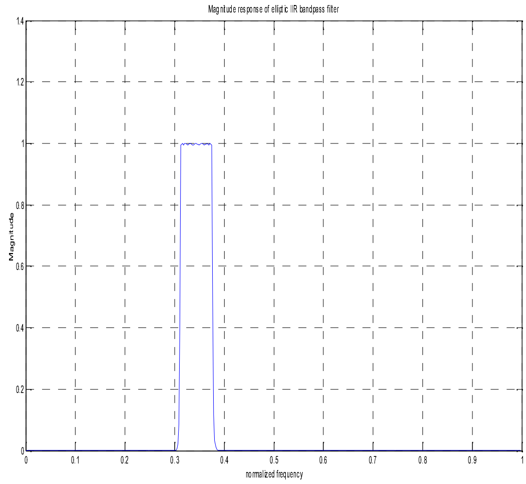


Fig. 12 Magnitude Response of Band-Pass IIR Filter at Last Stage

Table II.Comparison of the two designs

Filter Used / Hardware Utilization	One CIC +Three Half-band	One CIC+One Nyquist
Number of Multipliers Used	49	65
Number of Adders Used	54	72
Multiplications per Input Sample	0.9	0.5
Additions per Input Sample	5.1	4.75

VII. CONCLUSION

So it is clear from the table II that for the given specifications, calculations per input sample and the amount of hardware requirement changes by a noticeable amount. Use of a Nyquist filter in place of three Halfband filters for a specified decimation factor results in requirement of more numbers of multipliers and adders but at the same time calculations per input sample are decreased. Presence of bandpass filter at the final stage makes it possible to choose a signal of desirable bandwidth suitable for a particular application. Two designed decimation filter structure can be modified in accordance to the requirement of decimation factor.

REFERENCES

[1] E. B. Hogenauer, "An Economical Class of Digital Filters for Decimation and Interpolation", IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-29, pp. 155-162, April 1981.

[2] Y. Djadi, T. A. Kwasniewski, C. Chan and V. Szwarc, "A High Throughput Programmable Decimation and Interpolation Filter", Proceeding of the 1994 International Conference on Signal Processing Applications and Technology, pp. 1743- 1748.

[3] A. Kwentus, O. Lee, and A. N. Willson, Jr., "A 250 M Sample/ sec Programmable Cascaded Integrator-Comb Decimation Filter", in Proceeding VISI Signal Processing, IX, pp. 231-240, Oct-Nov. 1996.

[4] H. K. Yang and W. M. Snelgrove, "High Speed Poiyphase CIC Decimation Filters", Proceeding of 1996 IEEE International Conference On Communications, pp. II.229- II.233, Atlanta, US, May 1996.

[5] Yonghong Gao, Lihong Jia and Hannu Tenhunen, "An improved Architecture and implementation of Cascaded Integrator-Comb Decimation Filter", 1999 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, pp. 317-320, Victoria, B.C., Canada, August 1999.

[6] Jovanovic Dolecek and Fred Harris, "Design of CIC Compensator Filter in a Digital IF Receiver", 2008 IEEE.

[7] G. J. Dolecek and Fred Harris, "Design of Wideband CIC Compensator Filter for a Digital IF Receiver", Digital Signal Processing 19, ELSEVIER, pp. 827-837, April, 2009.

[8] Gordana Jovanovic Dolecek and Fred Harris, "On Design of Two-Stage CIC Compensation Filter", IEEE International Symposium on Industrial Electronics, Seoul, Korea, pp. 903-908, July 5-8, 2009.

[9] G. J. Dolecek and M. Laddomada, "An Economical Class of Droop Compensated Generalized Comb Filters: Analysis and Design", IEEE Transactions on Circuits and Systems-II, Vol. 57, No. 4, pp. 275-279, April 2010.

[10] Rajeev Ratan. Sanjay Sharma and Amit Kumar Kohli," Performance Evaluation of Multirate Filters for Digital Down Converters", International Journal of the Physical Sciences Vol. 6(12), pp. 2807-2817, 18 June, 2011

[11] Li Jia-hui, Wang-heng and Jin Yu-hong" One comb filter for lower power consumption", International Conference on Computer Science and Information Processing (CSIP) pp.965-968, April 2012 IEEE.

[12] V.Jayaprakasan and M.Madheswaran," Cascading Sharpened CIC and Polyphase FIR Filter for Decimation Filter", International Conference on Advances in Electrical and Electronics Engineering (ICAEE'2013) March 17-18, 2013 Dubai (UAE).

Stacked Dual Oxide Gate Structure MOSFET: Characteristics and Simulation

Amit Verma
Dept. of Electronics
NITTTR, Chandigarh, India
verma85.amit@gmail.com

Shivendra
Dept. of Electronics
NITTTR, Chandigarh, India

Rajesh Mehra
Dept. of Electronics
NITTTR, Chandigarh, India

***Abstract*—A 90 nm stacked dual oxides Gate structure MOSFET is designed and simulated using the Cogenda Genius TCAD simulation tool.This paper investigate the improved characteristics of various structures based on the various combination of SiO_2 and HfO_2 as the oxide between the metal and the semiconductor.This Gate structure takes advantage of material work function with the extracting the benefit of high k-Gate dielectrics(HfO2) while suppressing the its large voltage shift when layered with SiO_2.The I_d-V_{gs} and I_d-V_{ds} characteristics improves with the Stacked Dual Oxide Gate Structure.**

Keywords—Stacked Dual Oxide Gate structure, MOSFET, TCAD

I. INTRODUCTION

The metal oxide field effect transistor or MOSFET,is the most used semiconductor device today [1].The qualities which pushed the MOSFET to the status of the most promising active component for VLSI are its high yield ,low cost and dense packaging .One of its feature is its isolating propertyso that MOSFET can be placed side by side on the chip without the need for the providing isolation tubes.So it can be placed denser than the bipolar counterpart and require less processing steps[2].The MOSFET is a device used to amplify or switch electronic signal with the fast switching time and important device for VLSI circuit like microprocessor.[3].

The MOS transistor contains two types –the p-channel MOSFET (PMOS) and n-channel(NMOS).Both of this MOS transistor have their own characteristics that differentiate each other. In MOSFET, there are 3 terminal-Source, Drain and Gate.Metal is used to form the Gate terminal .Gate Oxide layer is used for isolation between the metal and the semiconductor so that there can be no current between Gate and Substrate of semiconductor due to tunneling of carrier.High quality electrical insulation provided by SiO_2 is used as Gate-oxide. Besides, SiO_2 is a very good insulator. The resistivity of SiO_2 is greater than 10^{15} (-cm) and large 8-9 eV energy gap.

II. THEORITICAL BACKGROUND

Scaling of MOSFET down sub-100 nm and below leads to hot carrier deterioration. As the devices scaled down,higher Electric field cause the degradation in the electron mobility[4].The presence of large electric field in MOSFET implies the presence of high energy carrier or the hot carrier that might get injected into the surrounding dielectrics film such as the Gate and sidewall oxides[5].The device characteristics drastically changes by the presence of mobile carrier in the oxides when used over prolong period of time.

The analysis of hot carrier reliability in terms of hot-electron-injected gate current .impact-ionization substrate current , electron velocity and temperature near the Drain end for their impact on device speed have been discussed in past [6,7].

III. HIGH-K GATE DIELECTRICS

Scaling down the thickness of SiO_2 leads to high direct tunneling gate leakage current which cause increase in the device power consumption.So a thicker equivalent oxides thickness is induced to reduce the leakage current by introducing the high-kdielectrics to real applications. Hafnium oxideHfO_2 [8], HfO_2-based materials[9] ,zirconium dioxide ZrO_2 and ZrO_2-based material are some high-k dielectrics.HfO_2 and HfO_2 based material are the most promising candidates combining high dielectric permittivity and thermal stability with low leakage current due to reasonably high barrier height that limit electron tunneling[10].The combination SiO_2-HfO_2 as stacked dual oxide Gate structure achieve high drive current performance and lower leakage current providing its usefulness for high performances [11].

Table I. shows the k value and Band gap for HfO_2 and SiO_2.

Material	k-value	Band gap
HfO_2	25	5.8
SiO_2	3.9	9

IV. DEVICE STRUCTURE

This device is like an ordinary n-MOSFET but the oxide structure is replaced by the stacked dual oxide Gate structure .This structure is placed in such a way that HfO_2 is placed over SiO_2.The overall structure is shown in Figure 1 with the Gate Length L(=90 nm) and HfO_2 film thickness (=2 nm) with SiO_2 film thickness (=2 nm).Figure 2 is the detailed view of the stacked Dual Oxide structure.

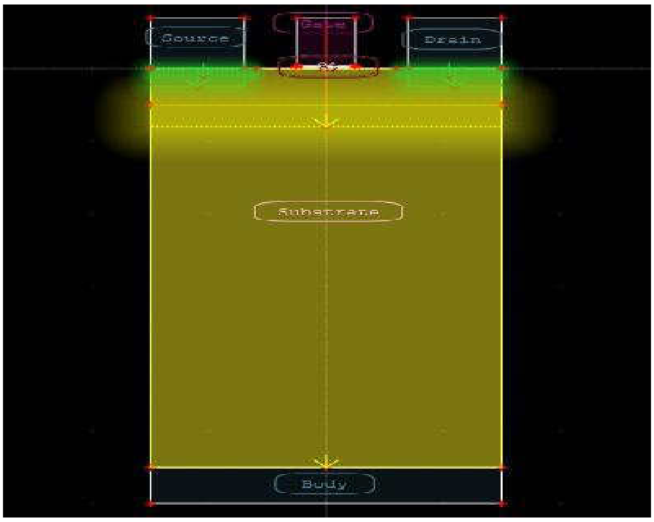


Fig.1.Structure of N-channel MOSFET

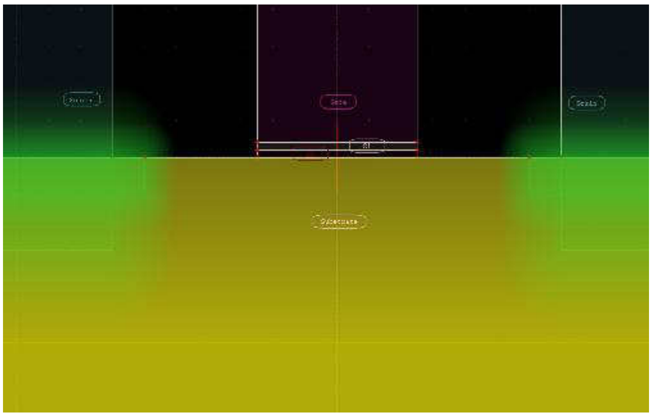


Fig.2 Detailed view of Stacked Dual oxide Gate Structure

In conventional MOSFET SiO_2 is used while in proposed stacked dual oxide Gate structure, SiO_2 is used in combination with high-kdielectic material HfO_2 as Gate insulator.

Table II. Design Parameters for the proposed design

Channel Length	90 nm
Device Width	1
Source/Drain Junction Depth(X_j)	25 nm
Substrate Doping	5 $*10^{16}$ cm^{-3}
Source/Drain Doping	1 $*10^{20}$ cm^{-3}
Permittivity of SiO_2	3.9
Permittivity of HfO_2	25
Effective Oxide Thickness(EOT)	$t_{ox1} + \frac{\varepsilon_{ox1}}{\varepsilon_{ox2}} t_{ox2}$
Substrate Bias (V_{sub})	0 V
Channel Doping	1 $*10^{18}$ cm^{-3}

V. Simulation And Discussion

Device simulation has been performed at 90 nm technology using Cogenda Genius TCAD. Cogenda Genius TCAD simulation tool is used to design the device structure. Several stacked dual oxide Gate structures are stimulated to show the performance between stacked dual oxide gate structure and standard MOSFET.

Table III. Different stacked dual oxide gate structures

Structure No	Device Configuration
Structure 1	MOSFET with SiO_2
Structure 2	MOSFET with SiO_2:HfO_2(1:1)
Structure 3	MOSFET with HfO_2:SiO_2(1:1)
Structure 4	MOSFET with HfO_2

The model used for the simulation isBoltzman for carrier statistics.The temperature of MOSFET for simulation is 300K.

For NMOS stacked dual oxide Gate structure if the small drain V_{ds}>0 is applied a drain current $I_d$will flow which is proportional to the V_{ds} from the drain to source through the conducting channel.This mode is called Linear mode or linear region.The I_d-V_{gs}characteristics for various dual oxide Gate structure is summarized in Figure 3(V_{ds}=1.5 Volts) .

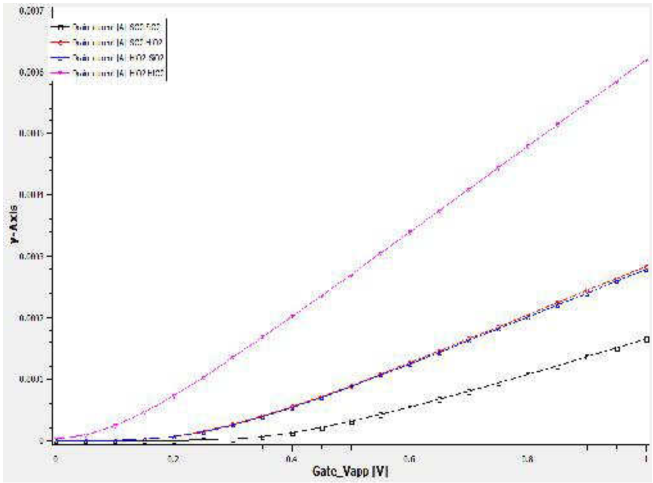


Fig.3. I_d-V_{gs} Characteristics of Different Gate Structures

Table 4 shows the threshold voltages for different structures mentioned in Table 3.and Figure 4 show a graphical view of all the structures for comparison.

Table III. Threshold Voltages of different Gate Structures

Device Configuration	Threshold Voltage
MOSFET with SiO_2	0.35 V
MOSFET with SiO_2:HfO_2(1:1)	0.21 V
MOSFET with HfO_2:SiO_2(1:1)	0.22 V
MOSFET with HfO_2	0.11 V

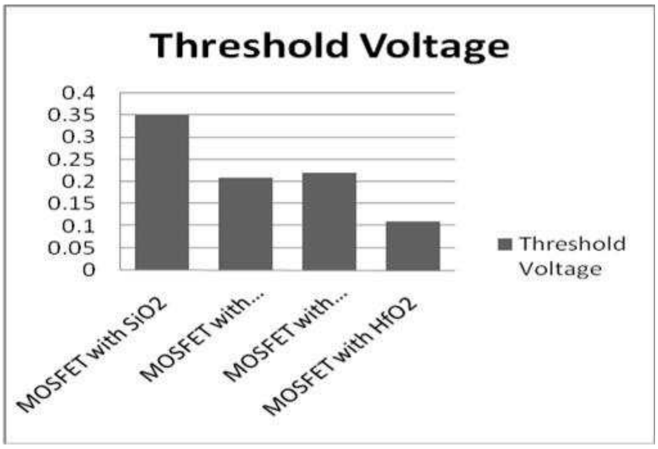


Fig.4. Graph between Different Structures and V_{th}

If the drain voltages is increased,the inversion layer charge and the channel depth at the drain end start to decrease. For $V_{ds}>V_{dsat}$, a depleted region is formed and the depletion region increases towards the source with the increasing drain voltages .This mode is called saturation mode or saturation region.The I_d-V_{ds} output characteristics of different stacked dual oxide gate structures are shown in Figure 5 (Structure 1), Figure 6 (Structure 2) , Figure 7 (Structure 3) and Figure 8 (Structure 4).

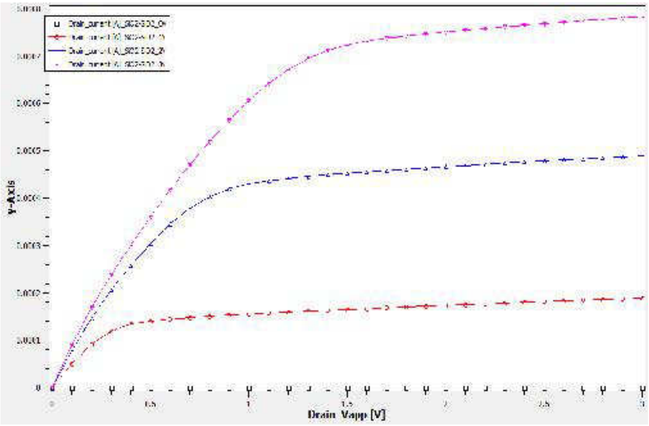


Fig.5. I_d-V_{ds} Characteristics of SiO_2-SiO_2 structure

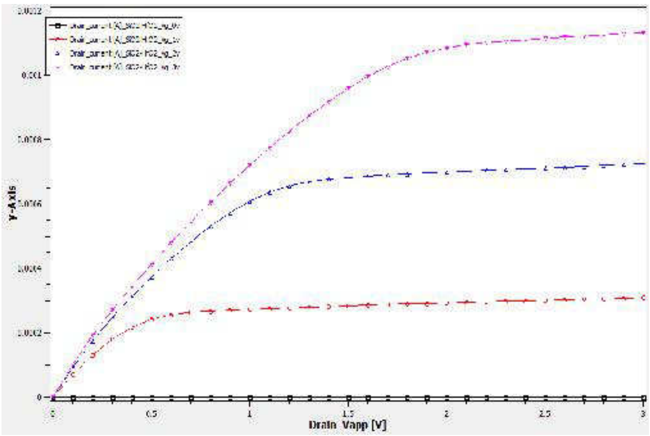


Fig.6. I_d-V_{ds} Characteristics of SiO_2-HfO_2 structure

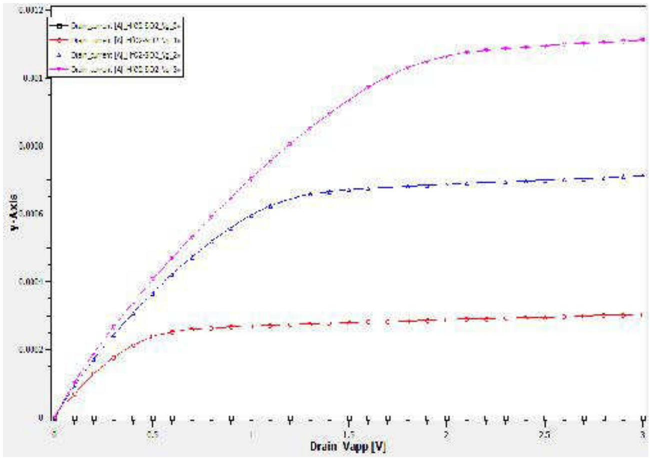


Fig.7. I_d-V_{ds} Characteristics of HfO_2-SiO_2 structure

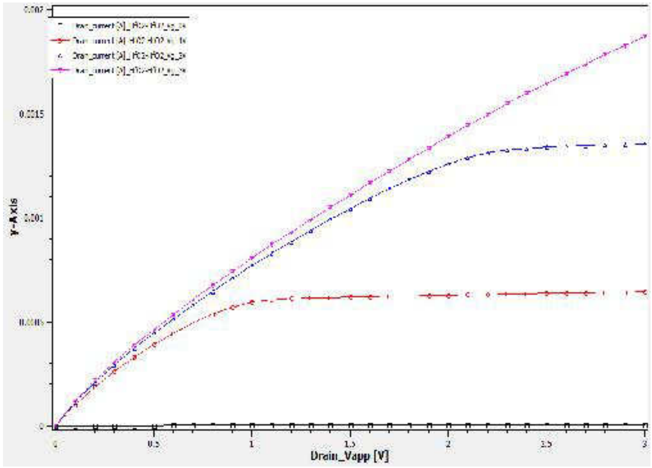


Fig.8. I_d-V_{ds} Characteristics of HfO_2-HfO_2 structure

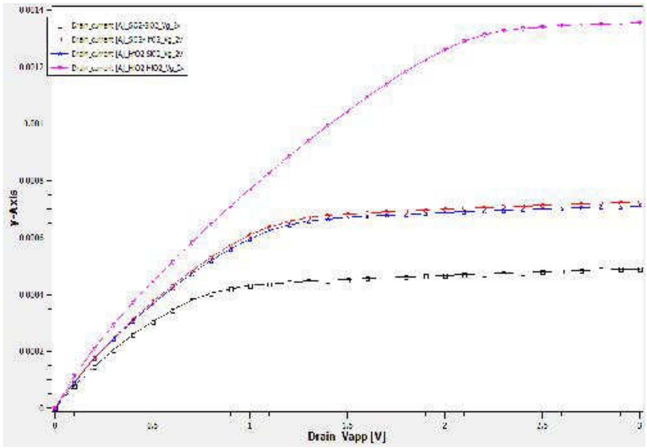


Fig.9. Comparison of I_d-V_{ds}characteristics of structures at V_{gs}=1 V

High gate to channel coupling produces more drain current than the conventional MOSFET (Structure 1).The relative comparison of all the structures can be studied in a condensed Figure 9 at V_{gs}=1 V.

VI. CONCLUSION

Simulation is performed for various gate structures where is observed that the structure using HfO2 or the high-k dielectrics provide better transconductance and so, is high drain current.There is also decrease in the threshold voltage in the stacked dual oxide gate structures than the conventional model of MOSFET.

REFERENCES

[1] B.L Anderson, R.L Anderson, "Fundamental of Semiconductor Devices" McGraw-Hill Education,2nd Edition ,2005, pp 373-375.

[2] Neamen,Donald A., "Semiconductor Physics and Devices: Basics Principles",University of New Mexico,McGraw Hill Higher Education,3rd Edition,2002, pp-1-18,367,449-485.

[3] Singh J. ,"Solid State Electronic Devices", McGraw Hill Higher Education,2nd Edition,2001, pp-370-371.

[4] H.Krautschneider, A. Kohlhase, and H. Terletzki,"Scaling down and reliability problems of gigabit CMOS circuits," Microelectron. Reliab.,vol. 37, no. 1, pp. 19–37, Jan. 1997.

[5] G. V. Groeseneken, "Hot carrier degradation and ESD in submicrometer CMOS technologies: How do they interact?" IEEE Trans. Device Mater. Rel., vol. 1, no. 1, pp. 23–32, Mar. 2001.

[6] C. H. Choi, K. Y. Nam, Z. Yu, and R. W. Dutton, "Impact of gate direct tunneling current on circuit performance: A simulation study," IEEE Trans. Electron Devices, vol. 48, no. 12, pp. 2823–2829, Dec. 2001.

[7] J. W. Han, J. Lee, D. Park, and Y. K. Choi, "Body thickness dependence of impact ionization in a multiple-gate FinFET",IEEE Electron Device Letters., vol. 28, no. 7, pp. 625–627, Jul. 2007.

[8] Gutowski M, Jaffe JE, Liu CL, Stoker M, Hegde RI, Rai RS, Tobin PJ: Thermodynamic stability of high-k dielectric metal oxides ZrO2 and HfO2 in contact with Si and SiO2. MRS Proceedings 2002.

[9] Dimoulas A, Vellianitis G, Mavrou G, Apostolopoulos G, Travlos A, WiemerC,Fanciulli M, Rittersma ZM: La2Hf2O7 high-k gate dielectric grown directly on Si (001) by molecular-beam epitaxy. ApplPhysLett 2004,15(85):3205–3207

[10] Zhao et al :Dielectric relaxation of high k-oxides Nanoscale Research Letters 2013 8:456.

[11] Chaujaret al.: TCAD Assessment Of Gewe-Rc MOSFET-Part I, 2602 IEEE Transactions On Electron Devices, Vol. 55, No. 10,pp-2602-2604, October 2008.

Effect of Substrate Variation on NMOS: Simulation using TCAD

Shivendra
Dept. of Electronics
NITTTR, Chandigarh, India

Rajesh Mehra
Dept. of Electronics
NITTTR, Chandigarh, India

***Abstract*— This paper uses COGENDA VISUAL TCAD simulation of a 90 nm N-channel MOSFET technology for fabrication of N-channel MOSFET using different type of substrate material. The effect of different type material used in fabrication of MOSFET on the threshold voltage and subthreshold current. In I_D-V_{DS} characteristic shows that the MOSFET operate in various region which are cut-off, triode, and saturation region .we can find out the value of subthreshold swing in cut-off region.The doping profile for all type substrate are at same concentration.**

***Keywords*— COGENDA VISUAL TCAD simulation, substrate material, threshold voltage, subthreshold current, doping profile.**

I. INTRODUCTION

According to the law given by Gorden E. Moore "the number of transistor that can be placed on an integrated circuit, doubles in approximately every two year". This trend is continuously going on. Now at present day The metal- oxide-semiconductor field–effect transistor or known as MOSFET is the most widely used in VLSI.The MOSFET is a device used as amplifier and as switching the electronic signals with fast response time.

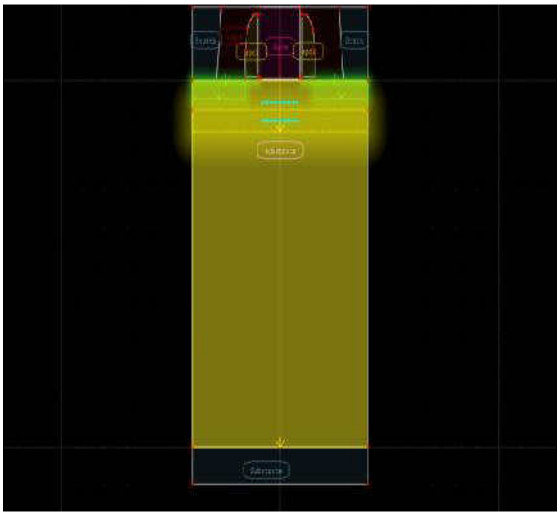


Fig. 1 Structure of Fabricated N MOS

The MOSFET is a four terminal device which are source(s), drain (D), gate (G) and body (B). The top layer of MOSFET is metal (polysilicon) and it is used to form gate electrode. For isolation between metal and substrate gate oxide(SiO_2) is used. So that ideally there is no current will flow between gate and substrate.SiO2 have very good insulating property so that silicon dioxide SiO2 has been used as a gate oxide in the MOSFET. SiO2 have high quality electrical insulator with resistivity of 10^15 ohm.

In MOSFET drain and source is heavily doped semiconductor where substrate are lightly doped semiconductor. In MOSFET there will be two p-n junctions which are between drain-substrate and source-substrate. Due to these junctions and tunneling effect between gate and substrate the subthreshold current or a leakage will flow.

II. THEORETICAL BACKGROUND

SUBTHRESHOLD CURRENT: When the voltage V_{gs} is less than threshold voltage, a weak inversion layer or channel is formed. Diffusion current is dominant current mechanism over drift current mechanism. Subthreshold current depend on doping concentration, mobility of material, applied voltage and temperature [1].

$$I_D = JdA = -qADn\frac{\partial n}{\partial y} \approx -qADn\frac{n(0) - n(L)}{L} \quad (1)$$

$$n(0) = n_i\, e^{q(\varphi_s-\varphi_b)/kT} \quad (2)$$

$$n(L) = n_i\, e^{q(\varphi_s-\varphi_b-V_D)/kT} \quad (3)$$

$$I_D = \frac{qADnni\, e^{-\varphi_B/kT}}{L}(1-e^{-qV_D/kT})e^{q\varphi_s/kT} \quad (4)$$

$$I_D \approx \frac{qADnni\, e^{-\varphi b/kT}}{L}(1-e^{-qV_D/kT})e^{q(V_G-V_T)/kT} \quad (5)$$

where φ_s=Energy level of substrate in electron volt

n_i=Intrinsic carrier concentration of semiconductor

φ_b=Energy level of body in electron volt

V_T= Threshold voltage MOSFET

K=Boltzman's constant

V_D=Voltage applied at drain terminal of MOSFET.

Subthreshold current is exponential function of applied gate voltage. Subthreshold current gets larger for smaller gates length. When the V_{DS} increase the leakage current or subthreshold current get increase [2, 6]. Subthreshold current is temperature dependent. When the temperature increase the subthreshold current get increase.

THRESHOLD VOLTAGE: If work function differences and oxide charges are present between gate and substrate, threshold voltage is shifted just like for MOS capacitor. Minimum voltage required to apply at the gate to form the inversion layer or channel is called threshold voltage. Threshold voltage depends on doping concentration of substrate, temperature, thickness of oxide, used material in substrate [1, 3].

$$V_T = V_{FB} + 2\varphi_B + \frac{\sqrt{2\varepsilon_s q N_A (2\varphi_B)}}{C_i} \quad (6)$$

If the substrate is biased with respect to the source V_{BS} the threshold voltage is also shifted.

$$\Delta V_T = V_T(V_{FB}) + 2\varphi_B + \frac{\sqrt{2\varepsilon_s q N_A (2\varphi_B + V_{BS})}}{C_i} \quad (7)$$

$$\Delta V_T = V_T(V_{BS}) - V_T(V_{BS}=0)$$

$$\Delta V_T = \sqrt{\frac{2qN_A\varepsilon_s}{C_i}}\left(\sqrt{2\varphi_B + V_{BS}} - \sqrt{2\varphi_B}\right) \quad (8)$$

Where N_A=acceptor ion concentration

ε_s=permittivity of material

C_i=Intrinsic capacitance

III. N-MOS SIMULATION

Simulation of MOSFET performance has been done since three decade ago. A two-dimensional simulation of MOSFET with 90 nm technology shows that the two-dimensional energy distribution was stably calculated using several mechanisms. This simulation is also helpful in analysis of device at an electron level. Based on the two dimensional technique, the threshold voltage, gate current, subthreshold current, drain current can be calculated at different drain to source (V_{DS}).

Table1: Doping profile and concentration

Name	Profile	Type	Peak conc./cm^3	Char. length/μm
Substrate	Uniform	Acceptor	5×10^{16}	-
Drain/source	Gaussian	Donor	1×10^{20}	0.02
P+	Gaussian	Acceptor	1×10^{18}	0.05

All TCAD simulations in this paper have been done with COGENDA VISUAL TCAD software. Four substrate are used to fabricate the MOSFET, these are Si, Ge, SiGe and SiC3C. All material has different Energy bandgap and mobility.

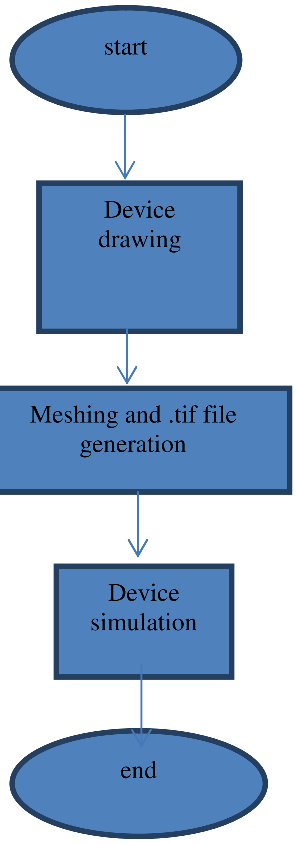


Fig. 2 Flow Chart of NMOS Design and Simulation

MOSFET operate in such a way that the current from the source to drain is carried by majority carriers ,by electron in NMOS and by holes in PMOS.Figure 1.1 shows a schematic of an NMOS device with p-type substrate. A potential is applied to the gate that produce inversion layer near-semiconductor-interface. Substrate material is most critical parameter to defined threshold voltage and subthreshold current [1].

Table 2 : material used in N-channel MOSFET

Region	Material	Mesh size(μm)
Source	Al	0.01
Drain	Al	0.01
Body	si	0.05
Gate	NPolySi	0.01
SUB	Al	0.025

IV. RESULT ANALYSIS

Fig. 3 shows the I_D-V_{GS} characteristics of the proposed MOSFET. In the figure, MOSFET with substrate material SiC3C have 1 V threshold voltage and MOSFET with substrate material Si and SiGe have 0.45 V threshold voltages [3, 4].

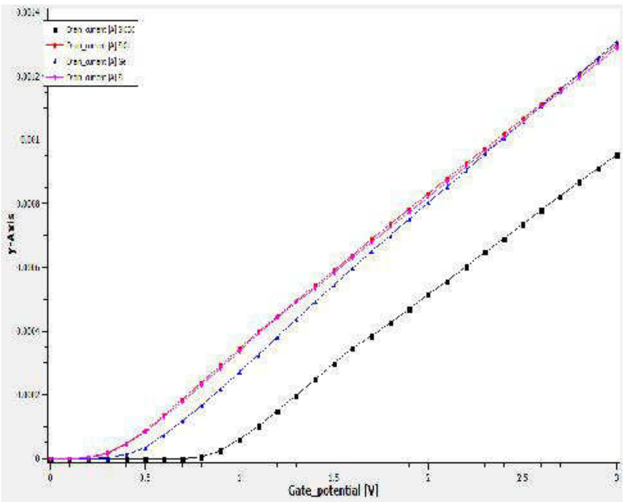


Fig. 3 threshold voltage of MOSFET with different substrate material

Fig. 4 shows the *leakage* current of the proposed MOSFET. . In the figure ,MOSFET with substrate material SiC3C have 2×10^{-16} A subthreshold current and MOSFET with substrate material Ge and SiGe have 2×10^{-8} A subthreshold current[2,6].

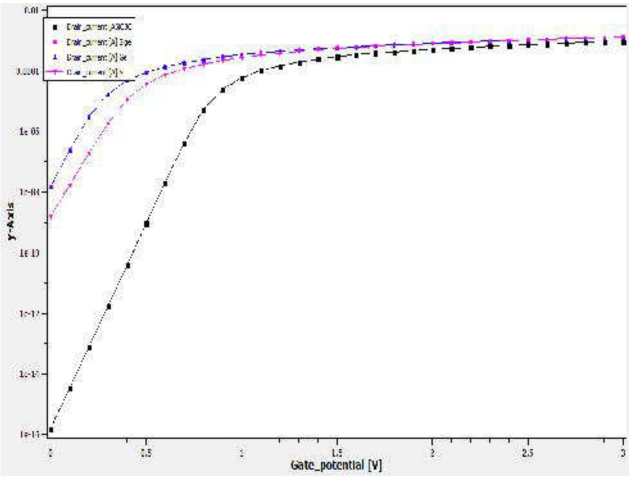


Fig. 4 Subthreshold current of MOSFET with different substrate Material.

Fig. 5 shows the I_D-V_{GS} characteristics of the proposed MOSFET.In the figure, MOSFET with substrate material SiC3C have minimum current in saturation and MOSFET with substrate material Ge and SiGe have same current in saturation [3,4].

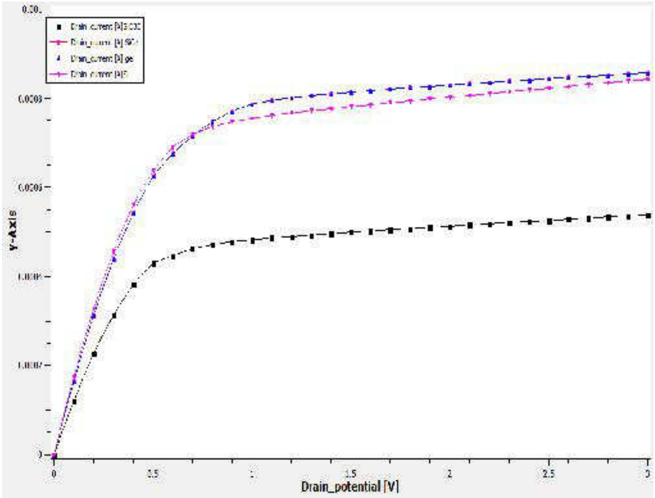


Fig. 5 I_D-V_{DS} characteristics of N MOS with different substrate material

V. CONCLUSION

The proposed method of varying the substrate material gives improved performance in terms of subthreshold current, threshold voltage, and I_D-V_{DS}. The results clearly show that there is a considerably large variation in the sub-threshold current, and threshold voltage. The reduced subthreshold current, and threshold voltage assures a reduction in the static power dissipation i.e. the device will dissipate lesser power. This will clearly improve the quality of the device in terms of cost and power consumption. The idea behind the proposed concept is successfully implemented in the simulation done using TCAD.

REFERENCE

[1] MIT Open Courseware 6.012 Microelectronic Devices and Circuits Fall 2009.

[2] K. Keerti Kumar and N. Bheema Rao, “Variable Gate Oxide Thickness MOSFET: A Device level solution for Sub-threshold leakage current reduction”, in Proc. ICDCS, 2012, pp. 495-498.

[3] S. Borkar, T. Kamik, S. Narendra, J. Tschanz, A. Keshavarzi, V. De, "Parameter Variations and Impact on circuits and micro architecture", DAC, 2003, pp.338-342.

[4] Z.H. Liu, C. Hu, J-H. Huang, T-Y. Chan, M-C. Jeng, P.K. Ko, Y.C. Cheng, "Threshold Voltage Model for Deep-Submicrometer MOSFET's," IEEE Trans. On Electron Devices, Vol. 40, No. 1, January 1993, pp. 86-95.

[5] C.H. Wann, K. Noda, T. Tanaka, M. Yoshida, C. Hu, "A Comparative Study of Advanced MOSFET Concepts," IEEE Transactions on Electron Devices, Vol. 43, No. 10, October 1996, pp. 1742-1753.

[6] Yee-Chia Yeo; Tsu-Jae King; Chenming Hu, “MOSFET gate leakage modeling and selection guide for alternative gate dielectrics based on leakage considerations,” IEEE Transactions on Electron Devices, Vol. 50, No. 4, April 2003, pp. 1027-1035.

Single Electron Transistor & Their Applications

Hitesh Arora
Dept. of ECE
CTIEMT, Shahpur
Jalandhar, India
hiteshh294@yahoo.com

Diksha
Dept. of ECE
CTIEMT, Shahpur
Jalandhar, India
dkaushal1105@gmail.com

Ravinder Kaur
Dept. of ECE
CTIEMT, Shahpur
Jalandhar, India
er.rkaur@gmail.com

Navneet Gill
Dept. of ECE
CTIEMT, Shahpur
Jalandhar, India
pallnavneet5@gmail.com

***Abstract*—In the work done, the challenges being faced by single electron transistor are analyzed. Different applications are considered to pattern the short coming and advantages of the SET (Single electron transistor).It has low voltage gain, high input impedance and sensitive to random background charges. The recent research in the field has generated some byproducts ideas which can revolutionize Random access memory and digital data storage technologies.**

***Keywords*—Quantum dots (QDs), Single electron transistor (SET), Quantum cellular automata (QCA).**

I. Introduction to SET

A single-electron transistor [1] is a capable contender for ultralow-power and ultrahigh-density circuit systems in the coming future. In addition, exceptional features of SETs have been applied to many functional logic circuits [2]–[4]. The single electron transistor is a three terminal device. Its electric characteristic is that it is governed by the movement of single electrons. It is a type of a switching device that it uses controlled electron tunneling to amplify current. In this the effect of coulomb blockage can be observed. It consists of two electrodes known as the drain and the source, connected through tunnel junctions to one common electrode with a low self-capacitance, known as the island,as shown in fig1.

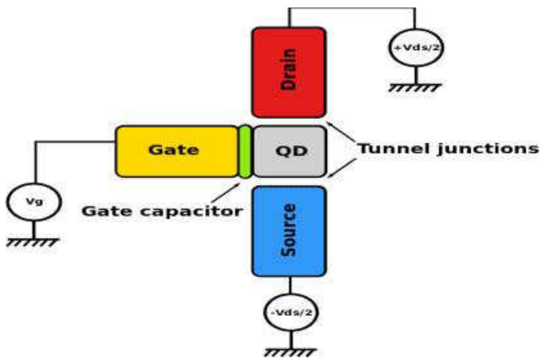


Fig. 1 Single Electron Transistor

The electrical potential of the island can be tuned by a third electrode, known as the gate, capacitively coupled to the island. In the blocking state no accessible energy levels are within tunneling range of the electron on the source contact. All energy levels on the island electrode with lower energies are occupied. When a positive voltage is applied to the gate electrode the energy levels of the island electrode are lowered. The electron can tunnel onto the island, occupying a previous vacant energy level. Then it can tunnel onto the drain electrode where it in elastically scatters and it reaches the drain electrode Fermi level.

The tunnel junction consists of two metal pieces separated by a very thin (~1nm) insulator. The only way for electrons in one of the metal electrodes to travel to the other electrode is to tunnel through the insulator. Since tunneling is a discrete process, the electric charge that flows through the tunnel junction flows in multiples of the charge of electrons e.

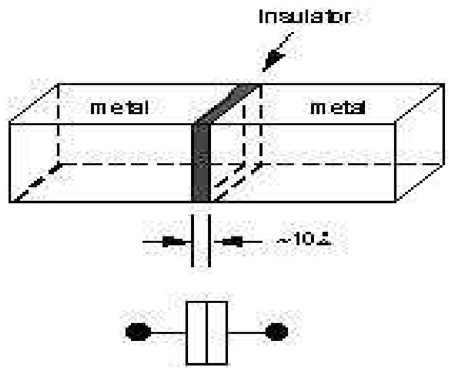


Fig. 2 Tunnel Junction

II. APPLICATIONS OF SET

Different applications considered to pattern the short coming and advantages of the SET are as follows:

A. Supersensitive electrometeran analog application

The high sensitivity of single-electron transistors supports them to act as electrometers in unique physical experiments. For example, they have made possible clear-cut observations of the parity effects in superconductors. Absolute measurements of extremely low dc currents (~10-20 A) have been proved. The transistors have also been used in the first measurements of single-electron effects in single-electron boxes and traps a improved version of the transistor has been used for the first proof of the existence of fractional-charge excitations in the fractional quantum hall effect. The supersensitive electrometer measures voltages or charge with a great accuracy when compared to electrometer [5]. This extremely high sensitivity has been recommended [6] as a basis for supersensitive electrometry.

B. *Charge sensor*

The single electron transistors are effective charge sensors for measuring out spin or charge qubits kept in quantum dots (QDs). To observe their capacitive parameters, which are associated to the signal to noise ratio (SNR) during qubit measurement, twin silicon single QDs were fabricated by means of a lithographic process on silicon on insulator substrate. As configuration and dimension of QDs could be determined by direct imaging, the theoretical capacitive parameters could be related to the measured values. Good agreement was found between the calculated and measured values, which authorize the validity of the calculation method. The results indicating that reducing the SET diameter decreases the capacitive coupling between qubits but raises the signal to noise ratio for both dc and radio frequency single shot measurements. Since these results are independent of the device materials, they are advantageous for establishing guidelines for the design of SET charge sensors in lateral QD-SET structures based on a two dimensional electron gas.

C. *Detection of Infrared Radiation*

The single-electron transistor can as well be used to detect infrared signals at room temperature. By exciting electrons over an electrically induced energy barrier, both the range of measurable wavelengths and the sensitivity of the device can be controlled. The sensor start working when an infrared signal excites conduction-band electrons in a 25-nm deep electron reservoir. A silicon insulator channel measuring 40 × 400 nm is placed next to the reservoir to intensify the number of excited electrons. A poly-silicon lower gate then turns off the transistor and electrically forms an energy barrier, generating a storage node on the other side. Electrons with energy greater than the height of the barrier are injected into the storage node, where they are read as changes in current flowing through the transistor.

D. *Ultrasensitive Microwave Detector*

A further application of Single Electron Transistor is an Ultrasensitive Microwave Detector; through two small capacitance tunnel junctions and a capacitive gate island is feebly coupled to a bias circuit at low bias voltages and temperatures, a single quasi particle can only be introduced to the island through photon-assisted tunneling. Once this occurs, the quasi particle is trapped on the island as it takes a relatively long time for this specific quasi particle to tunnel off. Meanwhile it is trapped, charge is transported through the system two electrons at a time. Since the photon-assisted transition merely switches the detector current on, this device is not limited to one electron tunneled through the system per absorbed photon. Due to this the device tremendously sensitive and potentially beneficial detector of microwave radiation.

E. *Temperature Standards*

Earlier theoretical analysis based on the conventional theory has shown that V= 5.44NkBT/e is amazingly stable with respect to almost any variations of the array parameters (with the important exception of a substantial spread in the junctions resistances), on condition that a remarkable opportunity to use the arrays for absolute thermometry, subsequently the fundamental constants are known with high accuracy. Each particular array may give high (1%) accuracy of within less than one decade of temperature variations, but for arrays with different island size (and hence different), these ranges can be shifted and overlap. Therefore, it is possible to have an absolute standard of temperature with a very expansive (say, two-decade) total range from several circuits fabricated on a single chip. This advancement is very encouraging, but since all this work is recent, some time is needed to see whether these new devices will be able to strive with the conventional temperature standards.

F. *Single-Electron Spectroscopy*

A new application of single-electron electrometry is the probability of measuring the electron addition energies (and hence the energy level distribution) in quantum dots and other nanoscale entities. There are two natural methods to carry out such measurements. The first is to use the quantum dot as the island of the single-electron box, capacitively coupled to the single electron transistor or other sensitive electrometer. The succeeding is to use the quantum dot directly as the island of a weakly biased single-electron transistor and measure the gate voltages providing the sharp increase of the source-drain conductance.

G. *DC current standards*

One of the possible applications of single-electron tunneling is fundamental standards of dc current intended for such a standard a phase lock SET oscillations or Bloch oscillations [7] in a simple oscillator with an external RF source of a fine characterized frequency f. The phase locking would offer the transfer of a certain number m of electrons per period of external RF signal and thus produce dc current which is fundamentally related to frequency as I= mef. This arrangement have drawback of coherent oscillation that are well ahead overcome by the use of such a constant RF source to drive devices such as single-electron turnstiles and pumps , which do not exhibit coherent oscillations in the autonomous mode.

H. *Prospect for Digital Applications*

The origination of single electronics as an applied discipline [7], [6], [8], the most exciting question was whether single-electron devices will be able to compete with semiconductor transistor circuits in mainstream, digital electronics. Several paths toward this goal have been explored.

I. *Voltage State Logics*

The single-electron transistors can also be used in the "voltage state" mode. Here in this mode, the input gate voltage U controls the source-drain current of the transistor which is used in digital logic circuits, on the contrary to the usual field-effect transistors (FETs). Hence it means that the single-electron charging effects are restricted to the interior of the transistor, while externally its appearance is like the usual electronic device switching multi-electron currents, with binary unity/zero presented with high/low dc voltage levels (physically not quantized). One substantial shortcoming of voltage state circuits is that neither of the transistors in each complementary pair is closed too well, so that the static leakage current in these circuits is fairly substantial, of the order of 10-4 e/RC.

III. Limitations in SET Implementations

A. Back Ground Charge

The major limitation with the single electron logic circuits is the randomness of the background charge. A single charged impurity entombed in the insulating environment polarizes the island, building on its surface an image charge Q0 of the order of electron e. This charge is effectively subtracted from the external charge Qe.

B. Room Temperature

The another problem with all the kinds of single electron logic devices is the constraint Ec~100kB T, which in run through means sub-nanometer island size for room temperature operation. In such a small conductors the quantum kinetic energy gives a dominant input to the electron additional energy even small variations in island shape will lead to uncertain and relatively significant variations in the spectrum of energy levels and therefore in the device switching threshold.

C. Out Side Environment Linking with SETs

The separate structures patterns which work as logic circuits must be organized in to larger 2D patterns. There are two notions, first is to integrate SET as well as related equipment's with the prevailing MOSFET, and this is attractive because it can upturn the integrating density. The next option is to give up linking by wire, instead utilizing the static electronic force between the basic clusters to form a circuit linked by cluster, which is called quantum cellular automata (QCA). The benefit of QCA is its first information transfer velocity between cells via electrostatic dealings only, no wire is required between arrays and the size of each cell can be as small as 2.5 nm, this made them very suitable for high density memory and next generation quantum computer.

D. Lithography Technique

Another drawback with single electron devices is the requirement Ec~100kBT, which in actually means sub-nanometer island size for room temperature operation. In VLSI circuits, this fabrication technology level is very high. Furthermore, even if these islands are fabricated by any sort of nanolithography, their shape will hardly be absolutely regular.

E. Co-tunneling

The pressure essence of the effect is that the tunneling of several electrons through diverse barriers at the identical time is possible as a single coherent quantam mechanical process. The rate of the process is crudely less than that for the single electron tunneling.

IV. Conclusion

This research paper emphases on the basic principle of single electron transistor its applications and restrictions with importance of single electron transistor in the era of nanotechnology to provide low power consumption and a high operating speed in the field of digital design for the fabrication of various electronic devices. SET has evidenced its value as a significant tool in scientific research.

References

[1] H. Grabert and M. Devoret, Single Charge Tunneling. New York, NY:Plenum, 1992. http://perso.neel.cnrs.fr/vincent.bouchiat/page_personnelle_de_Vincent_Bouchiat/course_papers_files/devoret.pdf

[2] J. R. Tucker, "Complementary digital logic based on the Coulomb blockade," J. Appl. Phys., vol. 72, pp. 4399–4413, Nov. 1992.

[3] Y. Takahashi, A. Fujiwara, K. Yamazaki, H. Namatsu, K. Kurihara, and K. Murase, "Multigate single-electron transistors and their application to an exclusive-OR gate," Appl. Phys. Lett., vol. 76, pp. 637–639, Jan.2000.

[4] H. Inokawa, A. Fujiwara, and Y. Takahashi, "Multipeak negative-differential- resistance device by combining single-electron and metal-oxide semiconductor transistors," Appl. Phys. Lett., vol. 79, pp. 3618–3620, Nov. 2001.

[5] J. Doe, M. S. Douglas and R. Hill, J. Electrochem. Soc., 137 (1990) 1902.

[6] K. K. Likharev, "Single-electron transistors: Electrostatic analogs of the DC SQUID's," IEEE Trans. Magnetics, vol. 23, pp. 1142–1145, Mar. 1987

[7] C. Wasshuber and H. Kosina, "A single-electron device and circuit simulator," Superlattices and Microstructures, vol. 21, pp. 37–42, Jan. 1997.

[8] K. K. Likharev, Dynamics of Josephson Junctions and Circuits. New York: Gordon and Breach, 1986.

Speed Control of HEV with PID and PD controller

Avtar Singh
Dept. of Electronics and Communication Engineering
Punjabi University
Patiala, India
avtar2030@gmail.com

Amrit Kaur
Dept. of Electronics and Communication Engineering
Punjabi University
Patiala, India
amrit_aks@yahoo.ca

***Abstract*— Hybrid electric vehicle (HEV) is getting more attention recently because it has no emission of toxic gases which leads to global warming. The main aim of this paper is to control the speed of non-linear hybrid electric vehicle by controlling the throttle position so as to get driving safety, improved fuel economy, reduced manufacturing cost and pollution. To control the speed of hybrid electric vehicle PD and PID controller is used. The tuning of these two controllers is done using MATLAB PID tuning toolbox. Results are compared of PID controller with PD controller.**

Keywords—Hybrid Electric Vehicle (HEV),PID, PD.

I. INTRODUCTION

A major source of green house effect is increasing number of vehicles. Scientists are finding ways to reduce it. So, in recent times hybrid electric vehicles are getting more attention because it has no emission which will lead to global warming. Moreover the sources used for generation of electric power battery of electric vehicles are non conventional source which are environment friendly. The HEVs uses both electric machine and IC engine for delivering the propulsion power [1][2]

The oppugns on which research work is still going on to achieve high performance efficiency, ruggedness, compact sizes, and less costs in power converters and electric machines, as well as in associated electronics [3].

This capability is mainly due to:

1. The possibility of downscale of engine.
2. The energy recovery ability of rechargeable.
3. Storage system during breaking phases.
4. The power can be split into thermal and electrical paths, so it enables to have an additional degree of freedom.

Also, as summarized below HEVs have more benefits than conventional ICEVs [4],

1. The torque generation in electric vehicles is very much prompt and veracious, so crash can be easily averted by decreasing the torque.
2. The output torque of these motors can be easily estimated.
3. A motor can be inclined to each wheel of electric vehicle.
4. Give high efficacy over broad torque and speed pasture
5. High sureness and robustness for vehicular environment.

The pioneering authors working on the intelligent energy management in HEVs have proposed compendious category and anatomize of the state of art control design for the same. The performance of HEVs merely depends upon the applied automation system. Many controller such as LQR controller, state feedback controller (SFC), Observer based controller have been designed [5][6]

The speed of hybrid electric vehicle can be controlled by controlling the servo motor which in turn is controlling the throttle position to have smooth torque.

In this paper, we propose control the speed of hybrid electric vehicles with conventional PID controller and PD controller, tuning of PID and PD controller is done with MATLAB toolbox PID tuner.

This paper is organized as follows: in section II detail of the HEV system is described. Section III presents summary of the stability analysis of the control system. In section IV the control strategy used is summarized, in section V results are shown and hence conclusion is drawn in section VI.

II. MODEL DESCRIPTION OF VEHICLE

A schematic diagram of electronic throttle control using a DC servo is shown in Fig. 1. Simulink model of system is shown in fig. 2.

The dynamics of the vehicle is given as [5],[8]

$$m\frac{dv}{dx} = F_e(\theta) - \alpha v^2 - F_g \tag{1}$$

$$\tau_f \frac{dF_e(\theta)}{dx} = -F_e(\theta) + F_{e1}(\theta) \tag{2}$$

$$F_{e1}(\theta) = F_1 + \gamma\sqrt{\theta} \tag{3}$$

F_e = Engine force, a function of throttle position

F_g = Gravity induced force, a function of the road grade.

θ =Throttle Position, v=vehicle speed

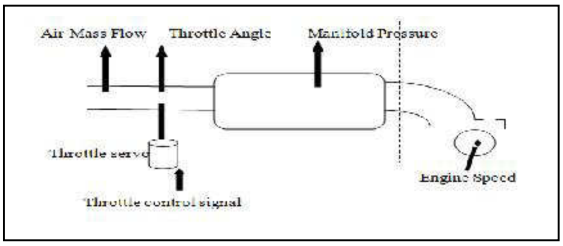


Fig. 1 Schematic diagram of electronic throttle control

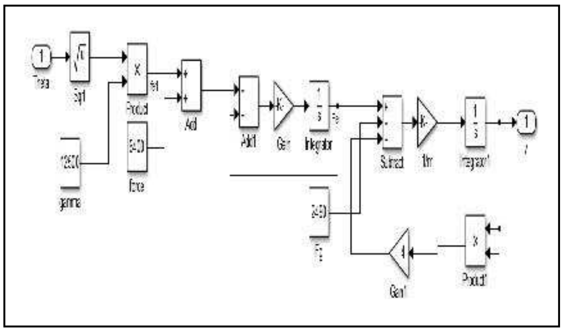


Fig. 2 Simulink model of vehicle.

- Assumptions

1. Gravity induced force (F_g) is 30% of weight of vehicle.
2. Engine time constant (τ_f) is taken as 0.2 that is in the range from 0.1 to1.0 sec.

The values of remaining parameters are given in Table 1 [5]

Table I. Numerical Values of Parameter

Constants	Notation	Value(SI unit)
Mass (m)	M	1000 Kg
Aerodynamic Drag Coefficient	α	4 N/(m/s)2
Engine Force Coefficient	γ	12500 N
Engine Idle Force	F_1	6400 N

By using given system dynamics and hence state space equation transfer function is obtained [5]

Transfer function

$$\frac{V(s)}{\theta(s)} = \frac{829000}{s^2+5s}$$

III. STABILITY ANALYSIS

The state space variable is given as [4]

$$A=\begin{bmatrix} 0 & 0.001 \\ 0 & 5 \end{bmatrix}, \quad B=\begin{bmatrix} 0 \\ 829000 \end{bmatrix}, \quad C=[1 \quad 0]\ , \quad D=[0]$$

The characteristic equation is given as

$$|\lambda I - A| = 0$$

When we get the Eigen values as λ_1 =0, λ_2 =-5. Then given system is controllable and observable. As the order of matrix M and N is equal to rank of the matrix =1;

$$M=829000\begin{bmatrix} 0 & 0.001 \\ 0 & -5 \end{bmatrix}, \quad N=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

IV. CONTROL STRATEGY

A. PID Controller

Proportional – integral- derivative control is a feedback control system having three parameters k_p, k_i and k_d. All these parameter have their own advantage. So that PID controller add all the advantages of the parameter such faster response time due to P-only control, decreased/servo offset due to I- only control and prediction of disturbance to the system by determining the change in error due to D- only control. The transfer function of PID is given as [9].

$$C(S)= K_p + \frac{K_i}{s} + k_d s = K_p\left(1 + \frac{1}{T_i s} + T_d s\right) \tag{4}$$

Where k_p =proportional Gain, k_i =Integral Gain, k_d =Derivative Gain, Ti=Reset Time= k_p / k_d, Td=Rate Time or derivative time= k_d / k_p

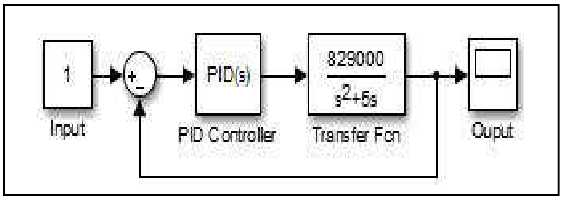


Fig. 3. Simulink Model Of plant with PID controller

B. PD Controller

Proportional – derivative control is a feedback control system having two parameters k_p and k_d. These two parameters have their own advantage. So that PD controller adds all the advantages of the parameter such faster response time due to P-only control and prediction of disturbance to the system by determining the change in error due to D- only control. The transfer function of PD is given as [9].

$$C(S) = K_p + k_d s \tag{5}$$

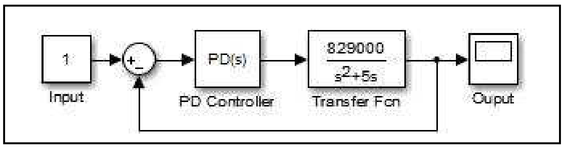


Fig. 4. Simulink Model Of plant with PD controller

V. RESULTS AND DISCUSSION

In this section, results are shown of open loop system without any controller and closed loop system with PID controller and closed loop response with PD controller.

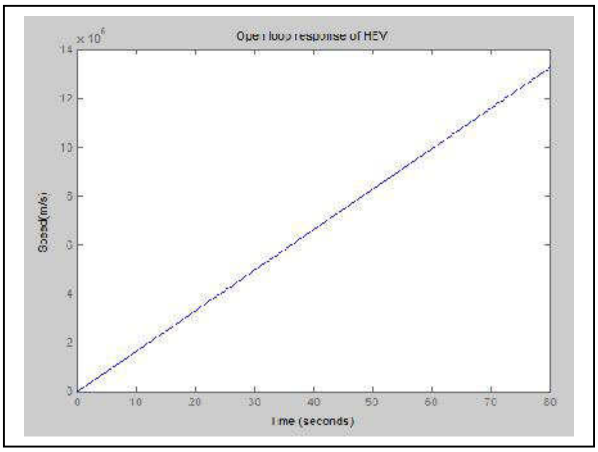


Fig 5. Open loop response of HEV

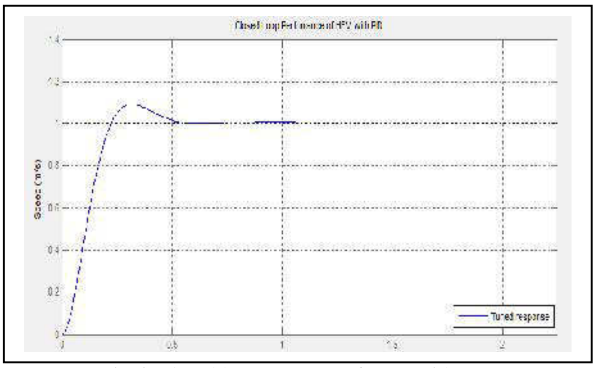


Fig 6. Closed loop response of HEV with PID

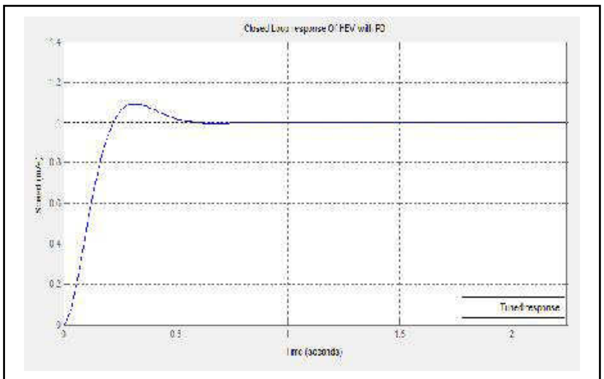


Fig. 7 Closed loop response of HEV with PD

Table II Performance Index of Various Controllers for HEV

Controller	% max overshoot	Settling Time	Rise Time
HEV-PID	9.08	0.486	0.149
HEV-PD	9.21	0.497	0.148

The tuning of PID controller is done with MATLAB simulink model with PID tool is inbuilt in MATLAB.

VI. CONCLUSION

A different kinds of controller is used with HEV for controlling the speed of hybrid electric vehicle. As shown in table II HEV-PID gives better results as compared to the HEV-PD. Maximum overshoot and settling time is less to achieve the desired speed so that current and torque will also be optimized and also its battery operation. Vehicle drive train efficacy may be improved and fuel efficacy of electric vehicles may also be optimized.

REFERENCES

[1] Husain I., 'Electric and Hybrid Vehicles', Design Fundamentals, 2nd edition, CRC Press, Taylor and Francis Group, New York, Chap 1, 2011

[2] Chau K.T., Chan C.C. and Liu C., "Overview of Permanent magnet brushless drives for electric and And hybrid electric vehicles," IEEE Trans. Ind. Electron ,Vol. 5, No. 6, pp. 2246-2257, jun. 2008.

[3] Emadi Ali, Joo Lee Young, Kaushak Rajashekare, "Power Electronics and Motor Drives in electric, Hybrid Electric, and lug-in hybrid Electric Vehicles", IEEE Transactions on Industrial Electronics, Vol. 5, No. 6, pp. 2237-2245, 2008.

[4] Hori Yoichi, "Motion Control Of Electric Vehicles and Prospects of Supercapacitors", IEEJ Transactions on Electrical and Electronic Engineering, Vol. 4, p. 231-239, 2009

[5] Kumar Yadav Anil,Gaur Prerna, Jha S.K., Gupta J.R.P and Mittal A.P., "Optimal Speed Control Of Hybrid Electric Vehicle", journal Of Power Electronics, Vol. 4, pp. 393-400, 2011.

[6] Liu X., Huang Q., Chen Y., and Li J., "Nonlinear Modelling And optimal Controlling Design for Radar Servo System", Third International Symposium on intelligent Information Technology Application. Pp. 327-330, 2009.

[7] Kim S.H. and Park C., "A Self Organizing Fuzzy Controller for wheeled Mobile Robot using an Evolutionary Algorithm", IEEE Transactions on Industrial Electronics, Vol. 48, No. 2, pp. 467-474, 2001.

[8] Junaid K.M. and Wang S., "Automatic Cruise Control Moddeling- A Lattice PWL Approximation Approach", Proceedings Of IEEE Intelligent Transactions system Conference, pp. 1370-1375, 2006

[9] Ogata K., 'Modern Control Engineering', (India: Pearson Education,2002) 681-751

Optimization Techniques on GPU

Rashmi Sharan Sinha
SBS State Technical Campus,
Ferozepur, Punjab
sinharashmisinha@hotmail.com

Satvir Singh
SBS State Technical Campus,
Ferozepur, Punjab
satvir15@gmail.com

Abstract—In this paper, we present a comprehensive survey on parallelizing computations involved in optimization problem, on GPU using CUDA. Many researchers have reported significant speedup using CUDA on GPU. Stochastic algorithms, Metaheuristic algorithms and Heuristic algorithms i.e., Mixed Integer Non-linear Programming (MINLP), Central Force Optimization (CFO), Genetic Algorithms (GA), Particle Swarm Optimization (PSO), etc. are exploring/exploiting the processing power of GPU. GPGPU shows tremendous speedups of 6x to 7x in Steady State Genetic Algorithm to 10,000x speedups in CFO. GPU have multithread cores with high memory bandwidth which allow for greater ease of use and also more radially support a layer body of applications.

Keywords—GPU, GPGPU, CUDA, MINLP, PGMOEA, CGA, CFO, Optimization Algorithms.

I. INTRODUCTION

General Purpose GPU Computing really took off when CUDA and Stream arrived in late 2006 [1]. GPU constitute a tremendous step towards a usable, suitable, scalable and manageable future-proof programming model [2]. Optimization workloads are very parallel, and so GPUs developed as large-scale parallel computation machines [3] [4] [5] [6]. Originally GPGPU processing was done by tricking the GPU by disguising computation loads as graphic loads [7]. With the advent and large availability of General Purpose Graphics Processing Units and the development and straightforward applicability of the Compute Unified Device Architecture platform, several applications are being benefited by the reduction of the computing time [8]. GPGPU-based architecture, aiming at improving the performance of computationally demanding optimizations for identifiable specific mapping parameters, one can reduce total execution time drastically and also, improve greatly the optimization process convergence. Application performance can be significantly improved by applying memory-access pattern- aware optimizations that can exploit knowledge of the characteristics of each access pattern [3]. To evaluate the effectiveness of our methodology, we have created a tool that incorporates our proposed algorithmic optimizations and report on execution speedup using selected benchmark kernels that cover a wide range of memory access patterns commonly found in GPGPU workloads [9]. Graphics Processing Units (GPUs) are widely used among developers and researchers as accelerators for applications outside the domain of traditional computer graphics. In particular, GPUs have become a viable parallel accelerator for scientific computing with low investment in the necessary hardware.

II. MINLP OPTIMIZATION

With the increasing advent of GPGPU using CUDA, the stochastic algorithm of advanced Genetic Algorithm is used to solve non-convex Mixed integer Non-linear Programming (MINLP) and non-convex Non-linear Programming (NLP) problems [10]. MINLP refers to mathematical programming algorithms that can optimize both continuous and integer variables, in a context of nonlinearities in the objective function and/or constraints.

MINLP problems involve the simultaneous optimization of discrete and continuous variables. These problems often arise where one is trying to simultaneously optimize the system structure and parameters. This is difficult because optimal topology is dependent upon parameter levels and vice versa [10]. In many design optimization problems, the structural topology influences the optimal parameter settings so a simple de-coupling approach does not work: it is often not possible to isolate these and optimize each separately. Finally, the complexity of these problems depends upon the form of the objective function. In the past, solution techniques typically depended upon objective functions that were single-attribute and linear (i.e., minimize cost). However, real problems often require multi-attribute objectives such as minimizing costs while maximizing safety and/or reliability, ensuring feasibility, and meeting scheduled deadlines. In these cases, the goal is to optimize over a set of performance indices which may be combined in a nonlinear objective function. Through this algorithm the intensity of each individual is beamed using entropy measures. The results of the tests shows a significant speedup of 42x with single precision and 20x with double precision over nVidia Fermi C2050 GPU [10].

III. PGMOEA

The general Purpose GPU is efficiently used in optimizing the multiple objective problems. The particle gradient Multiobjective Evolutionary Algorithm (PGMOEA) is used to solve optimization problems. PGMOEA is first experimented on CPU and then after parallelizing the algorithm executed upon GPU which formed a great speedup results [11]. The experiment is conducted upon two different examples. The first example shows a speedup of 9x with nVidia GeForce GTX285 then CPU result. While the second example is 10x faster than that of CPU [11]. The speedup comparison is shown below in Table 1.

Table I. Speedup Comparison

Algorithm	Example 1		Example 2	
	Time(s)	*Speedup*	*Time(s)*	*Speedup*
PGMOEA on GPU	0.97	9.95	0.83	10.64
PGMOEA on CPU	9.01	1.04	8.02	1.10

IV. CELLULAR GENETIC ALGORITHM

Genetic Algorithm have a subclass known as Cellular Genetic Algorithm (cGA) which provides the data of population structured in several specified topologies [12]. The cGA is compared upon CPU, single GPU and multiGPU. The nVidia GTX285 multiGPU test shows a speedup of 8 to 771 times then single GPU [12]. The multiGPU is more prominent in paralleling the algorithm and producing accurate results as there is a need of special maintenance to perform same experiment upon single GPU.

V. DIFFERENTIAL EVOLUTIONARY ALGORITHM

GPGPU is proved to be great architectural unit in reducing the processing time [13]. The Differential Algorithm which is one of the part of Evolutionary Algorithm is implemented upon CPU using C-CUDA. The motivating features of Differential Algorithm are easy for parallelization and convergence properties which intern gives an appropriate result. The algorithm is first tested upon CPU then on nVidia GTX285 with 1GB GDDR3 GPU with the speedup outcomes. GPU gives 20x to 35x faster results which proves GPU is much more effective and efficient than Differential Algorithm on CPU [13]. The Speedup comparison results are shown in Table 1.

VI. CELLULAR AUTOMATA

Cellular Automata have various real life application like physical system modeling, road traffic simulation, artificial life simulation, etc [14], [15], [16]. Cellular automata design evolved from evolutionary algorithm and a part of Genetic Algorithm which is complex in nature. The Algorithm is parallelized and implemented upon GPGPU shows an efficient reduction in execution time.

The rules of Cellular Automata take longer time period inevolution in sequential execution. The same Genetic Algorithm shows 31.34x to 314.94x speedup when executed upon nVidia GeForce FX280 GPU which is a significant reduction in execution time [17].

VII. ACCELERATING PSO

PSO is a metaheuristic algorithm works by having a swarm of particles [18]. These particles are moved around in the search-space according to a few simple formulae. The movements of the particles are guided by their own best known position in the search-space as well as the entire swarm's best known position [19]. When improved positions are being discovered these will then come to guide the movements of the swarm. Particle Swarm Optimization (PSO) is one of the type of Evolutionary Algorithm used to optimize the multiple objective problems. When an optimization problem involves more than one objective function, the task of finding one or more optimal solutions is known as multi-objective optimization [18]. The objects are having random velocities and positions. The algorithm is tested upon three different platforms of C, Matlab and C-CUDA. The parallel implementation of PSO on nVidia GTX 280 gives 17 to 41 times speedup in computing time in C-CUDA as compared with the C and Matlab as Shown in Fig.1 [20].

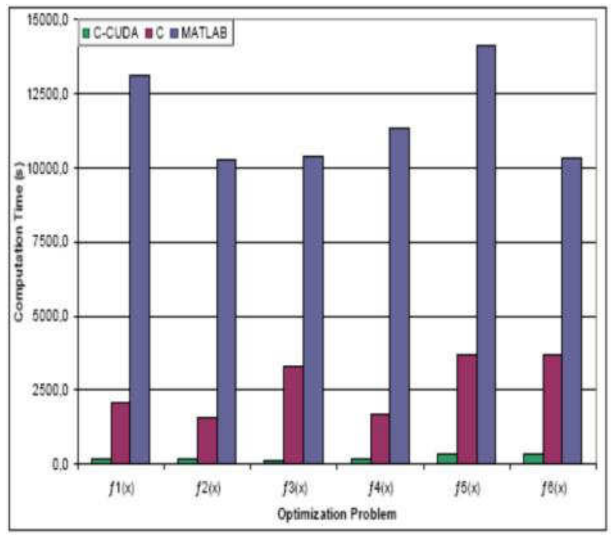


Fig. 1. Computing time for C-CUDA, C AND MATLAB

VIII. STEADY STATE GENETIC ALGORITHM ON GPU

The optimization problem is effectively solved by the means of Evolutionary Computing [21]. The steady state Genetic Algorithm used to access optimization algorithms. These algorithms basically have selection for the reproduction and selection of survival implementation with concurrent kernel execution [22].

The study is first performed upon CPU then with nVidia GeForce GTX480 GPU gives a speedup of 3x to 6x then the previous implementation on CPU [12]. The executed time is greatly reduced using general purpose GPU. The population individual data is accessed parallelaly which effectively speedup the process.

IX. BINARY-CODED AND REAL-CODED GENETIC ALGORITHM

Genetic Algorithm is tested and evaluated on parallel implementation on C-CUDA API on the parameters like population size, number of threads, problem size and problem of differing complexities with variation in the population individuals [12]. For an efficient implementation on GPGPU the solution is thoroughly implemented along with the operators like random number generation, initialization, selection operation, and mutation operations [13]. The nVidia GeForce 8800GTX shows overall speedup of 40-400 on three different test problems [23]. Thus parallel implementation is more effective then sequential process as compared with clock time and accuracy.

X. CENTRAL FORCE OPTIMIZATION (CFO)

The metaheurestic algorithm Central Force Optimization (CFO) is implemented upon GPGPU using local neighborhood and implemented CFO concepts [24]. The calculation of CFO is dependent upon the movement of probes which are scattered all over the space. The probes then slowly move towards the probe having highest mass or fitness. PR-CFO is the most evaluated algorithm with the measures of initial position and acceleration vectors, fitness evaluation and probe movements [25]. The test problems is having the dimension of 30 to 100 of four different examples of Pseudo random CFO (PR-CFO). The PR-CFO is tested with four test types i.e. Ring, Standard, CUDA, CUDA Ring. PR-CFO shows a speedup of 4 to 400 using CUDA. PR-CFO ring and PR-CFO CUDA ring on nVidia Tesla C1060 shows 10,000 times faster results as compared with standard PR-CFO algorithm [25].

XI. CONCLUSION

In this paper we present different optimization algorithm with tremendous speedup in the computation time. MINLP archived an overall speedup of 20x to 42x using nVidia Tesla C2050 GPU as compared to intel Core i7 920 CPU processor. The new binary-coded and real-coded Genetic Algorithm using CUDA leads to a performance improvement with the speedup of 40x to 400x. Central Force Optimization (CFO) results in reduction of computing time and a speedup of 10,000x. The Cellular Automata shows 314.97x as compared with the sequential implements.

REFERENCES

[1] K. S. Perumalla, "Discrete-event execution alternatives on general purpose graphical processing units (GPGPUs)," in Principles of Advanced and Distributed Simulation, 2006. PADS 2006. 20th Workshop on, 2006, pp. 74–81. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1630711

[2] A. Jablin, P. McCormick, and M. Herlihy, "Scout: High-performance heterogeneous computing made simple," in Parallel and Distributed Processing Workshops and Phd Forum (IPDPSW), 2011 IEEE International Symposium on, 2011, pp. 2093–2096.

[3] Fast Parallel Markov Clustering in Bioinformatics Using Massively Parallel Graphics Processing Unit Computing, 2010.

[4] Data Processing in Space Weather Physics Models in the Meridian Project, 2010.

[5] Design and Implementation of Remote Parallel Computing System Based on Multi-Platform, 2010.

[6] An efficient parallel algorithm for evaluating join queries on heterogeneous distributed systems, 2009.

[7] Profiling General Purpose GPU Applications, 2009.

[8] D. Luebke, M. Harris, N. Govindaraju, A. Lefohn, M. Houston,

[9] J. Owens, M. Segal, M. Papakipos, and I. Buck, "Gpgpu: generalpurpose computation on graphics hardware," in Proceedings of the 2006 ACM/IEEE conference on Supercomputing. ACM, 2006, p. 208.

[10] Evolving a CUDA kernel from an nVidia template, 2010.

[11] Advanced genetic algorithm to solve MINLP problems over GPU, 2011.

[12] Particle Gradient Multi-objective Evolutionary Algorithm Based on GPU with CUDA, 2010.

[13] A multi-GPU implementation of a Cellular Genetic Algorithm, 2010.

[14] Differential evolution algorithm on the GPU with C-CUDA, 2010.

[15] L. J. Durbeck and N. J. Macias, "The cell matrix: an architecture for nanocomputing," Nanotechnology, vol. 12, no. 3, p. 217, 2001.

[16] M. Gardner, "Mathematical games: The fantastic combinations of john conways new solitaire game life," Scientific American, vol. 223, no. 4, pp. 120–123, 1970.

[17] M. Tomassini, M. Sipper, and M. Perrenoud, "On the generation of high-quality random numbers by two-dimensional cellular automata," Computers, IEEE Transactions on, vol. 49, no. 10, pp. 1146–1151, 2000.

[18] GPU Accelerators for Evolvable Cellular Automata, 2009.

[19] J. Kennedy, R. C. Eberhart, and Y. Shi, "Swarm intelligence. 2001," Kaufmann, San Francisco, vol. 1, pp. 700–720, 2001.

[20] J. Kennedy and R. Mendes, "Population structure and particle swarm performance," in Evolutionary Computation, 2002. CEC'02. Proceedings of the 2002 Congress on, vol. 2. IEEE, 2002, pp. 1671–1676.

[21] Swarm's flight: Accelerating the particles using C-CUDA, 2009.

[22] F. Stentiford, "An evolutionary programming approach to the simulation of visual attention," in Evolutionary Computation, 2001. Proceedings of the 2001 Congress on, vol. 2, 2001, pp. 851–858. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=934279

[23] R. Arora, R. Tulshyan, and K. Deb, "Parallelization of binary and real-coded genetic algorithms on GPU using CUDA," in Evolutionary Computation (CEC), 2010 IEEE Congress on, 2010, pp. 1–8. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5586260

[24] Parallelization of binary and real-coded genetic algorithms on GPU using CUDA, 2010.

[25] A. Stefek, "Benchmarking of heuristic optimization methods," in MECHATRONIKA, 2011 14th International Symposium. IEEE, 2011, pp. 68–71.

[26] Central Force Optimization on a GPU: A case study in high performance metaheuristics using multiple topologies, 2011.

Design and Implementation of a Low Power and Constant-g_m CMOS OTA with Rail-to-Rail Operation

Vivek Kumar
Dept. of School of VLSI Design and Embedded System
National Institute of Technology
Kurukshetra, India
vivekchoudhary6@gmail.com

Kamal K. Kashyap
Dept. of Electronics and Communication Engineering
National Institute of Technology
Kurukshetra, India
kamalkant7185@gmail.com

Abstract— **In this paper, we propose constant dc output, power efficient and constant-g_m 1.8V CMOS operational transconductance amplifier with rail-to-rail input ranges. The proposed OTA has been integrated with CMOS process with transistor length of 180nm. The op-amp is compound structure of constant-g_m rail-to-rail input stage, folded cascode summing stage and class AB output stage. Here we have discussed the first two out of three parts. The designed OTA is a compound structure of NMOS and PMOS differential pair connected in parallel. The compound structure achieves rail-to-rail operation. We have used 1:3 times current mirror techniques for obtained constant-g_m. Simulation results show that the proposed OTA achieves variation of dc output is 4.5% , dc gain is 57.56 db, phase margin is 73.06deg and unity gain frequency is 249.47MHz . The proposed OTA designed with minimum source usages to reduce power consumption in circuit. The power consumption reduced by 30% compared to the previous rail-to-rail OTA and obtained less than 6% deviation in transconductance. The power consumed by the proposed OTA is only 1.5mW.**

Keywords—OTA; low power; constant-g_m; weak inversion; complementary input pair.

I. INTRODUCTION

In VLSI circuits the trend is towards combining more and more electronics devices on a single chip and the allowable lower supply voltage and lower power consumption. So in VLSI circuits smaller device dimension entail lower breakdown voltage, while higher device densities result in lower power dissipation per functional circuit cell. Today, the increasing use of portable equipment, whose powered by batteries or solar cell. So our requirements are, portable equipment dictates lower supply voltage and very low power consumption.

In this paper we discuss OTA based circuit because the OTA is a vital analog building block for many application is the largest and most power consumption block [1]. The op-amp must also have a constant gain bandwidth product over the whole input common-mode ranges. To obtain the constant unity gain frequency over the whole common-mode input range, the transconductance of the input stage of OTA has to be constant. Here we discuss two method for obtain a constant-g_m. The first method uses 1:3 times current mirror at the input stage of OTA to control the g_m. in this first method the g_m is directly proportional to tail current, common-mode input voltage and W/L ratio of input transistors. In proposed method we obtained variation in transconductance less than 6% and variation in dc output 4.5%.In second method g_m control by square-root current control [2]-[3]. In this method g_m is directly proportional to the sum of square root of the tail currents and obtained variation of g_m is about 8.0% over the whole common-mode input ranges. But in this circuit several current path uses, which considerable raise the power consumption of the input stage. Another disadvantage is that the g_m control is rather complex in nature. So we preferred first method to control g_m.

II. COMMON-MODE INPUT STAGE

Generally used input stage of operational amplifier is a single differential pair. A single differential pair either a N-channel input pair N1, N2 or P-channel input pair P1, P2, which is shown in Fig 1.

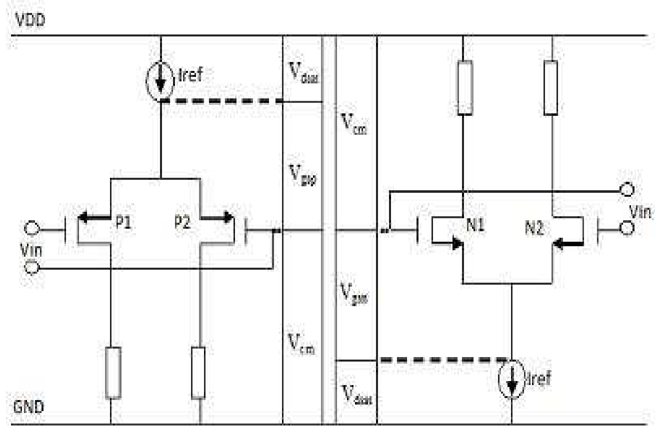


Fig. 1 Common-mode input range for single differential pair

The common-mode input range of P-channel input pair is given by-

$$GND<V_{cm}<V_{dd}- V_{dsat} - V_{gsp} \tag{1}$$

Where V_{cm} is the common-mode input voltage, V_{dsat} is a saturation voltage across the current source, V_{dd} is the supply voltage and V_{gsp} is the gate to source voltage of a P-channel input transistor.

The common-mode input range of N-channel input pair is given by:

$$V_{gsn} + V_{dsat} < V_{cm} < V_{dd} \tag{2}$$

Where, V_{dsat} is a saturation voltage across the current source whose connected to source of N-channel and V_{gsn} is the gate to source voltage across the N-channel input pair.

Minimum supply voltage required for obtained full rail-to-rail common-mode input range is given by-

$$V_{sup,min} = V_{gsp} + V_{gsn} + 2\, V_{dsat} \tag{3}$$

When the supply voltage is greater than the minimum supply voltage than three region obtained-

1. At the negative supply rail, P-channel transistors pair must be operate so, the total transconductance is given by $g_{mt} = g_{mp}$.
2. At positive supply rail, N-channel transistors pair must be operate so, the total transconductance is given by $g_{mt} = g_{mn}$.
3. At the intermediate stage both the N-channel and P-channel pair are operates so, the total transconductance is given by $g_{mt} = g_{mn} + g_{mp}$.

For increasing the common-mode input range, we use parallel connection of P-channel input pair and N-channel input pair shown in Fig 2.

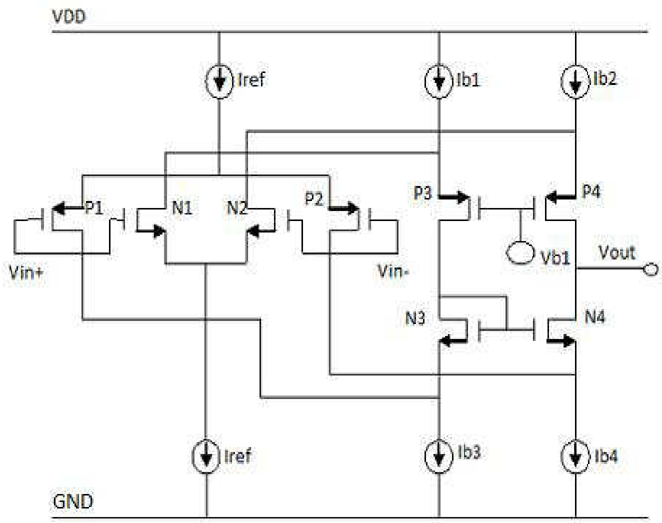


Fig. 2 OTA with rail-to-rail input stage

If the transistor operates in week inversion than transconductance is directly proportional to tail current [4].

$$g_{m,pn-weak} = \frac{I_p}{2n_pV_{th}} + \frac{I_n}{2n_nV_{th}} \tag{4}$$

Where, n_n and n_p are the slope factor of complementary transistor, V_{th} is the thermal voltage and I_n and I_p are the tail current of N-channel and P-channel pair transistor, respectively.

In week inversion saturation voltage is lower than strong inversion. So we use strong inversion for control constant-g_m [3],[5]. in the strong inversion tramsconductance is directly proportional to square root of tail current and W/L ratio of complementary transistors.

$$I_p = \tfrac{1}{2}\, \mu_p C_{ox}(W/L)_p(V_{gsp} - V_{th})^2 \tag{5}$$

$$I_n = \tfrac{1}{2}\, \mu_n C_{ox}(W/L)_n(V_{gsn} - V_{th})^2 \tag{6}$$

Here $V_{gs} - V_{th} = V_{gs,eff}$

$$g_{mnp} = g_{mn} + g_{mp} \tag{7}$$

$$g_{mnp} = \sqrt{2\mu_n C_{ox}(\frac{W}{L})_n I_n} + \sqrt{2\mu_p C_{ox}(\frac{W}{L})_p I_p} \tag{8}$$

$$g_{mnp} = \mu_n C_{ox}(\frac{W}{L})_n V_{gsn,eff} + \mu_p C_{ox}(\frac{W}{L})_p V_{gsp,eff} \tag{9}$$

Where, $V_{gs,eff}$ is the effective gate to source voltage of input complementary transistors.

III. Power Efficient Constant-G_M Technique

In previous research paper we have studied, that constant-g_m obtained by many technique, one is the current mirror technique. The main advantages of current mirror are its small die area and its low power consumption and also do not increase the noise in the input stage. With help one times current mirror we obtained constant-g_m about 41% variation [1]. So another technique we used three times current mirror.

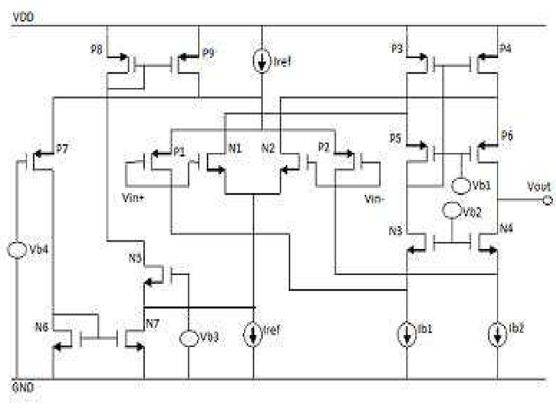


Fig. 3 OTA with rail-to-rail input stage and g_m control by 1:3 times current mirror

In this given figure 3 we used two three times current mirror and two current switches N5 and P7 to obtain a constant-g_m over whole common-mode input range.

$$\sqrt{2\mu_n C_{ox}(\frac{W}{L})_n I_n} + \sqrt{2\mu_p C_{ox}(\frac{W}{L})_p I_p} = 2P\sqrt{I_{ref}}$$

(10)

When, $I_p = I_n = I_{ref}$

And

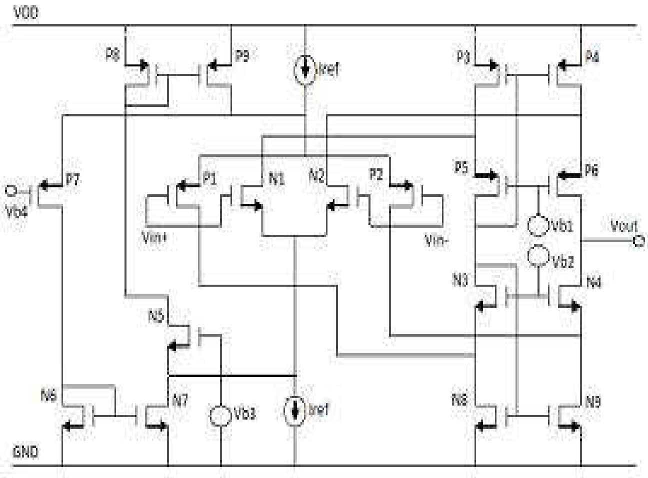
$$\mu_n C_{ox}(\frac{W}{L})_n = \mu_p C_{ox}(\frac{W}{L})_p = P^2$$

(11)

$$g_{m,np} = g_{m,n} + g_{m,p}$$

(12)

The size of input complementary pair have been chosen according to the eq(11) if the common-mode voltage below Vb3, transistor N1 and N2 are off, the current switch N5 is on and current switch P7 is off. The current switch N5 take away the tail current from N-channel input pair and feed into current mirror P8-P9. Than output of this current mirror current is added with the tail current of P-channel input pair and feed into P-channel. Then also we can explain this process for common-mode voltage above the Vb3. In the above Fig 3 g_m control by the two three times current mirror these current mirror operates by the current switches, is that both the current switch can conduct at very low supply voltage, so it is the main disadvantage of conventional circuit. To avoid this problem we use a single ended differential amplifier at input of current switch P7 and single ended output is allowed to the gate of P7 transistor. Than the current switch P7 will be fully dependent on biased voltage Vb5. For instance, insuring that the P7 transistor is always turned off at very low common-mode input voltage. Which current mirror circuit N6-N7 also not working at this stage. So we can reduce the power consumed by the switch P7 and current mirror circuit N6-N7 at very low common-mode input voltage.



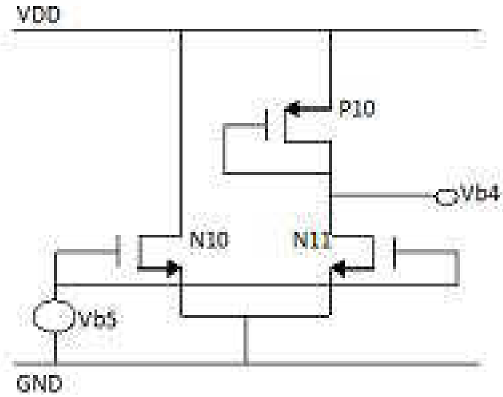


Fig. 4 Proposed OTA circuit with rail-to-rail input stage and constant-g_m over whole common-mode input ranges

IV. SIMULATION RESULT

The proposed operational transconductance amplifier has been implemented in 180nm CMOS process technology. The simulation result are obtained using SpectreS simulator from cadence. Figure 5 represent the AC response of the proposed OTA, in this Fig shown that we obtained DC gain 57.56db, unity gain frequency is 249.47MHz and The obtained phase margin is 73.06deg. Voltage reference and current reference are generated for biasing the transistors and different current sources [6]. Fig 6 represent the DC output of proposed OTA with common-mode input ranges and variation of obtained DC output is 4.5%. We obtained the approximately constant-g_m across whole common-mode input range, show in Fig 7. The maximum deviation in transconductance is 6%. Figure 8 represents the current gain; the maximum current gain is 142db. Figure 9 represent the transient response of the proposed operational transconductance amplifier for a sinusoidal input with rail-to-rail swing and its peak-to-peak amplitude varies from 0.35V to 1.60V. The power consumed by the proposed OTA is only 1.5mW. The overall performances of the proposed OTA are listed in Table1 and comparisons with other works in literature are listed in Table 2.

TABLE I. Summary of a proposed circuit simulated result

Parameter	Simulated result
Supply voltage	1.8 V
Technology	180nm
Temperature	27^0C
DC-gain	57.56db
Unity gain frequency	249.47MHz
Open loop phase margin	73.06deg
Output voltage swing	0.35-1.60V
Variation in g_m	Less than 6%
Maximum current gain	142db
Power consumption	1.5mW

Fig. 5 AC response of the proposed OTA

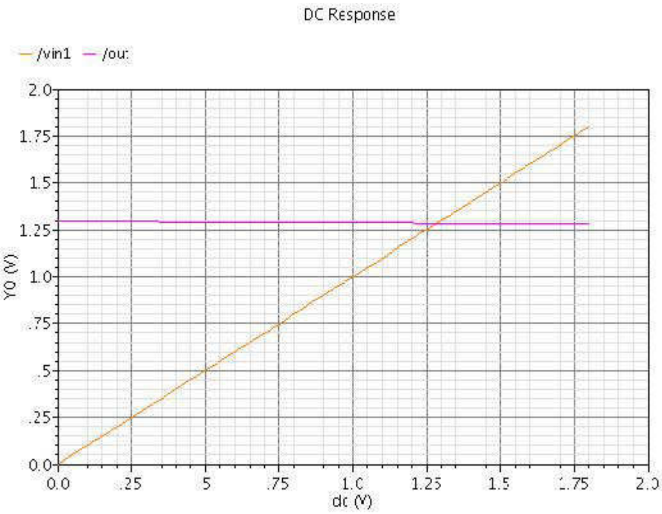


Fig. 6 Constant DC output voltage over the whole common-mode input ranges

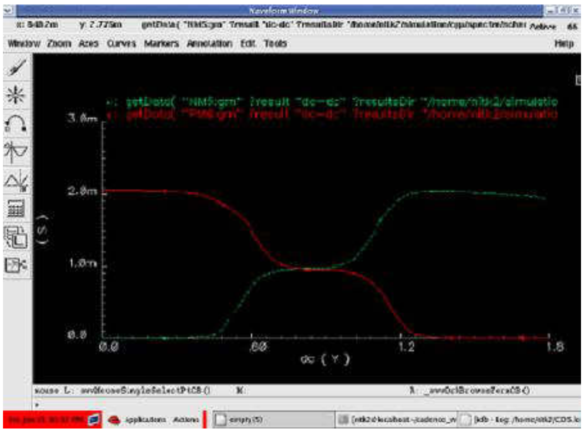


Fig. 7 Transconductance over the whole common-mode input ranges

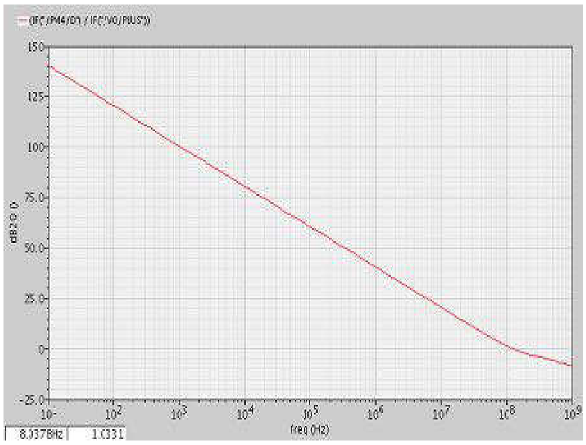


Fig. 8 Current gain of the proposed OTA

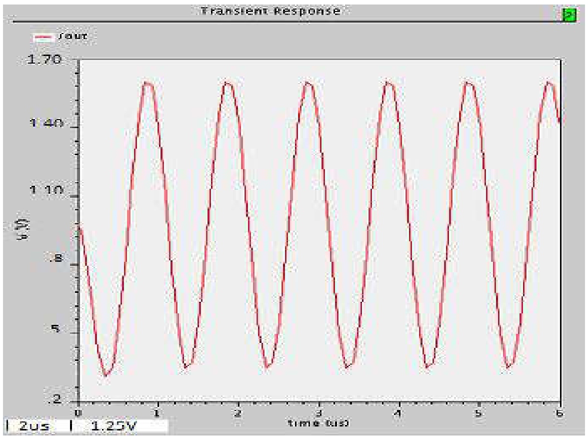


Fig. 9 Transient response of the proposed OTA for a sinusoidal input with rail-to-rail swing

V. CONCLUSION

The paper has presented the low power, high gain and constant-g_m rail-to-rail swing OTA. In this paper we used 1:3 times current mirror circuit to realize constant-g_m over the whole common-mode input range and we used minimum to minimum sources to reduce power consumption in proposed circuit. The power consumption in proposed circuit is only 1.5mW in 180nm CMOS process technology with 1.8 V power supply. We can use this proposed OTA in portable communication, sensor interface, charge pump based PLL, active filters etc.

TABLE 2: Comparisons with other works in literature

Refer ences []	Process technology /supply voltage	Power consum ption (mW)	UGF (MHz)	PM (deg)	g_m vaariati on
[1]	3.3 V	NA	0.165	NA	16.00%
[6]	180nm	3.24	700.7	63.85	NA
This work	180nm	1.5	249.47	73.06	Less than 6%

REFERENCES

[1]. Sansen, Analog design essentials, Springer, Dordrecht, The Netherlands, 2006.

[2]. R. Hogervorst, R. J. Wiegerink, P. A. L. de Jong, L. Fonderie, R. E. Wassenaar, and J. H. Huijsing, "CMOS low-voltage operational amplifier with constant-g_m rail-to-rail input stage," Analog Integrated Circuits and Signal Processing, vol. 5, 1993, pp. 135-146.

[3]. J. H. Huijsing, K. J. Dc Langen, R. Hogervorst, and R. G. H. Eschauzier, "Low voltage low power opamp based amplifier," Analog integrated circuits and signal processing, vol. 8, no. 1, 1995, pp. 49-67.

[4]. J. H. Huijsing and D. Linebarger, "Low voltage operational amplifier with rail-to-rail input and output ranges" IEEE J. Solid-State Circuits, vol. 20, 1985, pp. 1144-1150.

[5]. Il Kwan Chang, Jang Woopark, Se Jun Kim, Kae Dal Kwack "A global operational amplifier with constant-gm input and output satage," proceeding of the IEEE region 10 conference vol. 2, 1999, pp. 1051-1054.

[6]. M. K. Hati and Tarun K. Bhattacharyya, "design of a low power, high speed complementary input folded regulated cascade OTA for a parallel pipeline ADC," IEEE Computer Society International Symposium on VLSI(ISVLSI 11), July 2011, pp. 114-119.

Power Reduction in Arithmetic Algorithm Using Dynamic Voltage And Frequency Scaling (DVFS)

Deepak Garg
Dept. of VLSI Design and Embedded Systems
National Institute of Technology
Kurukshetra, India.
deepakgarg.vlsi@gmail.com

Rajender Kumar Sharma
Dept. of VLSI Design and Embedded Systems
National Institute of Technology
Kurukshetra, India.
mail2drrks@gmail.com

Abstract— **In the recent years, the main concern of VLSI Engineers is on the power reduction techniques. We are using Dynamic Voltage Frequency Scaling (DVFS) for reducing power in Spartan 3E FPGA Kit along with XPower Estimator tool. Here we have done practical analysis of sequential multiplier and simulating its power without DVFS and further we simulate the result with DVFS. Analyzing this result we found power has been reduced drastically. Our simulation results shows that 10.31% of total power and 31.58% of dynamic power is reduced using our approach as compared to non DVFS approach.**

Keywords— DVFS; Low Power; Spartan 3E Kit; Xilinx; XPower Estimator.

I. INTRODUCTION

The functions of mobile multimedia consumer electronics such as portable media players, smart phones, and digital cameras are rapidly diversifying as a result of digital convergence. Furthermore, the resolution of images and videos as well as the complexity of games handled by mobile electronics are continually growing. Concurrent execution of multiple applications and processing of high-definition audio, video and images require strong processor performance. However, as the power consumption of a processor increases in proportion to both the clock frequency and the square of the operating voltage, stronger processors generally require more power [1].

Dynamic voltage and frequency scaling (DVFS) is a technique in which we can change voltage and frequency at different interval of time according to the circumstances which helps to reduce the power. Now a day's power reduction is the main concern of the new technology. As in old technology to reduce power cooling circuits are used and after this we used metal sink technology in which we attach metal sheet with large area for cooling purpose which increased the operation cost. Over the last decade, manufacturers competed to advance the performance of processors by raising the clock frequency. However, the dynamic power consumption Pdynamic of a CMOS-based processor, the power required during execution of instructions is related to its clock frequency f and operating voltage Vdd as Pdynamic $Vdd^2.f$. And, the relation Vdd f also holds in these processors. As a result, the dramatically increased power consumption caused by high clock frequency has stopped the race, and they are now concentrating on other ways to improve performance at relatively low clock frequencies [2,3]. DVFS is a new technology in which we manage frequency according to our work requirement. DVFS is a hardware technology used to dynamically scale up or down the processor frequency according to the policy and the workload demand [4].

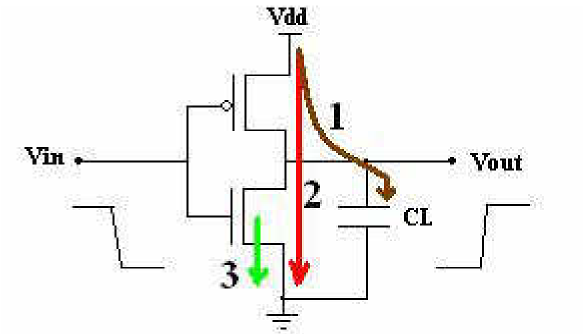


Fig. 1. Power dissipation in CMOS inverter [5].

Increasing or decreasing processor frequency will influence the general power consumption of a system. The total power dissipation in CMOS circuits is the summation of static power, short circuit power and dynamic power. These components of power dissipation are shown in Fig. 1 for a simple CMOS inverter.

There are three major sources of power dissipation in digital CMOS circuits, which are summarized in the following equation:

$$P_{total} = (C_L.V.V_{dd}.f_{clk}) + I_{sc}.V_{dd} + I_{leakage}.V_{dd} \quad (1)$$

The first term represents the switching component of power, where C_L, is the loading capacitance, f_{clk} is the clock frequency, and is the probability that a power consuming transition occurs (the activity factor) [6] [7]. In most cases, the voltage swing V is the same as the supply voltage V_{dd}; however, in some logic circuits, such as in single- gate pass-transistor implementations, the voltage swing on some internal nodes may be slightly less. The second term is due to the direct-path short circuit current I_{sc} which arises when both the NMOS and PMOS transistors are simultaneously active, conducting current directly from supply to ground [8], [9]. Finally, leakage current $I_{leakage}$, which can arise from substrate injection and subthreshold effects, is primarily determined by fabrication technology considerations [10].

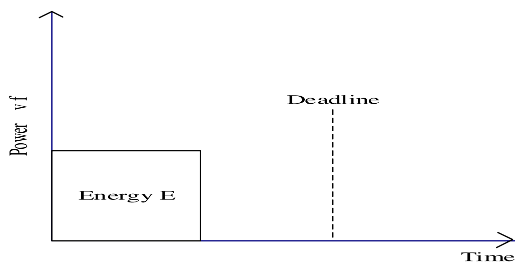
By the Ohm's law stating that power consumption is proportional to the square of the applied voltage, lowering the supply voltage could greatly save power consumption of the processor [11]. However, the effect of the scheduling algorithms is diminished when applications greedily raise the utilization rate of the CPU. Therefore, applications should maintain low utilization rates in order to reduce power consumption [12]. The rapid increase in the complexity and speed of each new CPU generation is outstripping the benefits of voltage reduction and feature size scaling [13].

Dynamic voltage frequency scaling technique is effective in reducing power dissipation by lowering the supply voltage and circuit frequency. The energy consumption of a program can be minimized by reducing the switching capacitance of each operation or by reducing the voltage at which these operations are performed or by decreasing the number of operation performed [14]. When frequency is being decreased, then there is performance is lowered [15]. When voltage is increased, then there is performance is raised. Scaling frequency down when it is not critical for performance also allows voltage to be scaled down [16]. Throughout this paper a noble approach is used to reduce the power consumption by using DVFS technique with XPower Estimator tool.

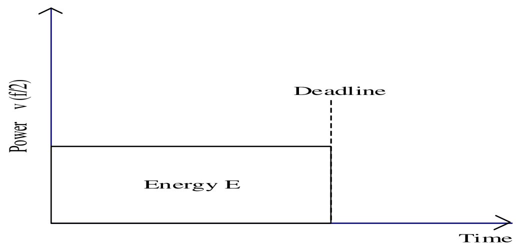
II. Analysis of DVFS technique

The high power consumption of a processor is becoming a critical problem for both battery-powered devices and high-performance computers [17]. It reduces circuit reliability, complicates the cooling technology, shortens the battery lifetime, and increases the production and operation costs of a CPU. One effective technique, called dynamic voltage scaling (DVS), achieves CPU power reduction through lowering the CPU supply voltage and clock frequency at runtime [18]. DVFS allows for the decrease of the clock frequency and voltage of the processor, leading to a decrease of the power consumption, at the cost of an increased latency [19] [20].

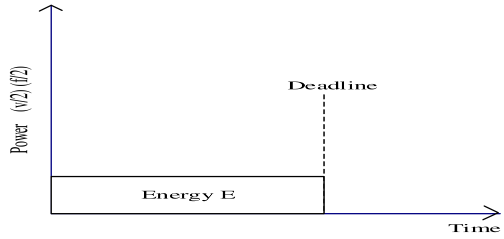
When the clock frequency f_{clk} is shrinked by half, this brings down the processor's power consumption and still allows task to accomplish by deadline, the energy remains the same as shown in Fig. 2(b). Turning down the voltage level vdd by half brings down the power level to a greater extent without any rise in execution time. Thus the energy consumption is cut significantly as shown in Fig 2(c) [21].



(a)



(b)



(c)

Fig. 2. Power consumption vs Energy consumption for a task, which is ready at t=0 and complete at t=T, with maximum frequency fclk [22].

III. Proposed Work

A. *Design Methodology*

The methodology of DVFS is shown in figure 3. Initially, we implement a multiplier using verilog. This multiplier is being operated at fix voltage and frequency. After this we analyze the same design with varying frequency and voltage. In our DVFS approach, there is a different frequency f_1 for multiplication and a different frequency f_2 for addition. Then give the different range of voltages.

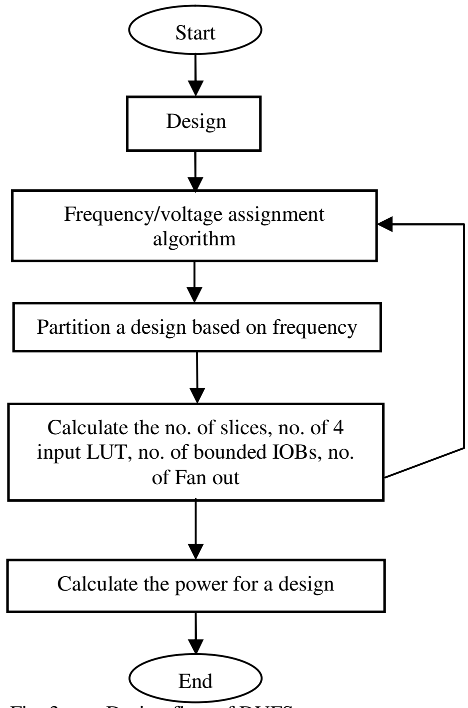


Fig. 3. Design flow of DVFS.

Using the simulation technique in Xilinx, we find the number of slices, number of input LUTs, number of bounded IOBs, number of Fanouts. After that we find the total power

(Dynamic + Static Power) for a defined number of slices, number of LUTs, number of bounded IOBs, number of Fanouts, frequency, voltage by using a XPower Estimator tool.

IV. Experimental Results and Comparison

We use a multiplier of 8 bit to test the proposed methods. Initially multiplication takes place without implementing DVFS technique and in this multiplication and addition takes place at same frequency and then the power consumption is calculated. But in our proposed DVFS approach multiplication is done at frequency f_1 and addition is taken at another frequency f_2 and then power is being calculated as shown in Fig. 4. All these modules are designed in Verilog HDL synthesized and implemented using Xilinx ISE 14.5 on Spartan 3E FPGA. Specifications of Spartan 3E FPGA kit that is used are shown in Table I.

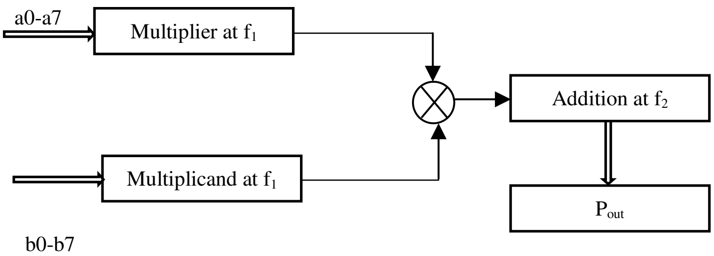


Fig. 4. Multiplier Architecture

TABLE I. SPECIFICATIONS OF SPARTAN 3E FPGA KIT

S.No.	Specifications	Values
1	Device Type	XC3S 500E
2	Package	FGG320 AGQ0625
3	Lot Code	A1414957AO
4	Speed Grade	4
5	Temperature Range	0^0-85^0C

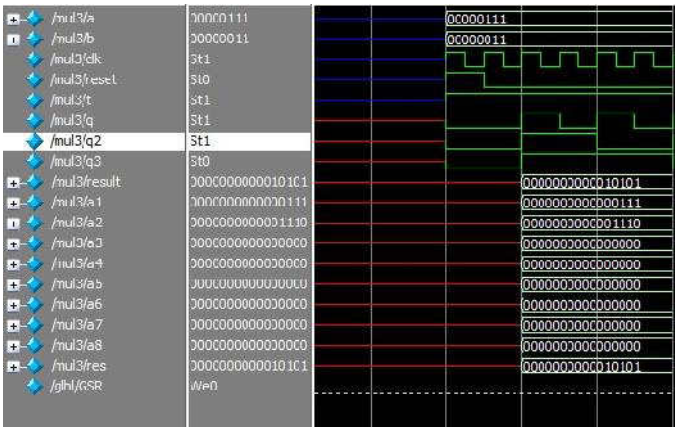


Fig. 5. Post Layout Simulation of 8 Bit Multiplier

Fig. 5 shows the post layout simulation result of the 8 bit Multiplier. The design was simulated using Modelsim PE 6.3. After programming and simulation there are utilization parameters like number of slices, number of 4 input LUTs, number of bounded IOBs, number of Fanouts are being noted down as shown in Table II.

TABLE II. MULTIPLIER VALUES WITH & WITHOUT DVFS

S.No.	Utilization Parameters	Numbers Without DVFS	Numbers With DVFS
1	Number of Slices	34	60
2	Number of 4 input LUTs	65	53
3	Number of bounded IOBs	36	38
4	No of Fanout	32	97

By considering these utilization factors the power is being calculated as shown in Table III. The power consumption is calculated to be 97mw for multiplier without DVFS using XPower Estimator tool.

TABLE III. POWER CONSUMPTION WITH & WITHOUT DVFS

Types of Power	Power Without DVFS (mw)	Power With DVFS (mw)
Quiescent	78	74
Dynamic	19	13
Total	97	87

When multiplication is done using DVFS technique multiplication of two numbers is done at frequency f_1 and addition is at frequency f_2 which is s shown in Fig.4. After programming and simulation there are utilization parameters like number of slices, number of 4 input LUTs, number of bounded IOBs, number of Fanouts which are given in Table II. By considering these utilization factors the power consumption is being calculated as shown in Table III. The power consumption calculated to be 87 mw for multiplier with DVFS using XPower Estimation tool. The result shows a 10.31% power reduction using DVFS technique.

Power consumption at 100MHz with different voltages for circuit with and without DVFS is shown in the fig 6. It is observed that the power consumption with DVFS is less than that of without DVFS.

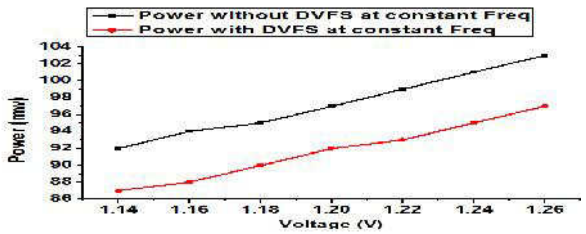


Fig. 6. Comparision of power consumption of circuit with and without DVFS at different operating voltages and fixed frequency.

Power consumption at 1.2 V and different frequencies for circuit with and without DVFS is shown in the fig 7. It is observed that the power consumption with DVFS is less than that of without DVFS.

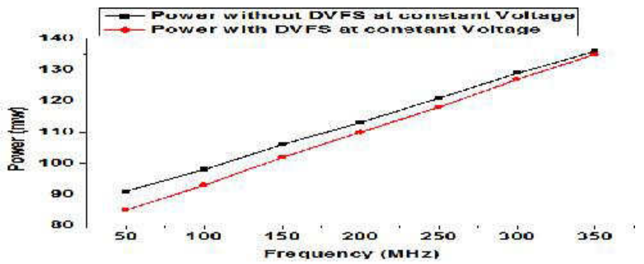


Fig. 7. Power comparison with and without DVFS at different operating frequencies and constant voltage.

The comparison of Slices, IOBs, Fanout, Power with DVFS and without DVFS is given in Fig. 8. The graph shows that with DVFS Slices and Fanout increased whereas No. of LUTSs and Power reduces significantly. This shows that Fanout got increased and we can connect more number of gates at output and reduced power helps in increasing battery life.

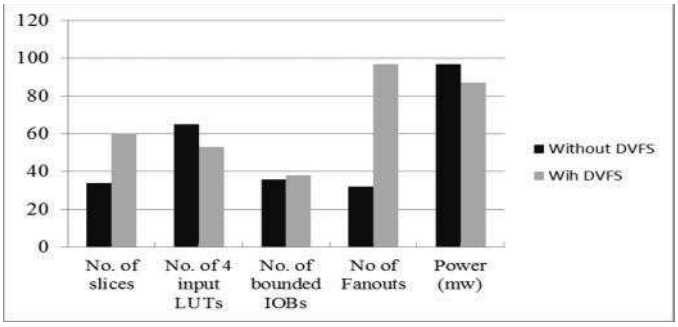


Fig. 8. Comparison of Utilization parameters with and without DVFS.

V. CONCLUSION

A DVFS technique is implemented to reduce the power consumption of Spartan 3E FPGA. An algorithm is designed using Xilinx ISE 14.5 and power is calculated using XPower Estimation tool at different voltages and frequencies. Our proposed algorithm helps in reducing 10.31% of total power and 31.58% of dynamic power.

ACKNOWLEDGMENT

I am extremely thankful to Pankaj Kumar for his valuable guidance, which has helped me to complete my project work efficiently. I express my gratitude to all my friends and well-wishers who have helped me throughout this work.

REFERENCES

[1] Y.Seo, J.Kim and E.Seo, "Effectiveness Analysis of DVFS and DPM in Mobile Devices,"Journal Of Computer Science And Technology, Vol. 27, No.4, pages 781-790, July 2012.

[2] E. Seo,J. Jeong,S. Park,J. Lee, "Energy efficient scheduling of real-time tasks on multicore processors," IEEE Transactions on Parallel and Distributed Systems, Volume 19, No. 11, pages 1540-1552, November 2008.

[3] A.M.K.Cheng, Y.Wang, "ADynamic Voltage Scaling Algorithm for Dynamic Workloads," Journal of Signal Processing Systems, volume 52, pages 45-57, 2008.

[4] C. M. Kamga, "CPU frequency emulation based on DVFS,"IEEE International Conference on Utility and Cloud Computing, pages 367-374, 2012.

[5] D. R. Suleiman, M.A. Ibrahim, I. I. Hamarash, "Dynamic Voltage Frequency Scaling For Microprocessors Power And Energy Reduction,".

[6] A. P. Chandrakasan, "Low Power CMOS Digital Design,"IEEE Journal of Solid-State Circuits, Vol.27, No.4, pages 473-484, April 1992.

[7] S. Kaxiras and M. Martonosi, "Computer Architecture Techniques for Power-Efficiency," Morgan and Claypool Publishers, 2008, ISBN: 9-781-598292-091.

[8] H. J. M. Veendrick, "Short-circuit dissipation of static CMOS circuitry and its impact on the design of buffer circuits," IEEE Journal of Solid-State Circuits, Vol. SC-19, pages 468-473, August 1984.

[9] N.H.E.Weste, and D.M. Harris, "CMOS VLSI Design: A circuits and systems perspective," 4th ed. , Addison Wesley, 2011, ISBN: 978-0-321-54774-3.

[10] J.P.Uyemura, "CMOS Logic Circuit Design," Springer, 2009, ISBN: 978-81-8128-312-2.

[11] O. Ergin, " Circuit Techniques for Power-Aware Microprocessors," Master Thesis, The State University of New York, USA, 2003.

[12] K. Kwon, S.Chae, K.Woo, "An Application level Enegy Efficient Scheduling for Dynamic Voltage And Frequency Scaling,"IEEE International Conference on Consumer Electronics (ICCE), pages 3-6, 2013.

[13] V. Tiwari, D.Singh, S.Rajgopal, G. Mehta, R. Patel and F.Baez, "Reducing Power in High-performance Microprocessors," In Proceedings of the 35th Conference on Design Automation, ACM, USA, Pages 732-737, June 1998.

[14] R. Gonzalez and M. Horowitz, "Energy Dissipation In General Purpose Microprocessors," IEEE Journal of Solid-State Circuits, Vol.31, No.9, pages 1277-1284, September 1996.

[15] Seo E, Park S, Kim J, Lee J., "TSB:A DVS algorithm with quick response for general purpose operating systems," Journal of Systems Architecture, volume 54, pages 1-14, 2008.

[16] V.Spiliopoulos, A.Bagdia, A.Hansson, "Introducing DVFS-Management in a full-System Simulator," IEEE International Symposium on Modelling, Analysis, & simulation of Computer and Telecommunication System, pages 535-545, 2013.

[17] C.H.Hsu, U. Kremer, "Compiler – Directed Dynamic Voltage/ frequency Scheduling for Energy Reduction in Microprocessors," published in springer on Designing Embedded Processors, pages 305-323, 2007.

[18] P.Pillai, K.G.Shin, "Real-Time dynamic voltage scaling for Low-Power embedded operating systems," In Proc. the 18th ACM Symposium on Operating Systems Principles, Banff, Alberta, Canada, pages 80-102, Oct. 21-24, 2001.

[19] K. Choi, R. Soma, M. Pedram, "Fine-grained dynamic voltage and frequency scaling for precise energy and performance trade-off based on the ratio of off-chip access to on-chip computation times," In Proc. Design, Automation and Test in Europe Conference and Exhibition, Los Angeles, California, USA,Vol.1, pages 4-9, Feb. 16-20, 2004.

[20] M.E.T.Gerards, J.L.Hurink, P.K.F.Holzenspies, "Analytic Clock Frequency Selection for Global DVFS," IEEE International Conference on Parallel, Distributed, and Network-Based Processing, pages 512-519, 2014.

[21] V. Venkatachalam, M. Franz, "Power reduction techniques for microprocessor systems," ACM Computing Surveys, Volume 37, No.3, pages 195-237, September 2005.

[22] N. Tuaycharoen, "RTOS-Based Dynamic Voltage Scaling," Master Thesis University of Maryland, 2003.

Temperature Dependent Delay Analysis of SWCNT Bundle as VLSI Interconnects

Amandeep Kaur
Dept. of Electronics and Communication Engineering
Thapar University
Patiala-147004, India
amansangha28@yahoo.com

Mayank Kumar Rai
Dept. of Electronics and Communication Engineering
Thapar University
Patiala-147004, India
mkrai@thapar.edu

Abstract—**The temperature dependent circuit parameters and performance analysis in terms of delay of single walled carbon nanotube (SWCNT) bundle interconnect with different tube diameters, have been analyzed and comparisons are made between results obtained through these analyses at 22nm technology node over a temperature range from 300K to 500K. The SPICE simulation results reveal that at temperature variation ranging from 300 K to 500K, propagation delay of SWCNT bundle interconnect is increasing with rise in temperature. Simulated results further reveal that with increase in tube diameter of SWCNT propagation delay decreases up to a certain temperature range. Based on these comparative results, SWCNT bundle of optimum tube diameter has been proposed according to the operating temperature for high speed VLSI interconnects.**

Keywords—Diameter, delay, temperature, SWCNT, VLSI.

I. INTRODUCTION

As the technology is scaled down, the copper interconnects for global signals are facing serious performance degradation. This is because of increase in the resistivity of copper (Cu) due to electron surface scattering and grain boundary scattering [1]. Due to their ultra high current density(10^{10}A/cm^2) [2] and insensitivity to electromigration, defect free, single-walled carbon nanotubes (SWCNTs) have often been proposed as promising candidates for use as interconnects in integrated circuits applications [2]-[5]. The CNTs are graphene sheets rolled up in the form of cylinders having the diameter in the range of nanometers. The reason behind choosing CNT as a future VLSI interconnects is its much lower resistivity than longer mean free path(λ) as compared to Cu. Although the intrinsic resistance of a metallic CNT is very high (6.4K) [2], these can be combined parallel to form a bundle and hence overall resistance can be reduced. Other recent studies have also analyzed the potential of CNTs as circuit elements [1], [6]-[15]. These works were primarily focused on influence of length and width, on bundle interconnect performance. The observed influences are due to control of equivalent circuit impedance parameters by bundle length and width. The impedance parameters of the equivalent circuit also depend on CNT diameter in the bundle [15]. An account of how delay and power dissipation in a SWCNT-bundle interconnect is controlled by the constituent tube diameter can be found in [11].

However, due to the present trends in device scaling, thermal issues have now emerged as a major challenging factor in the possible usage of nanotubes for designing high performance integrated circuits. Integrated circuits often operate at temperatures much greater than room temperature. So, if SWCNTs are to be employed in integrated circuit applications, it is critical to understand how their electrical characteristics will vary at temperatures greater than room temperature. Therefore, it is important to investigate the effect of temperature variation on the electrical transport of CNTs, if CNTs are proposed as an alternative to the conventional metal (copper) for future interconnects. Only a few studies investigated the effect of temperature variation on the performances of CNTs [17]-[21]. Their analyses show that electron transport is limited by the phonon scattering mechanism which affects the mean free path and hence the resistance of SWCNT. The temperatures variations have the significant effect on interconnect resistance which will affect the performance of CNT bundle in terms of delay.

In this paper, a complete temperature dependent impedance and performance analysis in terms of delay, for SWCNT bundle interconnects over a temperature range from 300 to 500K, have been presented. In addition, the performance SWCNT bundle has also been compared taking different tube diameters over a temperature range from 300 to 500K. Results obtained through simulation shows that the temperature variation in high speed ICs play an important role to optimize the interconnect performance in terms of propagation delay.

II. TEMPERATURE DEPENDENT IMPEDANCE PARAMETERS OF INTERCONNECTS

As the high performance ICs have great variation in temperature, it is necessary to include factor of temperature in the calculation of resistance of interconnect. The resistance of a CNT is dependent on the electron mean free path which is further dependent upon acoustic phonon and optical phonon scattering [20], where scattering is a function of temperature which affects the mean free path (MFP) as the temperature rises. The electron–electron scattering mechanism is negligible in CNT and the main source of scattering in CNT is the electron–phonon scattering [22]. Acoustic, optical emission and absorption scattering phenomena are the different types of electron–phonon scattering that present in metallic SWCNTs. Therefore, combining the effect of all type of scattering the effective mean free path (λ_{eff}) can be written as [18]

$$\lambda_{eff} = \{\lambda_{AC}{}^{-1} + \lambda_{OP}{}^{-1} + \lambda_{OP,abs}{}^{-1}\} \tag{1}$$

where λ_{AC} is the acoustic MFP dependent upon elastic electron scattering with acoustic phonons and λ_{OP} is the effective optical MFP due to inelastic scattering caused by optical phonon emission and $\lambda_{OP,abs}$ is the effective optical MFP due to scattering caused by absorption of phonon . Acoustic phonon scattering is the main origin of the ohmic resistance at low-bias conditions in metallic SWCNTs. λ_{AC} is proven to be directly dependable on both temperature and diameter. Therefore, λ_{AC} can be written as:

$$\lambda_{AC}(d,T) = \frac{\lambda_{AC,300}}{d_0}\frac{300\,d}{T} \tag{2}$$

where $\lambda_{AC,300}$=1600 nm is the reported acoustic MFP for tube diameter d_0 = 1.8 nm at 300 K [18]. In case of high-bias conditions, the electric field that is generated along the CNT (E = V/L) accelerates the electrons and increases their kinetic energy. When the electron energy reaches the optical phonon energy level, a phonon will be emitted with energy *h* (*h* ~ 0.16-0.2 eV) [18] .where as $\lambda_{OP,abs}$ is given by:

$$\lambda_{OP,abs} = \frac{\lambda_{OP,300}d}{d_0}\frac{N_{OP}(300)+1}{N_{OP}(T)+1} \tag{3}$$

$\lambda_{OP,300}$ = 15 nm is the measured spontaneous effective emission length for diameter d_0 at 300 K [17]. N_{OP} is the optical phonon occupation and is given by the equation:

$$N_{OP}(T) = 1/\exp\left({}^{h\Omega}/_{K_B T} - 1\right) \tag{4}$$

where K_BT is the thermal energy (K_B is the Boltzmann's constant). In addition, optical scattering can occur when an electron acquires the required energy by absorbing another optical phonon, respectively or after electrons gain sufficient energy from electric field (fld) [20], [22]. Therefore, $_{OP}$ can be written as:

$$\lambda_{OP,ems}(d,T) = \left(1/\lambda^{fld}_{OP,ems} + 1/\lambda^{abs}_{OP,ems}\right)^{-1} \tag{5}$$

where $\lambda^{fld}_{OP,ems}$ is the MFP due to the electric-field acceleration based scattering. $\lambda^{fld}_{OP,ems}$ can be written as:

$$\lambda^{fld}_{OP,ems}(d,T) = \frac{(h\Omega - K_BT)L}{qV} + \frac{\lambda_{OP,300}d}{d_0}\frac{N_{OP}(300)+1}{N_{OP}(T)+1} \tag{6}$$

The first term represents the distance that the electron needs to accelerate and gain the enough energy to emit a phonon, while the second term represents the distance that the electron travels after gaining the energy before emitting the phonon.

$\lambda^{abs}_{OP,ems}(d,T)$ measures the scattering effect due to absorbing an optical phonon and can be written as:

$$\lambda^{abs}_{OP,ems}(d,T) = \frac{\lambda_{Oz,300}d.(N_{Oz}(300)+1)}{d_0.N_{Oz}(T)} \tag{7}$$

The existence of the hot phonon phenomenon that is caused by severe self heating is not considered in this analysis. In the present study, low bias condition is considered, where the generated electric field is not large enough to create the hot phonons [20].

A. SWCNT Bundle:

The total resistance of SWCNT bundle [13] can be defined using temperature dependent resistance of individual SWCNT. Where, the resistance of an individual SWCNT with N conducting channel can be formulated as [20]

$$R_{CNT}(T) = R_C + \frac{h}{2N(d,T)e^2} \qquad (\text{for } L < \lambda) \tag{8}$$

$$R_{CNT}(T) = R_C + \frac{h}{2N(d,T)e^2}\left[\frac{L}{\lambda_{eff}(T)}\right] \qquad (\text{for } L > \lambda) \tag{9}$$

Where R_C is the contact resistance, *h* is the Planck's constant, e is the electron charge, $\lambda_{eff}(T)$ is the temperature dependent effective electron mean free path (MFP), which is the average distance between successive scatterings and formulated by "(1)". L is the length of interconnect and N is the number of available conducting channels for SWCNT. Here, for the sake of proper calculation of SWCNT resistance, R_C value is assumed to be 24 kilo-ohms [20].

At room temperature, many sub bands of SWCNTs with large diameters can extend to Fermi level to contribute to the electrical conductance by creating channels for electrons. Number of available conducting channels for SWCNTs with a diameter d at a temperature T is given by [23]:

$$N(d) \simeq \begin{cases} 2 & \text{for } d < \frac{d_T}{T} \\ a.d.T + b & \text{for } d > \frac{d_T}{T} \end{cases} \tag{10}$$

Where a=6.12*10^{-4}/nmK; b=1.275; d_T = 1300 nmK. The resistance of a SWCNT is given by "(8)" and "(9)". SWCNTs are usually connected parallel to decrease the resistance as shown in Fig. 3 and the number of CNTs in a bundle is given by "(11)" [13]:

$$n_{CNT} = \begin{cases} n_w n_h - (n_h/2) & \text{if } n_h \text{ is even} \\ n_w n_h - \left(\frac{n_h - 1}{2}\right) & \text{if } n_h \text{ is odd} \end{cases} \tag{11}$$

Where $n_w = \left\lfloor \left\{\frac{w-d}{x}\right\}\right\rfloor$ is the number of "columns", $n_h = \left\lfloor \frac{h-d}{(\sqrt{3}/2)x} \right\rfloor + 1$ is the number of "rows" in the interconnect bundle, w is the width of the bundle, d is the tube diameter, x is the separation distance between two SWCNT, h is the height or thickness of the bundle and n_{CNT}

is the total number of CNTs. Therefore, the total bundle resistance is formulated as:

$$R_{SWCNT\ bundle}(T) = R_{CNT}(T)/n_{CNT} \qquad (12)$$

The other impedance parameters (viz. capacitance and inductance) of SWCNT bundle are calculated by the appropriate expression available in the literature [12], [13].

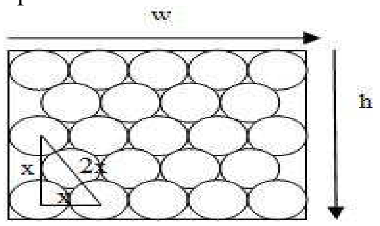


Fig. 1 Dense bundle of metallic SWCNTs [13].

TABLE I. Simulation parameters used for the calculations [12], [24]

Parameter	Value
V_{dd}	0.7V
Width (W) of global interconnect	32 nm
A/R(aspect ratio) for global interconnect	3
Thickness (H) of global interconnect	96nm
Separation(s) between adjacent bundle of global interconnect	32nm
Oxide thickness(t_{ox})	76.8nm
$_{OX}$(relative dielectric constant)	2.05

TABLE II. Resistance, Inductance and Capacitance values for SWCNT Bundle with different tube diameters (d_{SWCNT}=0.7- 1.0nm) for 22nm technology at 300K.

SWCNT Tube Diameter (nm)	Inductance (pH)	Capacitance (pF)	Resistance (K)
0.7	0.14	2.94	1.75
0.8	0.17	2.65	1.98
0.9	0.22	2.42	2.27
1.0	0.26	2.22	2.51
1.1	0.31	2.06	2.78

III. IMPEDANCE ANALYSIS

The impedance parameters of SWCNT bundle for different tube diameters (i.e. d_{SWCNT} = 0.7nm – 1.1nm) are calculated from “(1)” –“(12)”, over a temperature range from 300-500 K for low bias conditions [18]. These calculations are done for global length (L= 1000μm) at 22nm technology. Data given in table 1 are used for calculations [12], [24]. Fig.2 shows the dependence of temperature and tube diameter on interconnect resistance. It has been observed that the resistance increases with rise in interconnect temperature. It has been also observed that SWCNT tube with larger diameter has larger resistance. Table 2 also shows that the capacitance of SWCNT bundle is much higher than the capacitance of SWCNT bundle and the bundle capacitance is decreasing with increase in tube diameter. Therefore, these variations in temperature dependent impedance parameters have the significant effect on delay.

IV. PERFORMANCE ANALYSIS

The circuit considered for analysis consists of a CMOS inverter driving the distributed CNT interconnect bundle having a load capacitance of 1pF as shown in Fig.3. The frequency of input pulse given to driver is 0.1GHz, having rise time of 1ns. The performance of the setup is studied by SPICE simulation. The Predictive Technology Model [24] for 22nm is used for CMOS driver. Distributed RLC model is used for the simulation with optimized number of repeaters. Fig. 4 shows the variation of 90% propagation delay of SWCNT bundle with different tube diameter (d_{SWCNT}= 0.7nm -1.0nm) with rise in temperature. It is observed that propagation delay increases in all cases with rise in interconnect temperature due to dominance of resistance (see Fig.2). Results further reveal that at room temperature SWCNT bundle with larger tube diameter has the least propagation delay but as temperature rises, after a certain temperature, trend changes and the SWCNT bundle with smaller diameter has least propagation delay due to the dominance of resistance increase over capacitance at higher temperatures.

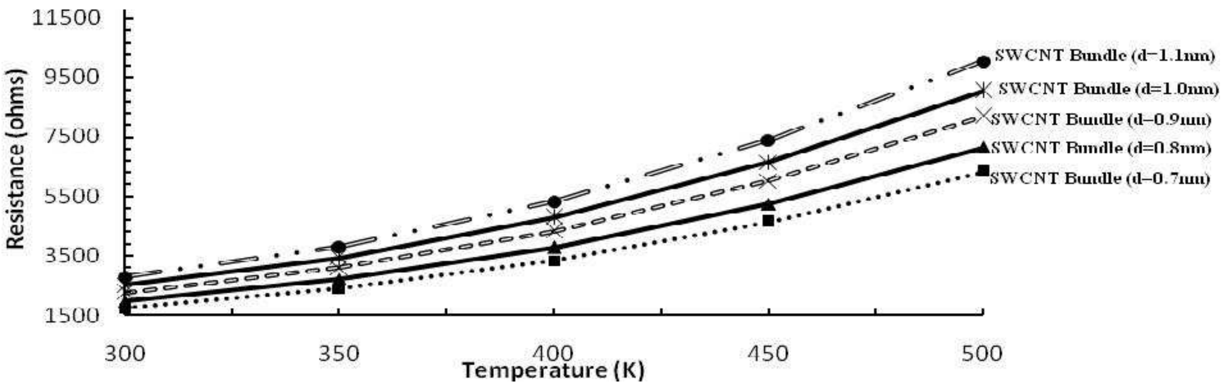


Fig. 2 Variation of resistance of SWCNT bundle with tube diameter (d_{SWCNT}=0.7nm – 1.1nm) with respect to temperature at 22nm technology.

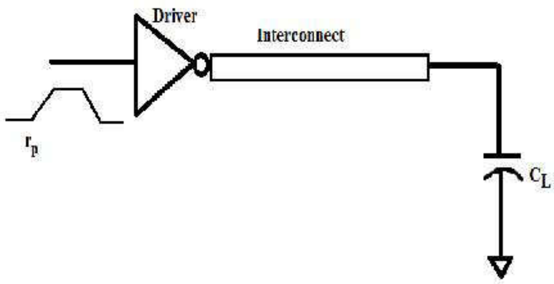


Figure 3 Interconnect with a load C_L derived by a CMOS inverter with an input r_p.

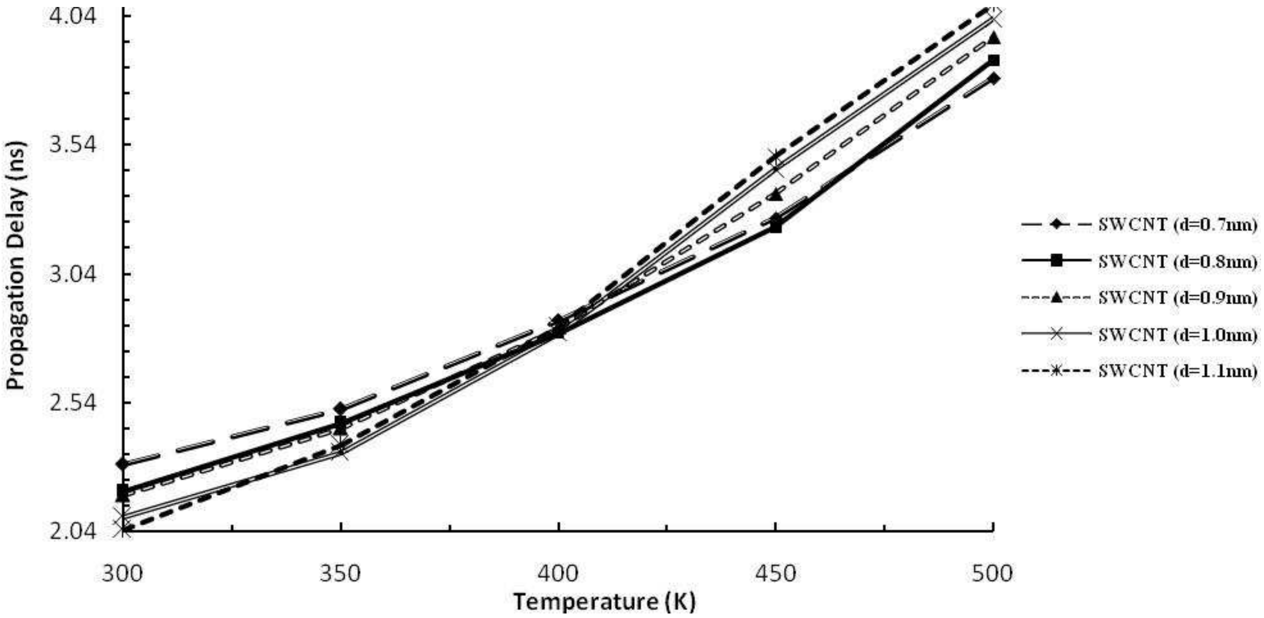


Fig. 4 Variation of propagation delay of SWCNT bundle with different tube diameters (d_{SWVNT}=0.7-1.1nm) with respect to variation in temperature at 22nm technology.

V. CONCLUSION

The influence of temperature variation on propagation delay of global interconnect of SWCNT bundle with different tube diameters has been investigated. It has been seen that the SWCNT bundle with larger tube diameter gives better performance in terms of propagation delay than SWCNT bundle with lower tube diameter up to a certain temperature range, after which trend changes. This is due to the dominance of rise in resistance associated with SWCNT bundle with rise in temperature over a constant capacitance value of SWCNT bundle. Results also reveal that the optimum value of tube diameter for SWCNT bundle should be chosen according to operating temperature range.

REFERENCES

[1]. W. Steinhogl et al. "Comprehensive study of the resistivity of copper wires and lateral dimensions of 100nm and smaller", Journal of Applied Physics, Vol. 97, 023706, 2005.

[2]. B. Q. Wei, R. Vajtai, and P. M. Ajayan, "Reliability and Current Carrying Capacity of Carbon Nanotubes," Applied Physics Letters, Vol.79, No.8, 2001, pp. 1172-1174,

[3]. F. Kreupl, A. P. Graham, M. Liebau, G. S. Duesberg, R. Seidel, andE. Unger, "Carbon nanotubes for interconnect applications," in IED MTech. Dig., 2004, pp. 683–686.

[4]. Paul L. McEuen, Michael S. Fuhrer, and Hongkun Park., "Single-Walled Carbon Nanotube Electronics; IEEE Transactions On Nanotechnology: 1, no.1, 2002, pp:78-85.

[5]. S. Frank, P. Poncharal, Z. L. Wang, W. A. de Heer. "Carbon Nanotube Quantum Resistors"; Science: 280.5370,1998, pp.1746-1774.

[6]. Azad Naeemi , Reza Sarvari and James D. Meindl., "Performance Comparison Between Carbon Nanotube and Copper Interconnects for Gigascale Integration(GSI)"; IEEE Electron Device Letters: 26, no.2: 2005, pp. 84-86,

[7]. A. Raychowdhury and K. Roy. "A Circuit Model for Carbon Nanotube Interconnects: Comparative Study with Cu Interconnects for Scaled Technologies"; ICCAD: 2004, pp. 237 -240.

[8]. W. Liang et al., "Fabri-parrot interference in a nano electron waveguide", Nature: 411: 2001, pp. 665- 669.

[9]. A.Naeemi and J.D.Meindl. "Impact of electron-phonon scattering on the performance of carbon nanotube interconnects for GSI"; IEEE Electron Device Lett.: 26, no.7, 2005, pp. 476-478,

[10]. M. Nihei, D. Kondo, A. Kawabata, S. Sato, H. Shioya, M. Sakaue, T. Iwai, M. Ohfuti, and Y. Awano. "Low-resistance multi-walled carbon nanotube vias with parallel channel conduction of inner shells"; In Proc. IEEE Int. Interconnect Technology Conf,2005,pp. 251–253.

[11]. Arthur Nieuwoudt and Yehia Massoud. "Predicting the Performance of Low-Loss On-Chip Inductors Realized Using Carbon Nanotube Bundles"; IEEE Transactions On Electron Devices: 55, no.1: pp. 298-312, 2008.

[12]. Mayank Kumar Rai and Sankar Sarkar, "Influence of tube diameter on Carbon nanotube interconnect delay and power output, Physica Satus Solidi A:208, no.3:pp. 735-739, 2011.

[13]. Navin Srivastava and Kaustav Banerjee. "Performance Analysis of Carbon Nanotube Interconnects for VLSI Applications"; ICCAD: pp. 383-390, 2005.

[14]. P. J. Burke. "Luttinger Liquid Theory as a Model of the Gigahertz Electrical Properties of Carbon Nanotubes"; IEEE Trans. Nanotechnology:1, no.3: pp. 129-144, 2002.

[15]. Mayank Kumar Rai and Sankar Sarkar, "Influence of distance between adjacent tubes on SWCNT bundle interconnect delay and power dissipation", Journal of Computational Electronics; DOI 10.1007/s10825-013-0490-7, 2013.

[16]. Massoud, Y., Nieuwoudt, A.: "Modeling and design challenges and solutions for carbon nanotube-based interconnect in future highperformance integrated circuits". ACMJ. Emerg. Technol. Comput. Syst. **2**(3), pp. 155–196, 2006.

[17]. Nieuwoudt, A., Massoud, Y.: "On the optimal design, performance, and reliability of future carbon nanotube-based interconnectsolutions". IEEE Trans. Electron Devices **55**(8), pp. 2097–2110 , 2008.

[18]. E. Pop, D. Mann, J. Reifenberg, K. Goodson, H. Dai, "Electro-Thermal transport in single-wall carbon nanotubes for interconnect applications", in: Electron Devices Meeting, IEDM Technical Digest, IEEE International, 2005, pp. 254–256, 2005.

[19]. E. Pop, D.A. Mann, K.E. Goodson, H. Dai, "Electrical and thermal transport in metallic single-wall carbon nanotubes on insulating substrates," Journal of Applied Physics, Vol. 101 (9) p. 093710 May 2007.

[20]. A. Hosseini, V. Shabro,"Thermally-aware modeling and performance evaluation for single-walled carbon nanotube-based interconnects for future high performance integrated circuits," Microelectronic Engineering 87, pp.1955–1962 2010.

[21]. E. Pop, D. Mann, J. Cao, Q. Wang, K. E. Goodson, and H. J. Dai, "Negative differential conductance and hot phonons in suspended nanotube molecular wires," Phys. Rev. Lett., vol. 95, p. 155505, 2005.

[22]. Ashok Srivastava et al. "A Thermal Model for Carbon Nanotube Interconnects", Nanomaterials, 3, pp. 229-241, 2013.

[23]. J.-Y. Park, S. Rosenblatt, Y. Yaish, V. Sazonova, H. stnel, S. Braig, T. Arias, P.Brouwer, P. McEuen, "Electron-Phonon Scattering in metallic Single-Walled Carbon Nanotubes", Nano Letters Vol.4 No.3 pp. 517–520, 2004.

Automated Temperature Control System using LabVIEW®and GSM Modem

Mukesh Kumar
Dept. of Electronics and Electrical Engineering
Lovely Professional University,
Punjab, India
zmukeshz@yahoo.co.in

Parveen Kumar
Department of Electronics and Communication Engineering, Punjab Technical University,
Punjab, India
parveen.eng@gmail.com

Rashmi
Department of Electronics and Communication Engineering, CT group of institutions, Shahpur, Jalandhar
Er.rashmi@ymail.com

Gopal Sharma
Department of Electronics and communication engg, CT Institute of Technology,
Shahpur, Jalandhar
Gopalenggr@gmail.com

***Abstract*- The objective of this paper is to design and develop simple low cost microcontroller based temperature monitoring system using GSM technique and LabVIEW software package, which gives the graphical values of received input data on the front panel. It provides wireless communication link to monitor and control equipments that are far away from the user. The temperature monitoring system undergoes four stages signal conditioning circuit, analog to digital converter, LabVIEW and with GSM Module the message is send to mobile. The system consists of a temperature sensor (LM35), a high-performance low-power Phillips 8-bit microcontroller (P89V51RD2) , GSM module (SIM300), ADC0804 and a computer terminal for storing and displaying the data. The LM35 sensor is connected to Phillips microcontroller through ADC and varying temperature is sent to GSM modem, which is simultaneously displayed in LCD and LabVIEW. ADC is used because microcontroller works with digital inputs. GSM modem can be used to send and receive SMS through AT commands. At the transmitter side, the user sends an SMS to the GSM modem using AT commands. The LM35 is an integrated circuit sensor that can be used to measure temperature with electrical output proportional to the temperature. A user friendly graphical user interface (GUI) was developed using the LabVIEW software package.**

Keywords — GSM module, LabVIEW, Sensor, LM35, microcontroller.

I. INTRODUCTION

The objective of this paper is to design and develop a wireless communication link to monitor and control equipments that are far away from the user. In an industry during certain hazards it will be very difficult to monitor the parameter through wires and analog devices such as transducers. It is a user-friendly application system. GSM plays a key role for transmitting and receiving the data from the user. The main advantage of this concept is to stoppage and some accidents caused by increased temperature can be avoided. The self designed GSM module is selected to finish the transmission and decoding of the data through "AT "command and coding of microcontroller [3]. Here, the temperature is monitored directly from LM35 which is simultaneously displayed in the LCD as well as LabVIEW and can be sent as a message by GSM technique. A typical temperature monitoring system using GSM and LabVIEW is shown in Fig. 1.

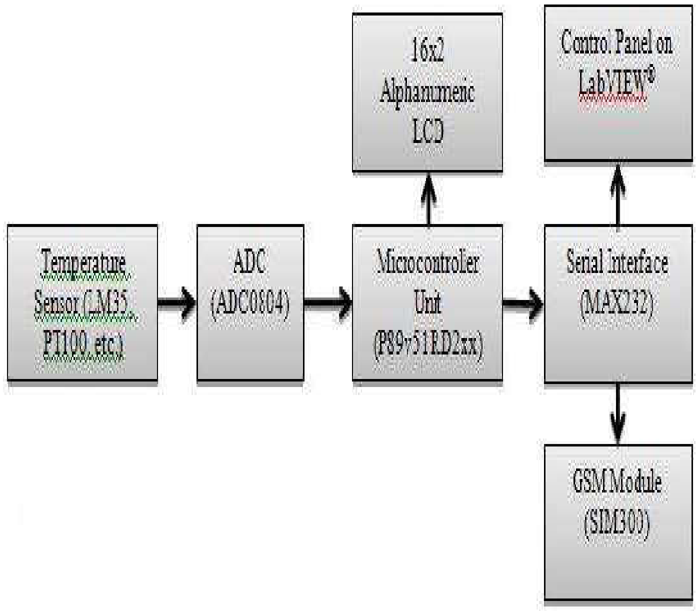


Fig. 1 Block Diagram of the Temperature Monitoring System

I. SOFTWARE DEVELOPMENT

Software design includes the development of the algorithm for the system, the writing of the separate routines for the different interfacing devices, and the test of the routines on the designed hardware. The interface between the microcontroller with the ADC, LCD, GSM module and PC was achieved using software modules. The software module is written in Embedded C in Keil4. Keil4 is an integrated toolset for the development of embedded application employing Atmel and Phillips microcontroller. The compile tool is used to convert embedded C Language to HEX File. The HEX file is stored into the Phillips microcontroller using ECEFlash. The flowchart of the control program is shown in Figure-3

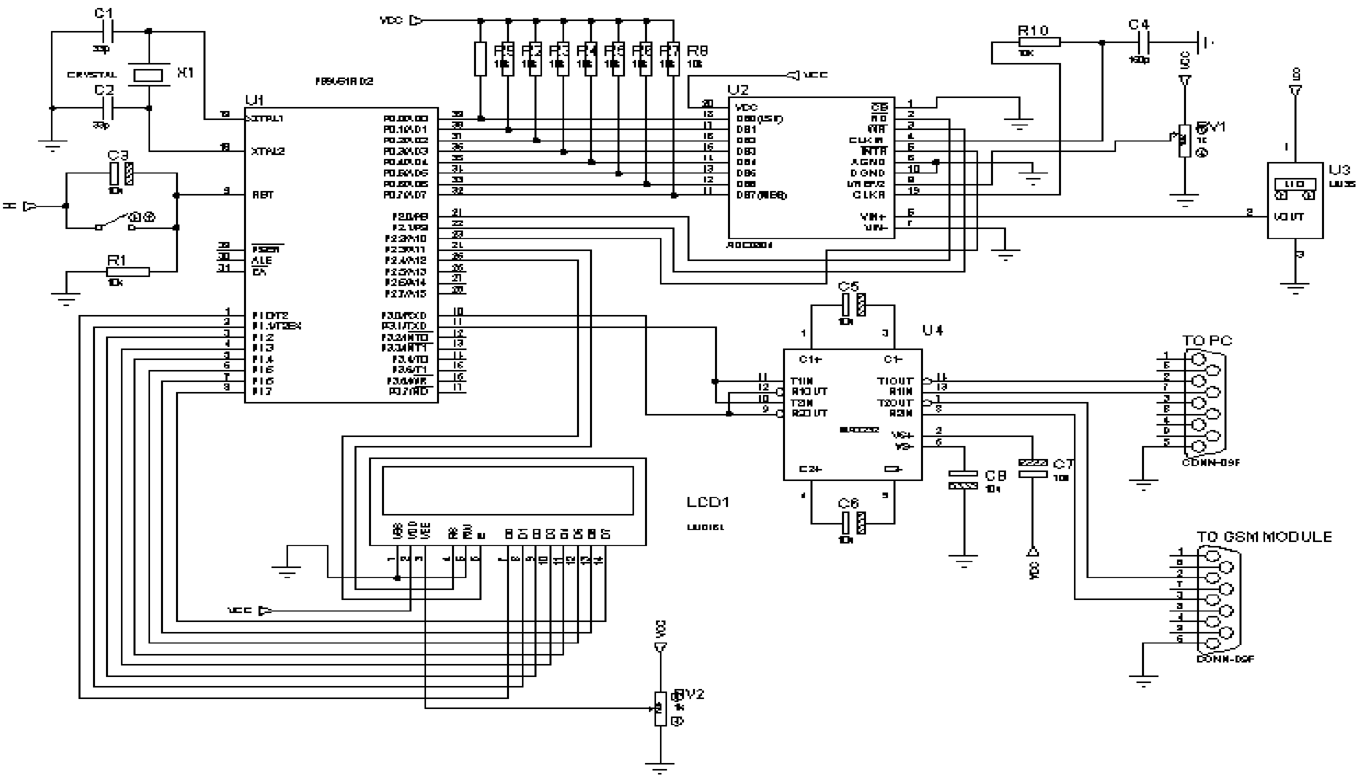


Fig. 2 Schematic of the Temperature Monitoring System

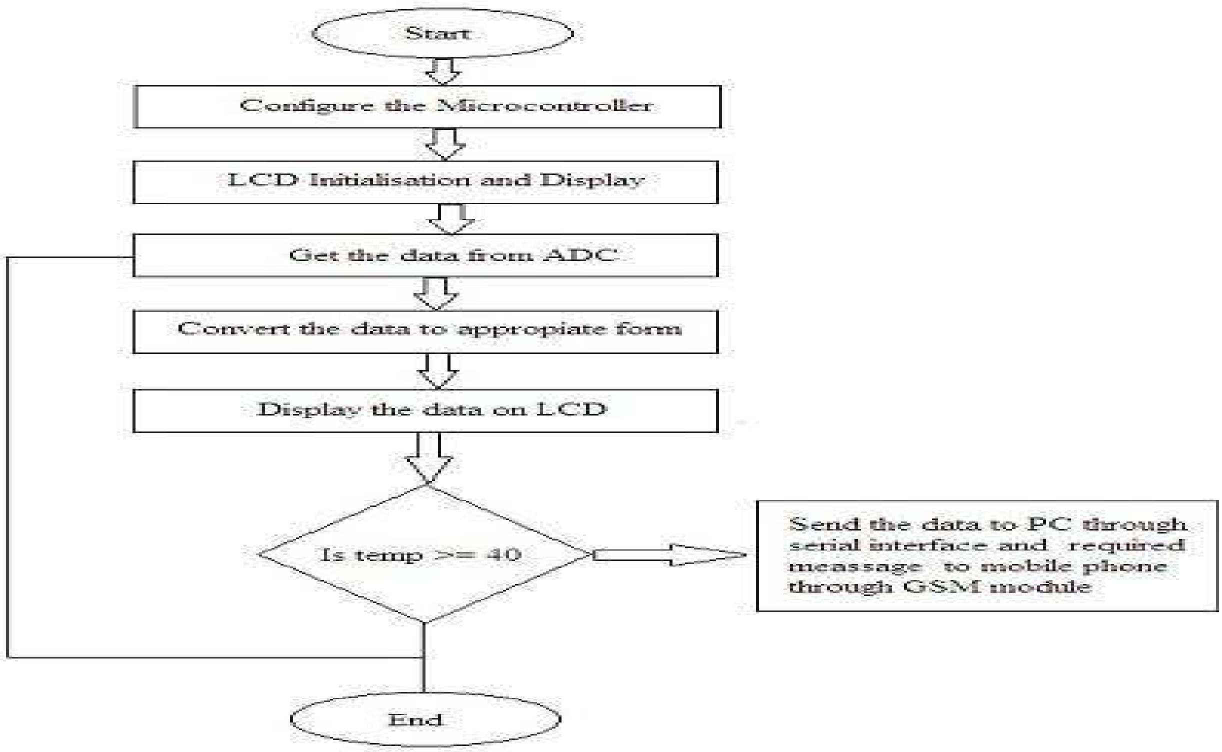


Fig. 3 Schematic of the Temperature Monitoring System

On the other side, a controlled graphical user interface application software was developed using a suitable language, such as the LabVIEW™ software developed by National Instruments Corporation. Moreover, this is a dataflow and graphics-based language and is appropriate for the design of a man-machine interface [7]. Figure-4 shows the block diagram that contains the graphical source code of a LabVIEW program, and Figure-5 illustrates the front panel window used as the user interface for the virtual instrument (VI).

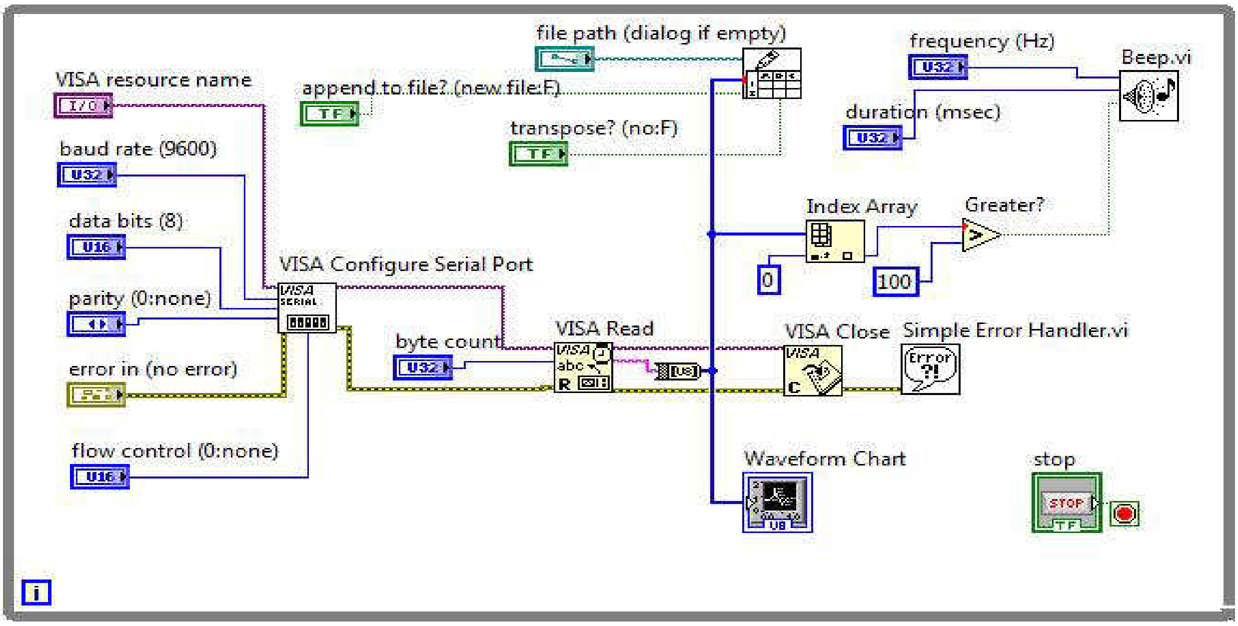


Fig. 4 Block diagram for Temperature Control Panel on LabVIEW®

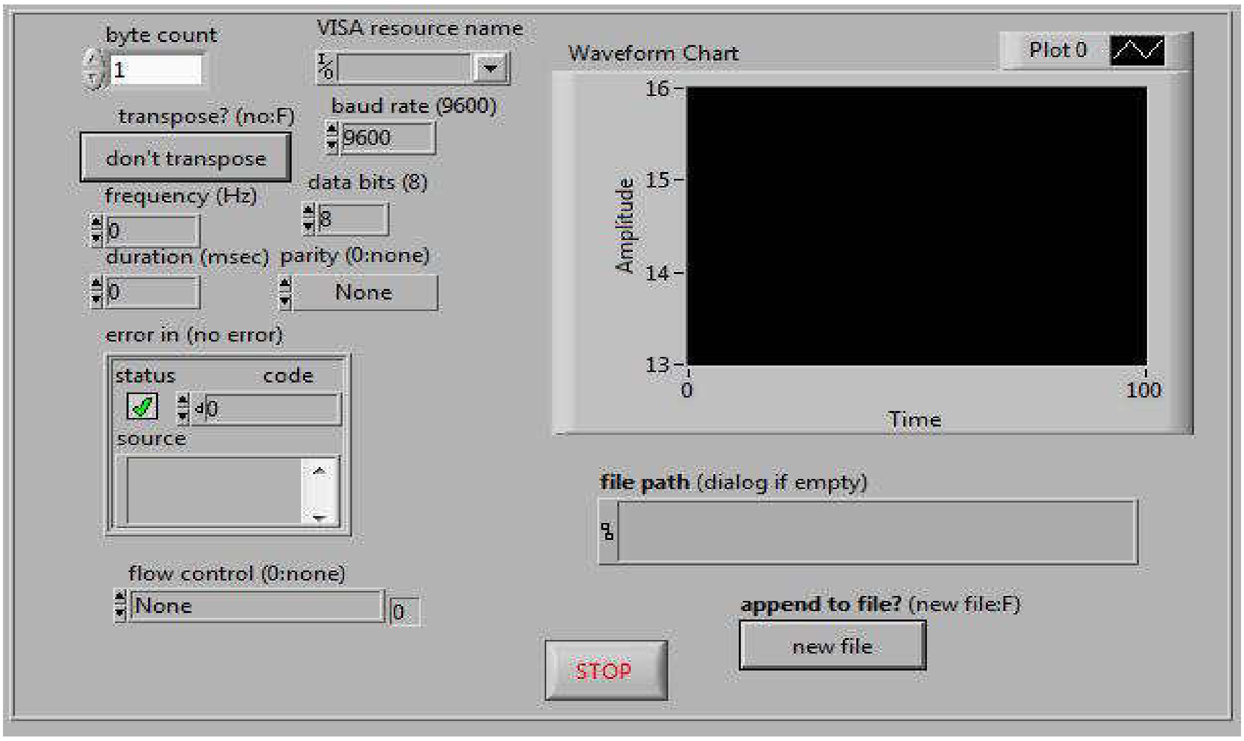


Fig. 5 Front Panel window of LabVIEW

In this overall system, LabVIEW interacts with the end-users in two different forms: (1) block diagrams, which are where the control function and the data flow are designed, and (2) front panel, in which the buttons, counters, data, and graphs are displayed and accessed through direct cursor manual control. The front panel has controls and indicators, which are the interactive input and output terminals, respectively, of the VI. The concept of the block diagram is to separate the graphical source code from the user interface in a logical and simple manner. The front panel objects appear as terminals on the block diagram. A group of command buttons were added to the front panel for user access, such as the stop and serial port number.

II. Conclusion

The temperature sensor (LM35) is connected to microcontroller and varying temperatures is sent to GSM modem to send SMS which is simultaneously displayed on LCD, LabVIEW's waveform chart and also logged into a measurement file for future references.Implemented system is successfully recoding the temperature variations and continuously verifying it with predefined threshold to provide us the alarming messages to control the undesired situations.

References

[1] Md. Abdullah Al Mamun, K. Sundaraj, N. Ahmed, Matiur Rahman, Mijanur Rahman and Nizam Uddin Ahamed "Design And Development Of A Pc-Based Automated Data Logging System For Measuring Temperature," in ARPN Journal of Engineering and Applied Sciences, Vol. 8, No. 11, November 2013.

[2] Dr.B.Ramamurthy, S.Bhargavi and Dr.R.ShashiKumar, "Development of a Low-Cost GSM SMS-Based Humidity Remote Monitoring and Control system for Industrial Applications," IJACSA,Vol. 1, No. 4, October 2010.

[3] T. Murugan, Azha.Periasamy and S. Muruganand, "Embedded based Industrial Temperature Monitoring Systems using GSM," International Journal of Computer Applications, Volume 58– No.19, November 2012.

[4] BaburaoKodavati, V.K.Raju, S.Srinivasa Rao, A.V.Prabu, T.Appa Rao, Dr.Y.V.Narayana "GSM and GPS based vehicle location and tracking system" IJERA Vol. 1 Issue 3, pp.616-625.

[5] Malik Sikandar Hayat Khiyal, Aihab Khan, and ErumShehzadi, "SMS Based Wireless Home Appliance Control System (HACS) for Automating Appliances and Security" Informing Science and Information Technology Volume 6, 2009.

[6] S. bin Ramli and N. btFisal. 2008. Design of Wireless Sensor Network for Environmental Monitoring (WiSNEM). UniversitiTeknologi Malaysia: Bachelor Degree Thesis.

[7] T. Jen-Hao, C. Shun-Yu, L. Jin-Chang and R. Lee. 2000. A LabVIEW based virtual instrument for power analyzers. In: Power Con 2000 International Conference on Power System Technology.1: 179-184.

[8] Goswami, T. Bezboruah and K.C. Sarma, "Design of an Embedded System for Monitoring and Controlling Temperature and Light", International Journal of Electronic Engineering ResearchVolume 1 Number 1 (2009) pp. 27–36.

[9] S.Josephineselvarani "online health monitoring systems using zigbee" IJCSE vol.3 No.4 Apr 2011.

[10] Syamkrishna, j.ravindra" Design And Implementation Of Remote Home Security System Based On Wsns And Gsm Technology" IJESAT Jan-Feb 2012.

[11] Amit Sachan, "Microcontroller Based Substation Monitoring and Control System with GSM Modem" IOSR Journal of Electrical and Electronics Engineering 2012.

[12] Amit Sachan, "GSM based SCADA monitoring and control system substation equipment".International Journal of Engineering Research & Technology 2012.

Influence of Fermi-Energy on Performance of Multi-Layer Graphene Nanoribbon Interconnects

Ankit Gupta
Department of Electronics and Comm. Engineering
Thapar University
Patiala, India
ankit.gupta1219@gmail.com

Mayank Kumar Rai
Department of Electronics and Comm. Engineering
Thapar University
Patiala, India
mkrai@thapar.edu

***Abstract*—An equivalent RLC model of multi-layer graphene nanoribbon interconnect is presented. Power dissipation of Multi-layer graphene Nanoribbon (MLGNR) interconnectis analyzed at 22nm technology node.Similar analysis is carried out for copper interconnect and results are compared with MLGNR for global length interconnects.SPICE simulation results reveal that power dissipation decreases with increase in Fermi-energy. MLGNR interconnect resistance and inductance increases with decrease in Fermi energy and increase in length. With the increase in Fermi energy power dissipation can be reduced. Theperformance can be improved by using higherFermi energy.**

***Keywords*—GNR, MLGNR, copper, power, length,fermi energy, VLSI.**

I. INTRODUCTION

With technology scaling, resistivity of copper interconnects is increasing with a high rate due to the combined effect of grain boundary and surface scattering [1],[13]. Graphene Nanoribbon is considered as the attractive candidates for future VLSI interconnects due to its capability to conduct high current densities, high thermal conductivities and large mean free paths [2]-[3]. Graphene nanoribbons can be formed from a mono layer of graphene sheet which consist of carbon atoms packed in a 2D honey comb lattice structure by patterning graphene into a thin strip [3].

GNRs can be classified depending on the shape of their edge (armchair or zigzag). Zigzag GNR's are always metallic and the armchair GNR's can be metallic when R=3p+2 or semiconducting when R=3p or 3p+1, where R is the number of carbon rings along its width and p is an integer[3]. Due to the planar nature of graphene, GNR has more straight forward fabrication processes and control over chirality [4] using high resolution lithography[5]. GNR's can be categorized as single-layer GNR (SLGNR) and multi-layer GNR (MLGNR). Due to the lower conductance of SLGNR, MLGNR is preferred for interconnect applications.

Previously, Conductance has been modeled using the simple tight binding model and linear response landauer formula [6],[7]. Resistance and Delay of doped GNR has been compared with other interconnect materials in [6],[8]. Compact physics based circuit models are being developed and delay associated to side contact and top contact MLGNR are compared with copper in [8],[9]. Stability of the GNR has been analyzed in [10] and crosstalk analysis is done for GNR interconnect and compared with copper and multiwall CNT in [11].

In this paper, we present a complete impedance analysis and power dissipation analysis of MLGNR interconnects with respect to the variation in lengthand results are compared with copper.The effect of Fermi energy on power dissipation has also been compared. The organization of this paper is as follows. Section Iintroduces briefs about the structures and properties of GNRs. In Section II the equivalent RLCmodel of Multi-layer GNR interconnects is presented. Impedance analysis is done in section III. In section IVPower is analyzed for local, intermediate and global lengths and compared with that of copper at 22 nm technology node andSection V concludes this work.

II. EQUIVALENT CIRCUIT OF THE MLGNR

The geometry of MLGNR is shown in Fig.1, where H is the thickness, w is the width of the MLGNR, y is the distance from ground plane with $_r$ as the dielectric constant of its medium, is the distance (0.34nm) between the adjacent layers and N is number of layers.

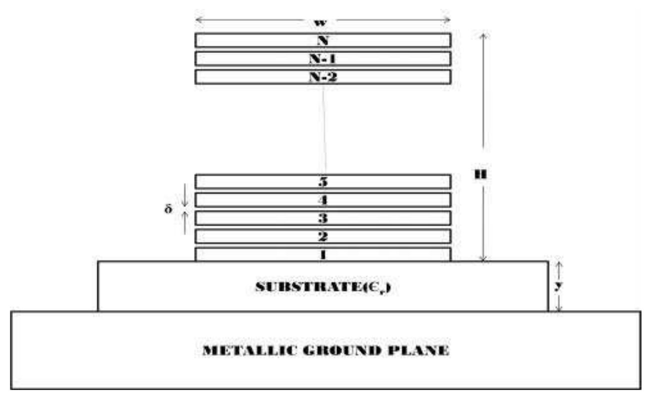


Fig.1.Geometry of MLGNR interconnect

The number of layers can be formulated as [11]

$$N = 1 + integer\left[\frac{H}{\delta}\right] \quad (1)$$

Its RLC model is shown in Fig.2. R_{mc} represents the imperfect contact resistancewith its typical value of 20k [12].

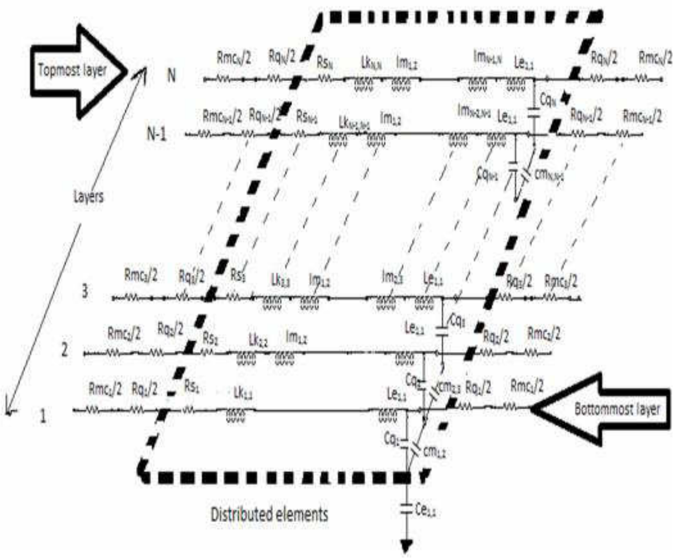


Fig.2. RLC Model for MLGNR interconnect

$R_q$is the quantumcontact resistance and is defined as [8]

$$R_q = \frac{h/2e^2}{N_{ch}} = \frac{12.9k\Omega}{Nch} \ (2),$$

Where $N_{ch}$is the number of conducting channels in each layer, h is the Plank's constant and e is the electronic charge.

N_{ch} is computed through the occupation probability of states corresponding to the peak and valley of the valence and the conduction subbands and is given by [10]

$$N_{ch} = \sum_{i=1}^{N_c}[1 + e^{(E_i - E_F)/K_T}]^{-1} + \sum_{i=1}^{N_v}[1 + e^{(E_F + E_i)/K_T}]^{-1} \ (3),$$

Where E_F is the Fermi energy, k is the Boltzmann constant, T is the temperature, $v_F$is the Fermi velocity, N_c is the density of states of electrons and N_v is the density of states of holes. R_s is the scattering resistance along its length L and is given by [8]

$$R_s = \frac{R_q}{\lambda}\Big/\mu m \quad (4)$$

Hence, the total resistance of the multi-layer GNR is given by

$$R_{eq} = \frac{R_{mc} + R_q + R_s}{N} \quad (5)$$

The inductance can be of two types, kinetic inductance(L_k) and magnetic inductance (L_e)[12].

$$L_K = \frac{h/4e^2\,\vartheta_F}{N_{ch}} \cong \frac{8nH}{N_{ch}}/\mu m \quad (6)$$

$$L_e = \frac{\mu_0 y}{w}\Big/\mu m \quad (7)$$

The distributed capacitance of GNRs contains both Quantum capacitance (C_q) and electrostatic capacitance (C_e). The quantum capacitance (C_q) per layer is given by[12]

$$C_q = \frac{4e^2/h\,.\,N_{ch}}{\vartheta_F} \cong N_{ch} * 0.2fF\Big/\mu m \quad (8)$$

The electrostatic capacitance (C_e) is calculated by using conformal mapping method and is given by [8].

$$C_E = \in M\left[\tanh\left(\frac{\pi w}{4y}\right)\right] \quad (9)$$

$$M(a) = \begin{cases} \dfrac{2\pi}{ln\frac{2.(1+\sqrt[4]{1-a^2})}{(1-\sqrt[4]{1-a^2})}}, & 0 \le a < \dfrac{1}{\sqrt{2}} \\ \dfrac{2}{\pi}\ln\dfrac{2(1+\sqrt{a})}{(1-\sqrt{a})}, & \dfrac{1}{\sqrt{2}} \le a \le 1 \end{cases} \quad (10)$$

Mutual inductance(l_m) and mutual capacitance (c_m) are exist between each pair of layers [12]

$$l_m = \frac{\mu_{o\delta}}{w}\Big/\mu m \quad (11)$$

$$c_m = \frac{\varepsilon w}{\delta}\Big/\mu m \quad (12)$$

III. IMPEDANCE ANALYSIS

The per unit length (p.u.l) Impedance parameters of MLGNR are calculated from Eqs.(1)-(12).All the physical dimensions of MLGNR and copper and taken from[14],[15] for 22nm technology node and line parameters of copper are calculated using [16]. Using the RLC model in Fig.2, we have calculated the total equivalent R,L and C for different lengths and Fermi energies (Ef). Mean free path is assumed to be 1µm for this analysis[8].Fig.3, illustrates a typical example of the dependence of MLGNR interconnect resistance on length (L) and Fermi energy (Ef) for global interconnects respectively. It shows that resistance increases with increase in interconnect length for both MLGNR and copper interconnects. Copper interconnects are of a higher resistance than MLGNR interconnects. Resistance of MLGNR with smaller fermi energies (Ef) is observed more.

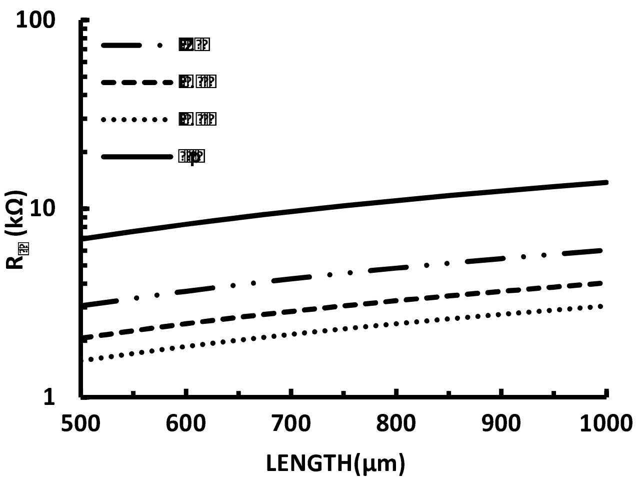


Fig.3. Equivalent resistance (R_{eq}) as a function of Fermi energy (Ef) for global lengths of multilayer GNR interconnect.

Fig.4 shows the dependence of MLGNR interconnect inductance(L_{eq}) on both length (L) and Fermi energy (Ef). Due to the effect of inductive coupling between the adjacent layers, Inductance (L_{eq}) in MLGNR interconnects is larger than its copper counterpart. Inductance (L_{eq}) in MLGNR interconnects increases with increase in length and decrease in fermi energy (Ef).

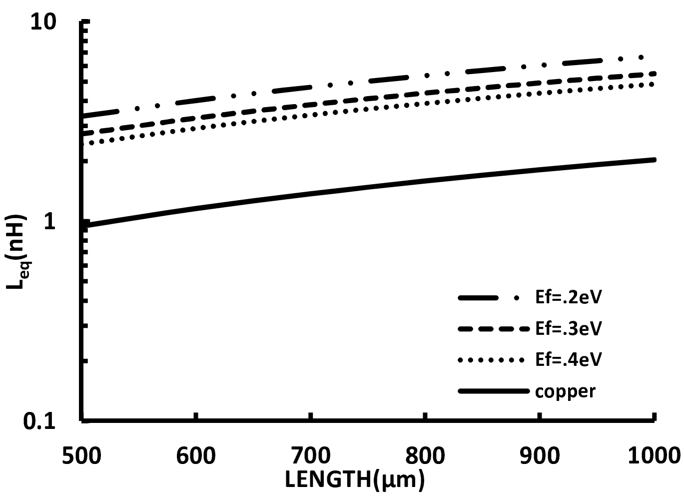


Fig.4. Equivalent inductance (L_{eq}) as a function of Fermi energy (Ef) for global lengths of multilayer GNR interconnect.

Fig.5. shows the dependence of MLGNR interconnect quantum capacitance (C_q) on length and femi energy (Ef). The effect of capacitive coupling between the layers is included. The quantum capacitance (C_q)in MLGNR interconnects increases with increase in length and decrease in fermi energy (Ef).

The distributed capacitance is due to quantum capacitance (C_q)and electrostatic capacitance (C_e). In MLGNR interconnects quantum capacitance (C_q) is much greater than electrostatic capacitance (C_e) [7]. The equivalent capacitance is the resultant of the series combination of quantum capacitance (C_q) and electrostatic capacitance (C_e). This is **shown in Table 1,**the equivalent capacitance is due to the dominance of lowerelectrostatic capacitance (C_e)in MLGNR interconnects.

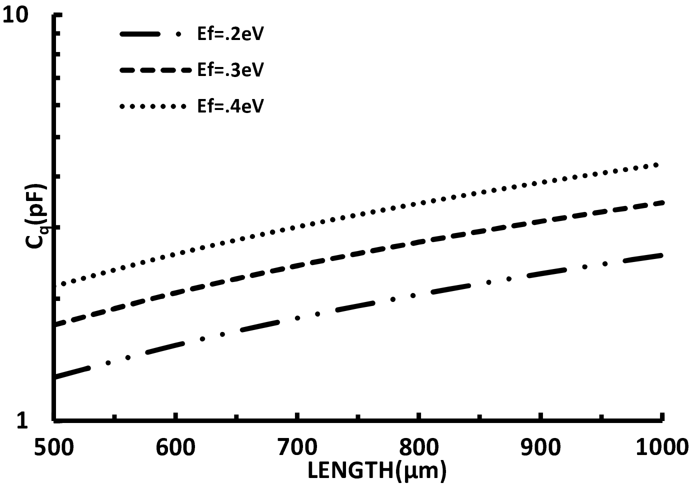


Fig.5. Equivalent Quantum capacitance (C_q) as a function of Fermi energy (Ef) for global lengths of multilayer GNR interconnect.

Table1

S.No.	Lengths	Equivalent capacitance of MLGNR(fF)			Equivalent capacitance of copper(fF)
		Ef=0.2eV	*Ef=0.3eV*	*Ef=0.4eV*	
1	Global(1mm)	22.48	22.53	22.56	14.82

IV. SPICE SIMULATION RESULTS AND DISCUSSIONS

The Simulation setup uses a CMOS inverter driving the MLGNR based equivalent distributed RLC based interconnect having a load of 300fF for local and intermediate lengths and 1pF for global lengths. The input to the inverter is provided by a 0.1 GHz pulse of 1 ns rise time. The performance of this setup is studied by SPICE simulation.The predictive technology model for 22 nm is used for CMOS driver [16]. The equivalent resistance, inductance and capacitance are used for the simulation with optimum number of repeaters.Similar analysis is carried out with copper using the same number of repeaters. The relative measure of MLGNR is obtained by normalizing MLGNR interconnect power by power of copper interconnects. Fig.6. shows the variation of power ratios (Normalized power) of multi-layer GNR and copper for global lengths of interconnect respectively.

Fig.6 shows that power ratios decreases with increase in Fermi energies. Fig.6. shows that power ratio decreases with increase in interconnect length and with increase in Fermi energies (Ef).This is due to the dominance of lower p.u.l. resistanceof MLGNR over inductance with approximately unchanged equivalent capacitance for higher values of Fermi energy(see Fig.3 and Table1).It has also been noted that copper interconnects dissipate more power for all interconnect lengths due to larger resistance compared to MLGNR(see Fig.3). Thus, these simulated results reveal that higher value of Fermi energy should be chosen in case of MLGNR for high speed interconnect applications.

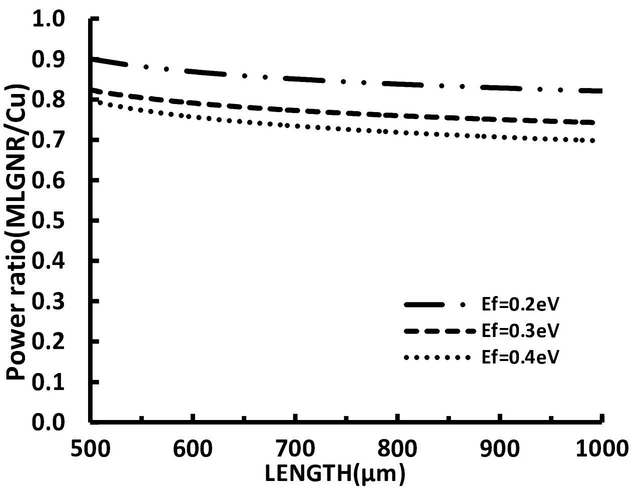


Fig.6 Power ratio (MLGNR/Cu) with varying length for global interconnects for different Fermi energies.

V. CONCLUSION

This paper analyzes the dependence of impedance parameters on Fermi energy.The influence of Fermi energy on power dissipation of MLGNR interconnect is critically examined andis comparedto power dissipation of copper interconnect. SPICE simulation results reveal thatthe use of larger Fermi energies improves the power dissipation of MLGNR interconnect.

REFERENCES

[1] W. Steinhogl, “Comprehensive study of the resistivity of copper wires and lateral dimensions of 100nm and smaller”, Journal of Applied Physics, Vol. 97, 023706 (2005).

[2] A. A. Balandin, S. Ghosh, W. Bao, I. Calizo, D. Teweldebrhan, F. Miao,and C. N. Lau, “Superior thermal conductivity of single-layer graphene,”Nano Lett., vol. 8, no. 3, pp. 902–907, Feb. 2008.

[3] H. Li et al., ‘‘Carbon Nanomaterials for Next-GenerationInterconnectsand Passives: Physics, Status, and Prospects,’’ IEEE Trans. Electron Devices, vol. 56, no. 9,2009, pp. 1799-1821.

[4] A Naeemi, J D Meindl, “Conductance Modeling for Graphene Nanoribbon (GNR) Interconnects”, IEEE Electron Device Letters (2007), vol. 28, no. 5, pp. 428-431, May 2007.

[5] X. Chuan, L. Hong, and K. Banerjee, "Modeling, Analysis, and Design of Graphene Nano-Ribbon Interconnects," Electron Devices, IEEE Transactions on, vol. 56, pp. 1567-1578, 2009.

[6] A. Naeemi and J. D. Meindl, “Conductance modeling for graphene nanoribbon (GNR) interconnects,” IEEE Electron Device Lett., vol. 28, no. 5, pp. 428–431, May 2007.

[7] C. Xu, H. Li, and K. Banerjee, “Graphene nano-ribbon (GNR) interconnects:A genuine contender or a delusive dream?” in IEDM Tech. Dig.,2008, pp. 201–204.

[8] A.K.Nishad and R. Sharma, " Analytical Time-Domain Models for Performance Optimization of Multilayer GNR Interconnects, IEEE Journal of selected topics in Quantum Electronics, VOL. 20, NO. 1, Jan 2014.

[9] V. Kumar, S. Rakheja and A. Naeemi, “Performance and Energy per– Bit Modeling of Multilayer Graphene Nanoribbon Conductors”, IEEE Transactions on Electron Devices, vol. 59,no. 10,pp. 2753-2761, 2012.

[10] S. H. Nasiri,M. K. M. Farshi, and R. Faez, “Stability analysis in graphene nanoribbon interconnects,” IEEE Electron. Device Lett., vol. 31, no. 12, pp. 1458–1460, Dec. 2010.

[11] D. Das and H. Rahaman, "Crosstalk and Gate Oxide Reliability Analysis in Graphene Nanoribbon Interconnects," in Electronic System Design (ISED), 2011 International Symposium on, 2011, pp. 182-187.

[12] J.-P, Cui, W,-S. Zhao, W.-Y. Yin, and 1. Hu, "Signal transmission analysis of multilayer graphene nano-ribbon (MLGNR) interconnects," IEEE Trans. Electromagn. Compat" vol. 53, no, 4, Nov, 2011.

[13] W. Steinhogl, et al., “Size-dependent Resistivity of MetallicWires in the Mesoscopic Range,” Physical Review B, 66,075414, 2002.

[14] Hong Li, Wen-Yan Yin, Kaustav Banerjee and Jun-Fa Mao, “Circuit Modeling and Performance Analysis of Multi-Walled Carbon Nanotube Interconnects,” IEEE Transactions on Electron Devices, vol. 55,no. 6,pp. 1328-1337, 2008 .

[15] International Technology Roadmap for Semiconductors (ITRS) reports, 2006. [Online]. Available: http://w[illegible]

[16] Predictive Technology Model, 2008, [Online]. Available: [illegible]//p[illegible]asu.edu/ .

Design and Simulation of UWB Stacked Patch Antenna using Folded Feed

Mohd Shahab
Deptt. of Electronics and Communication Engg.
Sant Longowal Institute of Engineering and Technology
Sangrur, India
shahabbly@gmail.com

Lakhvinder Singh Solanki
Deptt. of Electronics and Communication Engg.
Sant Longowal Institute of Engineering and Technology
Sangrur, India
lakhvinder_singh_solanki@yahoo.co.uk

***Abstract*—In this paper, a design strategy is presented to achieve enhanced impedance bandwidth for rectangular stacked patch stacked with H shape rectangular patch & fed with a folded patch feed. It has been found that by selecting appropriate width of the folded feed and the size of slots on lower patch, there is considerable improvement in the impedance bandwidth & gain of antenna. We achieve an impedance bandwidth of 121.74 % across the frequency range of 3.6GHz - 14.8GHz. The antenna occupies the compact dimension of 0.419 g by 0.358 g by 0.209 g where g is central operating frequency wavelength. Antenna pattern shows good performance with congrous radiation pattern and significant gain over whole frequency band**.

Keywords—Folded patch fee, patch antenna; UWB antenna.

I. INTRODUCTION

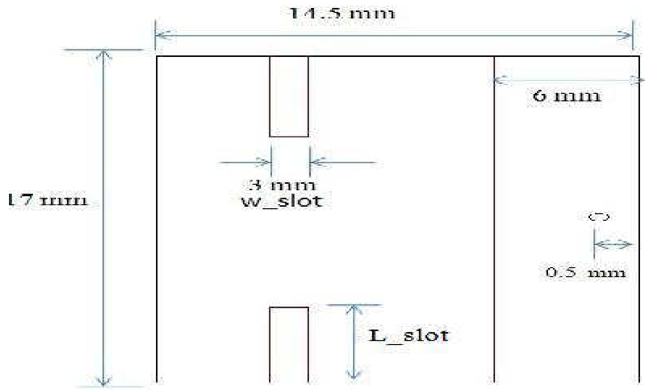
[illegible] ([illegible] B) [illegible] [illegible]. Ultra wideband was formerly known as "pulse radio", but the FCC and the International Telecommunication Radiocommunication Sector (ITU-R) currently define UWB in terms of a transmission from an antenna for which the fractional bandwidth greater than 20%. Also according to Defense Advanced Research Projects Agency (DARPA), UWB antennas require to have a fractional bandwidth greater than 25%.

According to the U.S. Federal Communications Commission (FCC) the authorized unlicensed use of UWB lies in the frequency range from 3.1GHz to 10.6 GHz and power spectral density emission limit for UWB transmitters is −41.3 dBm/MHz. [2]. The advantages of UWB systems over conventional (narrowband) wireless communication systems are its low transmitting power level and high data rates features. They require consistent gain response for sensitivity and stable radiation patterns over the entire bandwidth.. The antenna aspects of UWB systems differ significantly from those narrowband systems. The design of practical antennas that radiate efficiently over an ultrawide bandwidth continues to be a challenging problem. Due to the attractive merits of low profile, lightweight, ease of fabrication and wide frequency bandwidth, patch antennas are currently under consideration for use in ultrawideband (UWB) systems. However, conventional patch antenna suffers typically from a few percent of bandwidth. For this reason much effort is made to develop techniques and find configurations to broaden its impedance bandwidth. It is a common practice to use the stacked patches to increase gain and/ or impedance bandwidth of the microstrip antennas. This reduces the impedance variation of the antenna with frequency, thus enhancing the impedance matching across a broad frequency band. Various arrangements of the stacked patch structures have been investigated. In [5], the maximum bandwidth is 10% and the structure consists of stacked patches with a shorting wall to achieve low profile with two layers of dielectric substrate. For UWB application a dual layer stacked patch antenna with 56.8% bandwidth has been proposed in [6] which has a dimension of 26.5 x 18 x 11.5 mm^3. More recently an ultrawideband suspended plate antenna consisting with two layers has achieved an impedance bandwidth of 72.7% [7]. A new antenna feed method [8] namely the folded-patch feed has been shown to improve the impedance bandwidth of a patch antenna. The rectangular U-shaped-slot patch antenna with a folded patch feed with the dimension of 18 x 15 x 7 mm^3 is shown to have an impedance bandwidth of 53.5% (VSWR ≤ 2) [9]. A stacked patch antenna with a folded feed [10] is presented for ultrawideband application. This design achieves an impedance bandwidth of 111.08%.

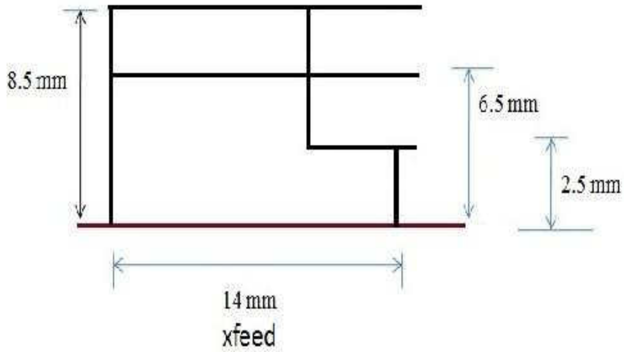
In this paper, a new antenna has been proposed with modified folded patch feed and slots on lower patch which improves the impedance bandwidth. Impedance bandwidth of 111.08% has with [10]. Proposed antenna configuration further improves impedance bandwidth upto 121.73% and also improves gain variation over the frequency band.

II. ANTENNA DESIGN AND STRUCTURE

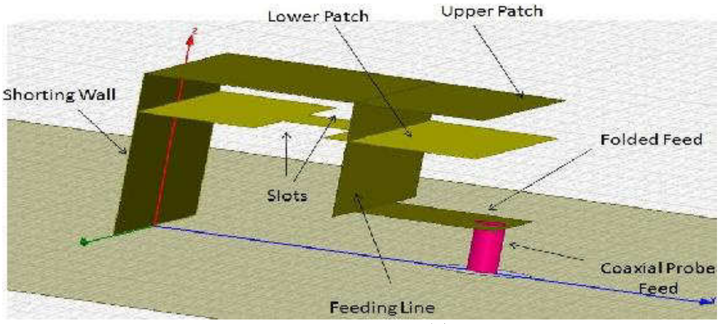
The configuration of the proposed rectangular shaped stack patch fed by modified folded patch feed antenna is illustrated in Fig.1. The proposed antenna design has a dimension of 17 x 14.5 x 8.5 mm³ and whole radiating element is centered on top of a 90 mm x 90 mm ground plane. The radiating patch including the shorting wall is made of 0.2 mm copper sheet as the substrate of the whole proposed antenna is air. The antenna is supporting by coaxial probe and the shorting wall. To support the whole radiating element above the ground plane as well as to increase a few % impedance bandwidth and to reduce overall size of antenna we connected a shorting wall to the ground plane. The probe feed with radius of 0.65 mm is located on the horizontal central line of the folded patch feed with 14 mm away from the left edge. With the aid of folded patch feed design, the probe length shortened, leading to smaller probe inductance. The overall height of the proposed antenna is 8.5 mm, while the length of probe is 2.5 mm only because of this the impedance bandwidth of the antenna increases significantly. The probe length keeps the real part of the input impedance as close to 50 . The above specific shape patch for the slots and the stacked patch configuration allowed us to obtain a satisfactory matching across the frequency band of interest. The width of the slots has been optimize between 3mm to 4mm. The capacitance at the feed point increases by horizontal section of feeding plate which compensate the increase in inductance due to long probe across a broad impedance bandwidth. The extra electromagnetic coupling in between stack patch of design further enhance the impedance bandwidth.



(a)



(b)



(c)

Fig. 1. Geometery of the proposed antenna a) top view, b) side view, c) 3D view

III. ANTENNA PERFORMANCE AND ANALYSIS

The simulation of antenna has been done by Finite-Element Analysis package, Ansoft HFSS. The simulated return loss plot is shown in Fig. 2. It is clearly evident that the bandwidth of the proposed antenna is 121.74 % (10 dB return loss) over the frequency range of 3.6 GHz-14.8 GHz.

As shown in Fig. 3, the gain of the antenna is above 6 dB and considerable over the frequency range. VSWR for the said antenna is plotted in Fig. 4 and is found between 1 and 2. The radiation pattern at 7.4 GHz and 12.7 GHz is shown in Fig. 5. The Fig. 3, 4 and 5 has been drawn for slot width 3mm, slot length 6mm and xfeed taking 14mm. The variation of return loss with slot width values 2mm, 2.5mm, 3mm, 3.5mm and 4mm is shown in Fig.5, it has been found that as the values of slot width increases above 3mm the impedance bandwidth decreases.

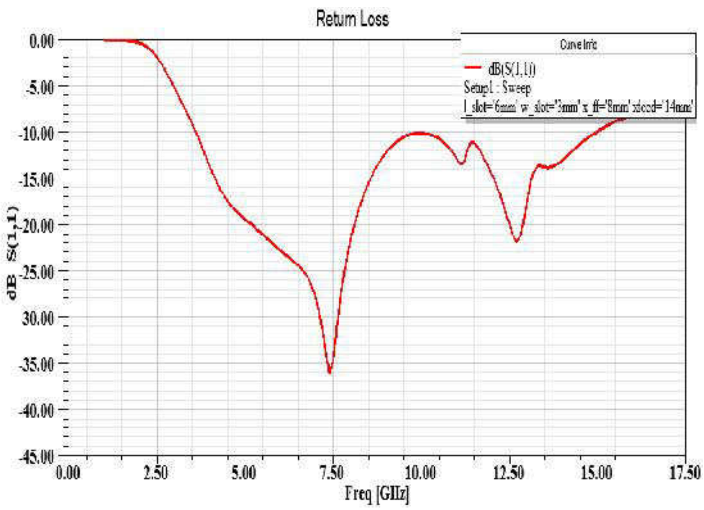


Fig. 2. Return loss of the proposed antenna

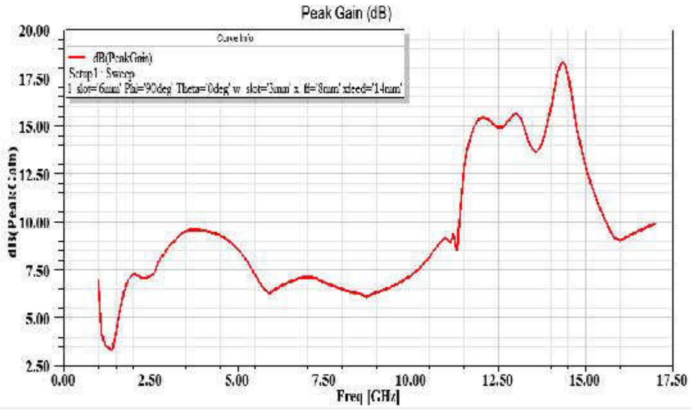


Fig. 3. Peak Gain of the proposed antenna

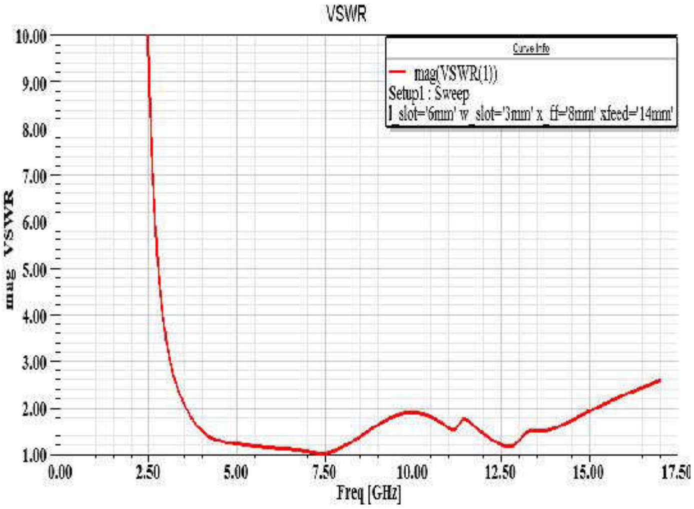


Fig. 4. VSWR of the proposed antenna

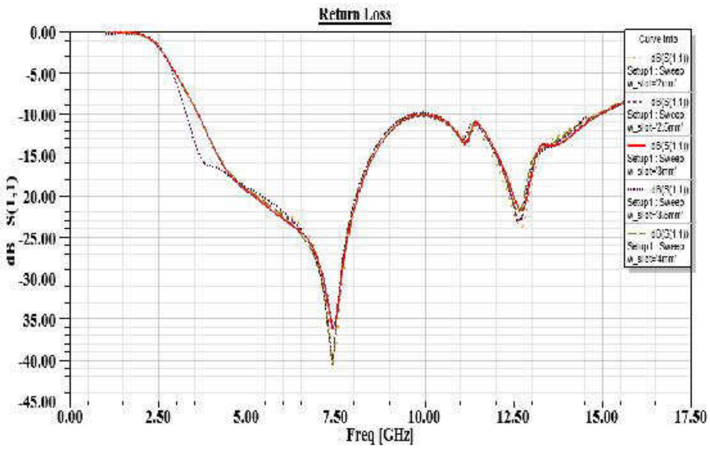
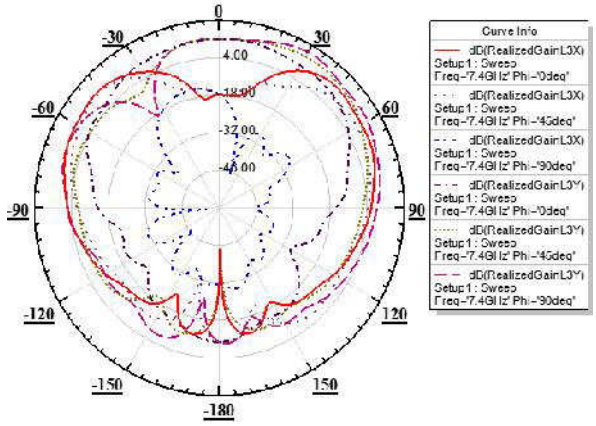
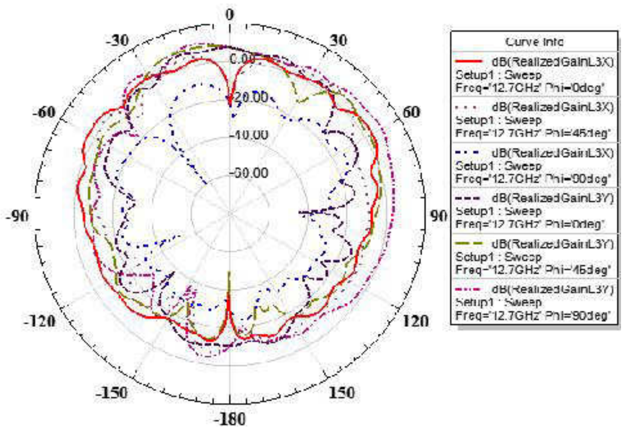


Fig. 5. Return loss variation with slot width



(b)



(c)

Fig. 5. (a) Radiation Pattern of the proposed antenna at 7.4 GHz, (b) 12.7 GHz

IV. CONCLUSION

In this paper, an impedance bandwidth of 121.74% has been achieved over a frequency range of 3.6 GHZ – 14.8 GHz which is 10.66% wider than previously reported design [10] that having impedance bandwidth 111.08%. We optimize the size of slots on the lower patch, width of folded feed and probe feed position. The variation of return loss with slots width has been done and found that by increasing the value above 3mm the impedance bandwidth decreases. This antenna has small size and can be fabricated easily with a copper sheet. The results and dimensions of antenna are mentioned in the Figures.

V. REFERENCES

[1] P. Salonen, M. Keskilammi, and M. Kivikoski, “New slot configurations for dual-band planar inverted-F antenna”, Microwave Opt. Technol. Lett., Vol. 28, pp. 293–298, Mar. 2001.

[2] P. Li, J. Liang, and X. Chen, “Study of printed elliptical/circular slot antenna for ultrawide band application,” IEEE Trans.Antennas Propag., Vol. 54, no. 6, pp. 1670–1675, 2006.

[3] S. D. Tragonski, R.B. Waterhouse, and D.M. Pozer, “Design of wide-band aperture-stacked patch microstrip antennas,” IEEE Trans. Antennas Propag., Vol. 46, no. 9, pp. 1245-1251, 1999

[4] R. B. Waterhouse, “Design of probe fed stacked patched.” IEEE Trans. Antennas Propag., Vol. 47, no. 12, pp. 1780–1784, 1999.

[5] J. Ollikainen, M. Fische, and P. Vainikainen, “Thin dual-resonant stacked shorted patch antenna for mobile communications,” Electron. Lett., vol. 35, no. 6, pp.437–438, 1999.

[6] M.A. Matin, B.S. Sharif, and C.C. Tsimenidis, “Dual layer stacked rectangular microstrip patch antenna for ultra wideband applications,” IET Microw. Antennas Propag., Vol. 1, no. 6, pp. 1192-1196, 2007.

[7] X. N. Low, Z. N. Chen, and W. K. Toh, “Ultrawideband suspended plate antenna with enhanced impedance and radiation. performance,” IEEE Trans. Antennas Propag., Vol. 56, no. 8, pp. 2490-2495, Aug. 2008.

[8] C.Y. Chiu, H. Wong, and C.H. Chan, “Study of small wideband folded-patch-feed antennas,” IET Microw. Antennas Propag.,vol.2, no. 1, pp. 501-505, 2007.

[9] C.Y. Chiu, K. M. Shum, C.H Chan, and K.M. Luk, “Bandwidth enhancement technique for quarter-wave patch antenna,” IEEE Antennas Wireless Propag. Lett., Vol. 2, pp. 130-132, 2003.

[10] Mohammad Tariqul Islam, Mohammed Nazmus Shakib, Norbahiah Misran and Baharudin Yatim, “A Stacked Patch Antenna for Ultrawideband Operation” 9781-4244-2931-8/09 IEEE, ICUWB 2009 (September 9-11, 2009).

[11] M.N. Shakib, M.T. Islam and N. Misrann “Stacked patch antenna with folded patch feed for ultra-wideband application” IET Microw. Antennas Propag., 2010, Vol. 4, Iss. 10, pp. 1456–1461.

[12] Dawar Awan, Shahid Bashir and Nerijus Riauka “Parametric Study of UWB Antenna Loaded with Stacked Parasitic Patch and Reflector “ 2013 Loughborough Antennas & Propagation Conference 11-12 November 2013, Loughborough, UK.

[13] David Gibbins, Maciej Klemm, Ian J. Craddock, Jack A. Leendertz, Alan Preece, and Ralph Benjamin “A Comparison of a Wide-Slot and a Stacked Patch Antenna for the Purpose of Breast Cancer Detection” IEEE Transactions On Antennas And Propagation, Vol. 58, No. 3, March 2010.

Design Analysis of a Circular Polarized Slot Loaded Microstrip Patch Antenna for Multiband Applications

Jatinder Pal Singh
Dept. ECE, CTIEMT, Shahpur
Jalandhar, India
sjatinderpal87@gmail.com

Amandeep Singh
Dept. ECE, CTIEMT, Shahpur
Jalandhar, India
eramandeep23@gmail.com

Nancy Gupta
Dept. ECE, CTIEMT, Shahpur
Jalandhar, India
nancy.gupta24@gmail.com

Abstract—**The designing of an antenna is a critical area for the present wireless communication system. It is important to maintain an agreement amongst gain, bandwidth, return loss and VSWR characteristics of antenna. In this paper an exclusive design concept for the circular patch antenna is introduced. Initially a circular microstrip patch antenna is designed and the performance of this new design is enhanced after applying different types of slotted perturbations in one half of antenna. The antenna performance is transformed from single band to multiband by employing 'Y' shape slotted structures. The proposed antenna represents quad band behavior at resonant frequencies 2.44 GHz, 6.48 GHz, 7 GHz and 8.25 GHz with a good value of return loss of -12.29 dB, -16.48 dB, -15 dB and -27.81 dB respectively. This design also represents VSWR between the specified range of 1 to 2 for the said resonant frequencies with an acceptable value of gain in dB. The antenna is designed and simulated with FEM based electromagnetic field solver. The proposed antenna design is suitable for WLAN, Radio astronomy, Passive sensors and Point to Point defense system wireless applications.**

Keywords—Multiband, Patch antenna, Wireless applications, 'Y' shape.

I. INTRODUCTION

In recent years the requirement for wireless communication systems has increased rapidly demanding quality of service, protection of information and increased data rate for the wireless local area networks (WLANs). The main aim of advance wireless communication system is high speed networking service for different media communication. Microstrip patch antennas offer numerous advantages such as low cost, light weight, low profile, ease of fabrication and conformity [1-2]. Instead of these advantages it also has some disadvantages like narrow bandwidth and low gain.

Several techniques have been suggested to overcome the limitations of the simple microstrip patch antenna. Applying perturbations is one such technique amongst those all suggested for designing the microstrip patch antenna. To comply with the requirements in communication systems such as the Wireless local network (WLAN), Global Positioning Systems (GPS), Radio astronomy, Point to Point high speed wireless applications it is often mandatory that antennas are of compact size, low cost and capable of operating at multiband frequencies [3].

The task of designing antenna becomes very challenging while considering trade-offs associated with operation over multiple frequency bands, constraints on size and limitations of commercial low cost materials Low cost and conformal antennas support the operation of many modern communication systems [4]. The development of patch antennas progresses day-by-day as James et al. [5] presented microstrip patch antenna array for microwave life detection system. This system was developed by Chen and Huang in the year of 2000 and scope of this system is to search human objects under earthquake rubble or behind the barrier. The operating frequency of antenna of this life rescue system is 1.150 GHz.

Agata et al. suggested a novel printed bow-tie antennas for dual-band dual-mode mobile handset applications. They introduced this antenna for dual-band GSM/DCS 1800 operation on a single handset employing dual-band antenna concepts[6]. Behera et al. presented a balanced amplifying microstrip patch antenna at 2.4 GHz. He discussed the advantages of an ideal balanced configuration which comprises good isolation with improved stability, good input and output external matching, cancellation in the load of products and harmonics are considered in designing a balanced configuration for the active antenna [7].

Onofrio Losito proposed high efficiency and broadband microstrip leaky-wave antenna. He was able to achieve tremendous antenna characteristics such as wider band of 33% for VSWR < 2, higher gain and higher efficiency [8]. Roy et al. proposed miniaturized broadband microstrip antennas for HIPERLAN/2 application. They reported U-slot loaded and V-slot loaded proximity coupled microstrip antennas. The performances of two antennas are investigated in frequency band of high performance wireless local area network.

Mohammad Tariqul Islam et al. presented a high gain microstrip patch antenna. His design comprises contemporary techniques; probe feeding, inverted patch structure and stacked multiple slotted patch. The composite effect of integrating these techniques and by introducing the novel multiple shaped patch, offer a low profile, broadband, high gain, and compact antenna element [9]. Asghar et al. proposed a circular microstrip patch array antenna for c-band altimeter system. The aim of this antenna construction was to obtain a high gain, an acceptable pattern and a reasonable value of SWR for altimeter system application [10]. Lai et al. presented and discussed a circular patch microstrip array antenna for ku-band wireless application. The proposed antenna is simplest in construction, low cost and miniaturized in size. This design is suitable for the wireless applications such as WLAN, Radio astronomy and Passive sensors for satellite services [11].

II. DESIGN METHODOLOGY

To design and analyze the proposed antenna, High Frequency structure Simulator (HFSS) electromagnetic software tool is utilized. HFSS is a high performance full-wave electromagnetic (EM) field solver for arbitrary 3D volumetric passive device modeling. It integrates visualization, simulation, solid modeling and automation in an easy-to-learn environment where solutions to 3D EM problems are quickly and accurately obtained [12].

Ansoft HFSS employs the Finite Element Method (FEM), adaptive meshing, and brilliant graphics to give unparalleled performance and insight to all 3D EM problems. Ansoft HFSS can be used to calculate parameters such as S-Parameters, Resonant Frequency, and Fields. HFSS is an interactive simulation system whose basic mesh element is a tetrahedron thus allowing us to solve any arbitrary 3D geometry.

III. ANTENNA LAYOUT AND STRUCTURE

The structure of the proposed antenna design is to be studied in this section. Figure 1 represents the architectural view of the proposed antenna. This antenna is made up of three layers in which first layer comprises the conductive ground plane, second layer comprises the substrate i.e. dielectric material and third layer comprises the conductive patch. Simply we can say the circular patch antenna of specified dimensions is separated from the ground plane by a substrate of ‘Rogers RT/duroid 5880’ of finite thickness. Co-axial probe feed is preferred to analyze the antenna for better impedance matching. Slotted perturbations are employed in upper half of the circular patch antenna to transform the antenna characteristics from single band to multiband.

The dimensions of the slots are to be chosen through literature as well as iterative trials so that the proposed antenna geometry could be able to radiate at multi resonant frequencies. With this only one particular patch antenna is utilized for almost four applications as aforesaid. As the resonating characteristics of the rectangular patch depends upon its length and width similarly resonating characteristics of the circular patch depends upon the radius of the patch which is described from literature review, the effective radius of circular patch is found by [13]:

$$r_{eff} = r\sqrt{1 + \frac{2h}{\pi r \varepsilon_r}\left[ln\left(\frac{\pi r}{2h}\right) + 1.7726\right]} \qquad (1)$$

Where r = physical radius of circular patch

h = height or thickness of the substrate

r = dielectric constant

The effective area of the patch is then given by [14]:

$$A_{eff} = \pi r_{eff}^2 \qquad (2)$$

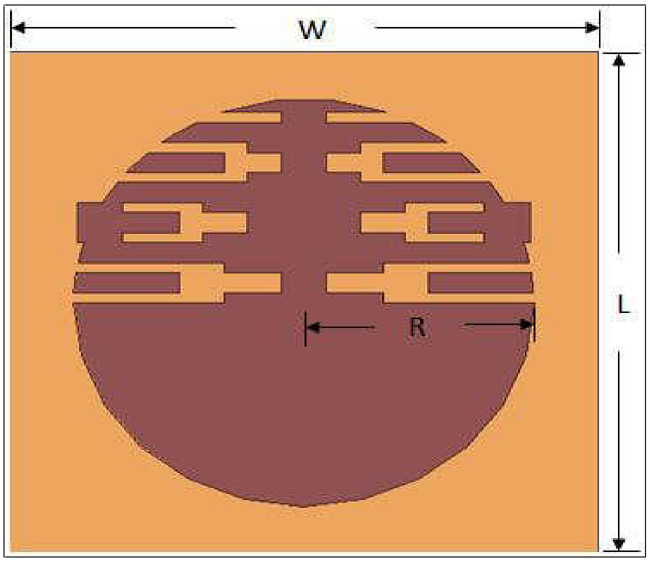


Fig. 1. a) Geometry of proposed antenna,

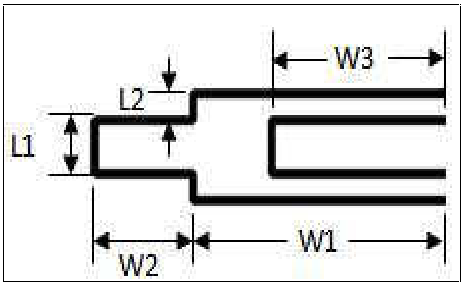


Fig. 1. b) Elementary slotted structure 1

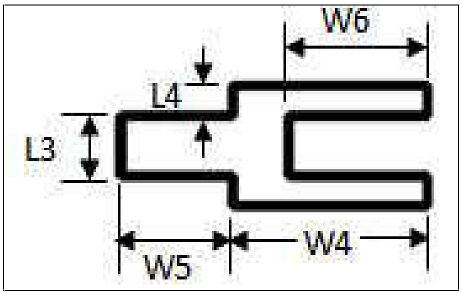


Fig. 1. c) Elementary slotted structure 2

The effective radius and effective area of the proposed antenna are calculated by the equation 1 and 2. The effective radius of the circular patch is somewhat higher than the actual radius because of the fringing phenomenon in the microstrip patch antennas. As the effective radius changes due to fringing, definitely there will be change in the area of the circular patch that actual area is to be calculated by equation 2. To enhance the antenna performance different slotted structures of figure 1b and figure 1c are embedded in the upper half of the proposed antenna as shown in figure 1 a.

The dimensions of the patch antennas are playing very important role in deciding the performance characteristics of the analyzed antenna. The initial dimensions are to be calculated as per the mathematical expression presented by [15] and further optimized values of all the parameters are obtained by having iterative trials on the software. All the optimized dimensions of the proposed antenna geometry are presented in the table 1 for the better understanding. The dimensions of the ground plane i.e. length and width of the ground plane are same as the length and width of the substrate material, therefore ground plane dimensions are not included in the table 1. The dielectric constant of the substrate material plays a vital role in deciding the resonant frequency as well as impedance bandwidth of the patch antenna is represented by the ‘ r’. Radius of the circular patch is represented by the ‘R’ and height of the substrate material is denoted by ‘h’ and their typical values are mentioned in table 1.

TABLE I. DESIGN PARAMETERS OF THE PROPOSED ANTENNA

Antenna parameters	Dimensions
L	50.10 mm
W	52.00 mm
R	22.33 mm
r	2.2
h	1.58 mm
L1	2.00 mm
L2	1.00 mm
L3	2.00 mm
L4	1.00 mm
W1	13.00 mm
W2	5.00 mm
W3	9.00 mm
W4	7.00 mm
W5	4.00 mm
W6	5.00 mm

IV. RESULTS AND DISCUSSIONS

In order to analyze the performance of the proposed antenna design, First of all the height of the Substrate should be selected so that proposed design must radiate at the specified frequencies. Variation in the substrate thickness is chosen according to the availability of the material in the international market. Figure 2 shows the response of the proposed antenna with respect to the frequency by varying the thickness of the substrate from 1.3 mm to 2.65 mm.

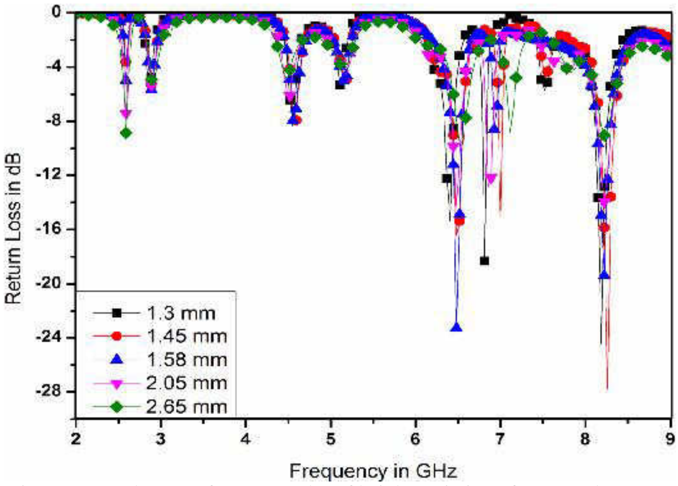


Fig. 2. Return loss Vs. frequency plot of height variation of proposed antenna geometry

It is observed that with substrate thickness 1.3 mm antenna resonates at frequencies 6.4 GHz, 6.81 GHz and 8.18 GHz with return loss characteristics of -15.36 dB, -18.32 dB and -24.44 dB respectively and shown by black squared line in figure 2.

The return loss characteristics changes if substrate thickness is 1.45 mm, antenna resonates at 6.48 GHz, 7 GHz and 8.25 GHz frequencies with S11 -10dB of -16.43 dB, -15.05 dB and -27.81 dB respectively and represented red circled line in figure 2. If thickness increases to 1.58 mm the antenna resonates at frequencies 6.51 GHz, 7 GHz and 8.25 dB with S11 -10dB of -19.82 dB, -11.12 dB and -18.56 dB respectively and shown by blue triangled line in figure 2.

Similarly the antenna performance is observed by varying the thickness upto 2.65 mm and observed that the substrate thickness of 1.58 mm is preferably opted as it provides the required return loss values at specified frequencies. Tabular form of results of return loss versus frequencies is presented in table 2 for more clarity.

TABLE II. RESONANT FREQUENCIES AND RETURN LOSS CHARACTERISTICS AT DIFFERENT THICKNESS OF SUBSTRATE

Substrate height (mm)	Fr (GHz)	S11 (dB)	Fr (GHz)	S11 (dB)	Fr (GHz)	S11 (dB)
1.3	6.4	-15.36	6.81	-18.32	8.18	-24.44
1.45	6.48	-16.43	7	-15.05	8.25	-27.81
1.58	6.51	-19.82	7	-11.12	8.25	-18.56
2.05	6.48	-14.22	6.85	-12.88	8.22	-13.89
2.65	6.51	-9.75	7.11	-8.84	8.22	-9.04

Figure 3 represents the performance of the proposed antenna with different ground sizes varies from 48 mm to 55 mm. At ground width of 48 mm antenna resonates at frequencies 6.48 GHz, 6.96 GHz and 8.22 GHz with S11 -10dB values of -17.26 dB, -10.22 dB and -19.24 dB respectively and shown by black squared line in figure 3.

Further ground width increases to 50 mm and observed that return loss characteristics changes and only two resonating frequencies are found at 6.57 GHz and 8.25 GHz, which means antenna offers dual band performance at 50 mm ground size. The ground width increase to 52 mm and the performance of proposed design is shown by the green triangled plot of figure 3.

TABLE III. RESONANT FREQUENCIES VERSUS RETURN LOSS CHARACTERISTICS FOR DIFFERENT GROUND SIZES

Ground Size (mm)	Fr (GHz)	S11 (dB)	Fr (GHz)	S11 (dB)	Fr (GHz)	S11 (dB)	Fr (GHz)	S11 (dB)
48	2.48	-6.3	6.48	-17.26	6.96	-10.22	8.22	-19.24
50	2.48	-9.3	6.57	-16.39	NA	NA	8.25	-22.87
52	2.44	-12.29	6.48	-16.43	7	-15.05	8.25	-27.81
55	2.44	-0.26	6.51	-15.63	6.96	-13.96	8.22	-17.7

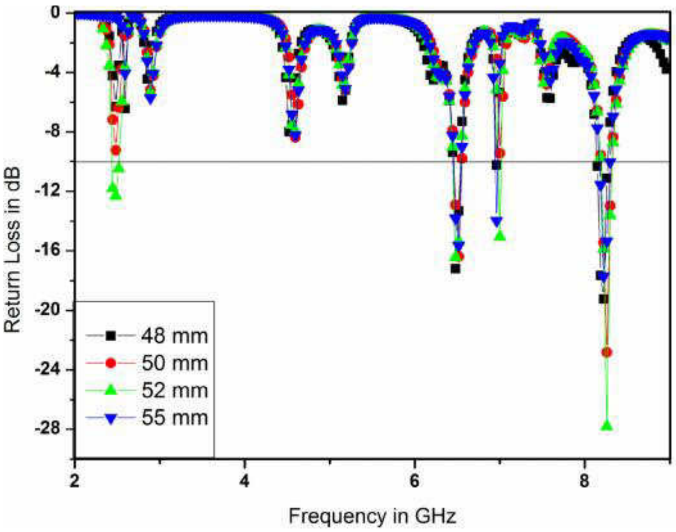


Fig. 3. Return loss Vs. frequency plot of ground size variation of proposed antenna geometry

It is observed that at 52 mm antenna represents the multiband (quad band) characteristics at resonant frequencies 2.44 GHz, 6.48 GHz, 7 GHz and 8.25 GHz with the return loss S11 -10dB values of -12.29 dB, -16.43 dB, -15.05 dB and -27.81 dB respectively. Similarly antenna is analysed for 55 mm ground width and the result is represented by triangled blue plot of figure 3. It is not easy to distinguish the performance of the antenna from the line plots presented in figure 3; therefore all these results are also represented in table 3 for more convenience.

No doubt we have discussed the analyses of proposed antenna to achieve the perfect values of the substrate thickness and ground plane width size for the particular antenna, but without knowing the gain pattern of the antenna we cannot claim that our proposed antenna is a good radiator or absorber. The 3D gain plot of the proposed antenna is represented in figure 4.

It is observed from the 3D gain plot of the proposed antenna that it radiates almost equally in all directions normal to the patch means it exhibits Omni directional characteristics. The gain of 6.33 dBi is obtained from the antenna which is very sufficient for the small distance wireless applications.

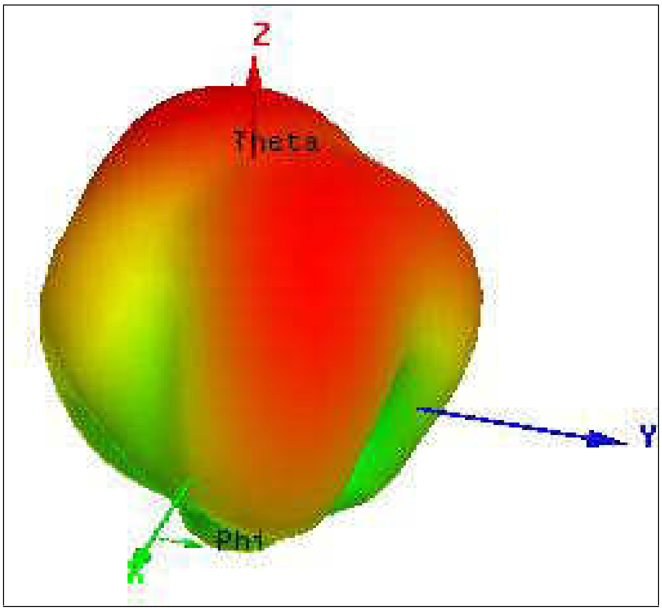


Fig. 4. 3D gain plot of proposed antenna geometry

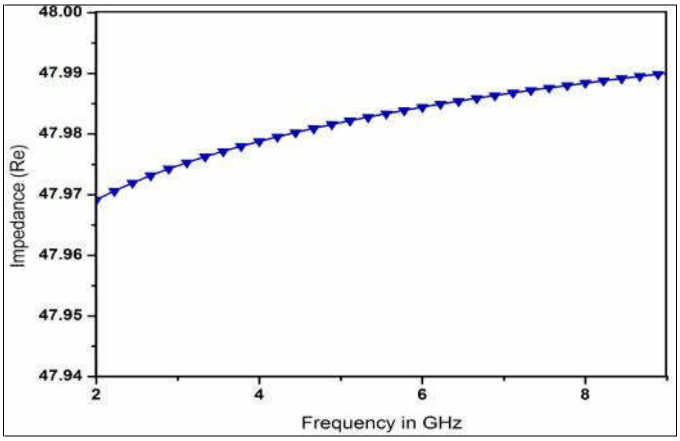


Fig. 5. Input impedance Vs. frequency plot of proposed antenna geometry

Input impedance is very important factor while analysing the antenna because it tells about the matching characteristics of the feedline and the patch antenna. If the input point at the patch is not matched with the feedline impedance then most of the signal is reflected back from the feed point which causes the development of voltage standing waves in the feedline. The perfect match of feedline and feed point is obtained at 50 ohm impedance. Figure 5 represents the input impedance plot at different resonant frequencies and it is observed that the impedance value varies nearly to 50 ohm for all the resonating frequencies.

V. CONCLUSION

A novel circular patch antenna design is investigated and the parameters are chosen by mathematical as well as iterative trial with the simulation software. The analysed design exhibits multiband behaviour at different resonant frequencies of 2.44 GHz, 6.48 GHz, 7 GHz and 8.25 GHz with the return loss S11 -10dB values of -12.29 dB, -16.43 dB, -15.05 dB and -27.81 dB respectively. It is also observed that the proposed design offers a good value of gain of 6.33 dB and it is also matched at input impedance approximately equal to 50 ohm. The proposed antenna design is suitably recommended for the applications like WLAN (2.44 GHz at -12.29 dB), Radio astronomy (6.48 GHz at -16.43 dB), Passive sensors (7 GHz at -15.05 dB) and Point to Point defence system (8.25 GHz at -27.81 dB) wireless applications.

REFERENCES

[1] W. He, R.Jin and J. Geng., "E-Shape patch with wideband and circular polarization for millimeter-wave communication", IEEE Transactions on Antennas and Propagation vol.56, no.3, pp. 893-895, 2008.

[2] Zhang, Y.P. and J.J., Wang. "Theory and analysis of differentially-driven microstrip antennas", IEEE Transactions on Antennas and Propagation vol.54, no.4, pp.1092-1099, 2006.

[3] Mahmoud N. Mahmoud, Reyhan Baktur," A Dual Band Microstrip Fed Slot Antenna", IEEE Transaction on Antennas and Propagation, vol. 59, no. 5, pp. 1720-1724, May 2011.

[4] J. R. James, P. S. Hall, "Handbook of Microstrip Antennas", vol. 1, Peter Peregrinus Ltd., London, 1989.

[5] K. M. Chen, Y. Huang, J. Zhang, and A. Norman, Microwave life detection system for searching human subjects under earthquake rubble or behind barrier, IEEE Trans. Biomed. Eng., vol. 27, no. 1, pp. 105-114, Jan. 2000.

[6] S. Agata and S.H.David, "Novel Printed Bow-Tie Antennas for Dual-Band Dual-Mode Mobile Handsets", IEEE, London, UK, 1998.

[7] S.K. Behera, R.K. Mishra, and D.R. Poddar, "Balanced amplifying microstrip patch antenna at 2.4 GHz", IEEE Applied Electromagnetics Conference, AEMC 2007, pp. 1-4, Kolkata, Dec 2007.

[8] O. Losito, "High Efficiency and Broadband Microstrip Leaky-Wave Antenna", Active and Passive Electronic Components, vol. 2008, Article ID 742050, pp. 1-6, 2008.

[9] M.T. Islam, M.N. Shakib, and N. Misran, "Broadband Microstrip Patch Antenna", European Journal of Scientific Research, vol.27, no.2, pp.174-180, 2009.

[10] A. Keshtkar, Ahmed Keshtkar and A. R. Dastkhosh,., "Circular Microstrip Patch Array Antenna for C-Band Altimeter System" , International Journal of Antennas and Propagation, vol. 2008, Article ID 389418, pp. 1-7, 2008.

[11] T.F. Lai, W.N.L Mahadi, and N.Soin , "Circular Patch Micro strip Array Antenna for KU-band" World Academy of Science, Engineering and Technology, pp. 48, 2008.

[12] M.S.Alam, M.T. Islam, N. Misran,"Design analysis of an electromagnetic band gap microstrip antenna", Am. J. Applied Sci., Vol. 8, Issue 12 , 2011.

[13] C.J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private communication, May 1995.

[14] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interfaces(Translation Journals style)," IEEE Transl. J. Magn.Jpn., vol. 2, Aug. 1987, pp. 740–741 [Dig. 9th Annu. Conf. Magnetics Japan, pp. 301, 1982.

[15] R. Garg, P. Bhartia, I. J. Bahl, A. Ittipiboon,"Microstrip Antenna Design Handbook", Artech House, 2001.

Broadband Slotted Equilateral Triangular Microstrip Patch Antenna for Multiband Communication

Nancy Gupta
Dept. ECE
CTIEMT, Shahpur
Jalandhar, India
nancy.gupta24@gmail.com

Navneet Singh
Dept. ECE
CTIEMT, Shahpur
Jalandhar, India
navneets34@gmail.com

Romil Kumar
Dept. ECE
CTIEMT, Shahpur
Jalandhar, India
kumar_romil@yahoo.co

***Abstract*—In today's era, Microstrip antennas play an essential part in wireless communication systems. The work done here results in multiband and broadband Equilateral Triangular Microstrip Patch Antenna, which is designed and examined using Finite-Element-Method based electromagnetic solver. Further this design is improved by introducing multiple slotted perturbations in the antenna geometry. The anticipated antenna resonates at three frequency bands while the latter frequency band being broadband.**

Keywords—Equilateral triangular microstrip patch antenna (ETMSA), Coaxial probe feed, Return los, Multiband, Broadband

I. INTRODUCTION

The continuously increasing trend towards mobility and information access has made reliable and efficient wireless communications a prime necessity. Antennas play an incredible role in wireless communication technology. Microstrip patch antennas are becoming reliable approach as modern antennas due to numerous advantages offered by them. Some of these include planer configuration, low profile, light weight, easy to fabricate, conformal, polarization configurability, easy integration with microwave integrated circuitry (MIC) etc. [1].

Microstrip antennas basically consist of radiating patch on one side of dielectric substrate, having conducting ground plane on another side. The radiation in microstrip antenna occurs primarily due to fringing fields between the radiating patch and the conducting ground plane. The conducting patch can have various possible shapes such as square, rectangle, circular, elliptical, triangular etc. which has great impact on the radiation recital of antenna [2-3]. Although Rectangular and circular patch antennas are generally known to have good radiation characteristics, but the Equilateral Triangular Microstrip Patch Antenna (ETMSA) has the added advantage of occupying less metalized area on the substrate than other existing configurations. This has made ETMSA's more attractive due to increased demand of small antennas for personal communication equipments [4]. In 1978, Helszajn and James introduced theoretical and experimental investigation on Equilateral Triangular Microstrip Antenna as disk resonator, filter and circulator [5]. A.A. Deshmukh et. al. proposed a broadband equilateral and half equilateral triangular antenna using proximity feed coupling technique [6].

Although having various advantages, microstrip antennas suffer from a major drawback of narrow impedance bandwidth and there has been continuous thrust in this area of research to improve the operational bandwidth for these antennas. Also multiband operation has been attracting researchers continuously so that single antenna can be used for two or more resonant frequencies for various wireless and radar applications [7].

In this paper, an equilateral triangular antenna is designed and analyzed using finite element based electromagnetic solver HFSS (High Frequency Structure Simulator). In addition, this design is perturbed by using three narrow rectangular slots so as to obtain broadband behavior.

The paper is organized as follows. Section II discusses the geometry and design of equilateral triangular antenna along with equations for resonant frequency and effective side length. Section III discusses various results obtained using HFSS simulator followed by section IV that ends with the brief conclusion.

II. ANTENNA DESIGN LAYOUT

An Equilateral slotted microstrip patch antenna is designed using HFSS simulator software. The proposed Equilateral Triangular Microstrip Antenna is shown in *fig. 1.*

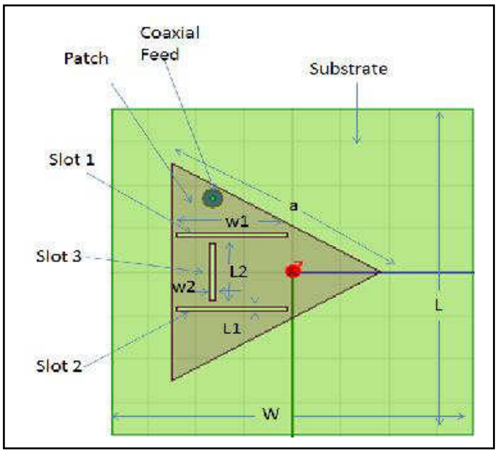


Fig. 1. Slotted equilateral microstrip Antenna

The substrate used for the proposed design is FR4-Epoxy having dielectric constant of 4.4. The height of substrate is taken to be 1.6 mm. The conducting ground plane is having same dimensions of substrate i.e. 75 X 75 mm. The square air box is considered for radiation whose size is 120 X 120 mm.

The various dimensions for the proposed geometry are given in table 1.

TABLE I. ANTENNA GEOMETRY DIMENSIONS

Antenna Parameters	Dimensions
Substrate Height	1.6 mm
L	75 mm
W	75 mm
a	25 mm
L1	1 mm
W1	23 mm
L2	13 mm
W2	1 mm

The resonant frequency for equilateral triangular antenna as discussed in [8-10] can be determined as follows:

$$f_{m,n,l} = \frac{2c}{3a\sqrt{\epsilon_r}}\sqrt{(m^2 + mn + n^2)} \quad (1)$$

Where m, n and l are mode integers due to the electric and magnetic boundary conditions. ϵ_r is dielectric constant of substrate, c is speed of light in free space, a is side length of equilateral triangle.

To account for the effects of fringing field, Dahale and Lee in [8] proposed a correction in side length, resulting in correction of 2 percent error in resonant frequency

$$a_{eff} = a + \frac{h}{\sqrt{\epsilon_r}} \quad (2)$$

Where a_{eff} is effective side length, a is actual side length, h is height of substrate.

III. RESULTS AND DISCUSSIONS

The ETMSA is analyzed for many feed locations using coaxial probe feeding technique to provide signal to the antenna with proper impedance matchnig. Coaxial probe feed is popular among many feeding techniques due to advantages such as easy to fabricate and good impedance matchnig. Proper impedance matching at desired feed location results in maximum coupling of signal from feeding line to the antenna and thus results in maximum radiation efficiency. *Fig. 2* shows the return loss plot for the proposed ETMSA. As it can be seen from plot, the minimum return losses occurs at three bands (5.75 -6.12 GHz, 7.22 GHz, 7.65-8.92 GHz). The first band 5.75-6.12 GHz results in bandwidth of 375 MHz while latter band 7.65-8.92 GHz results in broadband of 1275 MHz. Thus this proposed slotted ETMSA not only act as multiband antenna but also as a broadband antenna covering frequency range from 7.65 to 8.92 GHz. Moreover, this latter frequency band is also having minimum return loss of upto -33.104 dB.

The gain at various operational frequencies discussed in return loss plot, is simulated through 3-D polar plot in HFSS and is shown in *fig. 3(a,b,c and d)*. As it can be seen from *fig. 3* that proposed antenna is offering positive gain at all the operating bands i.e. proposed antenna can be used effectively at any of the three bands discussed in *fig. 2*.

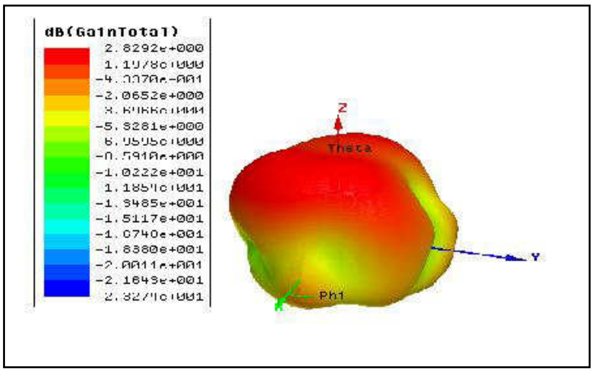


Fig. 3. a) Gain plot at frequency 6.025 GHz

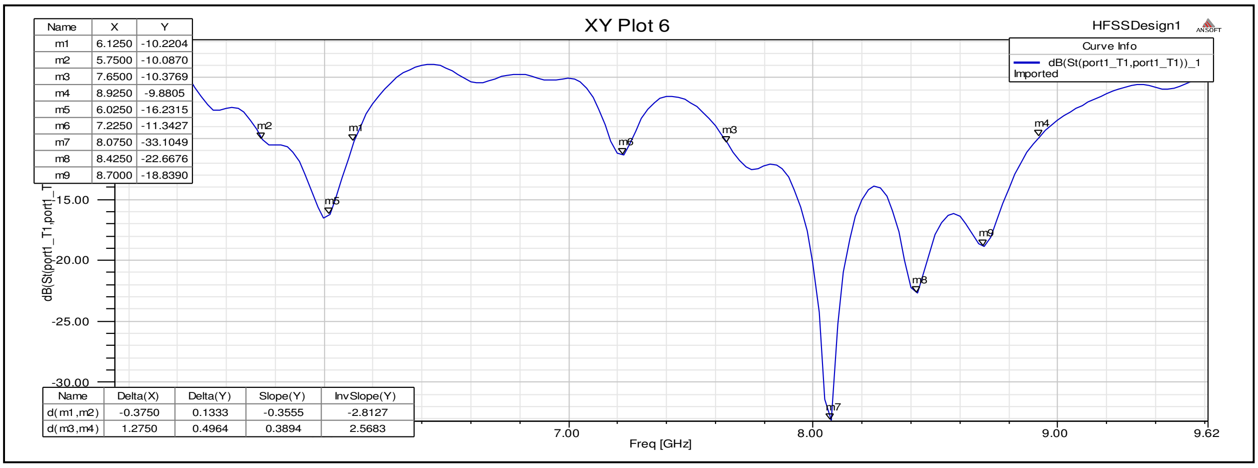


Fig. 2. Return loss plot for slotted traingular microstrip antenna

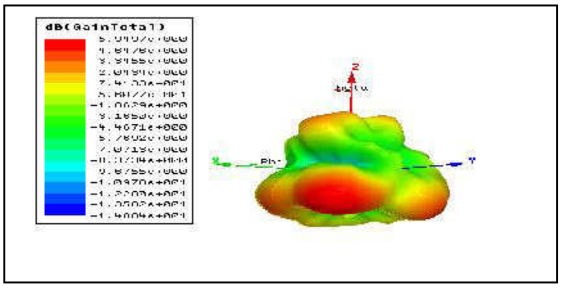


Fig. 3. b) Gain plot at frequency 8.075 GHz

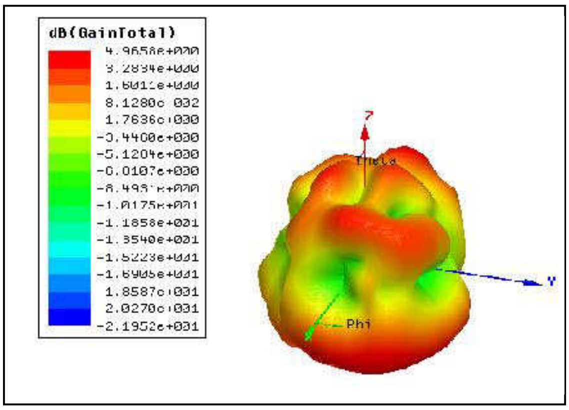


Fig 3 c) Gain plot at frequency 8.425 GHz

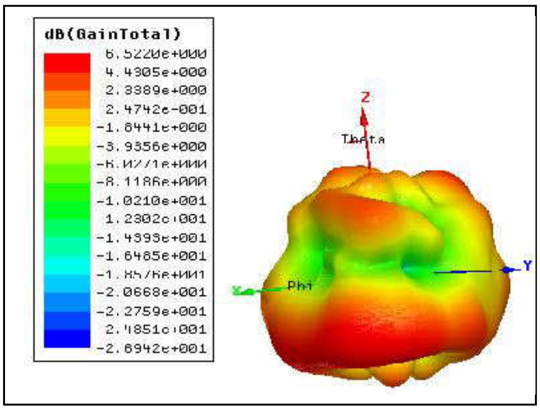


Fig. 3. d) Gain plot at frequency 8.7 GHz

Fig. 4 (a,b,c,d) represents the radiation pattern at azimuthul angle (0 and 90 degree) at the above said frequency bands which shows the radiation characteristics of antenna. It can be analyzed from above figure that radiation for the anticipated design is quite omnidirectional in the frequency range 6.025 GHz while at higher operating frequencies; proposed antenna's radiation is quite directional in some particular directions. This is probably because of some undesired higher order modes getting excited at higher frequencies, resulting in some distortions in resulting radiation patterns.

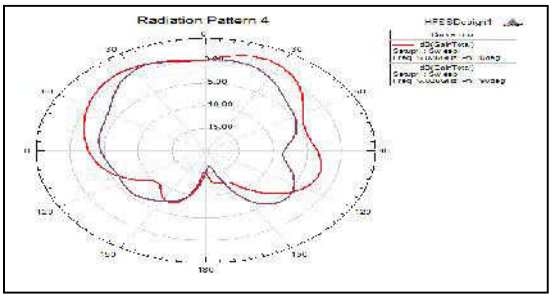


Fig. 4. a) Radiation pattern at 6.025 GHz ($=0^{o}$ and 90^{o})

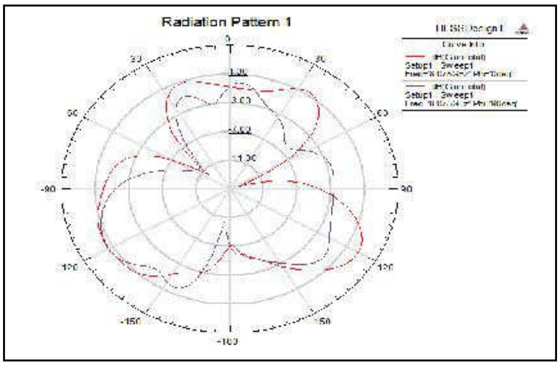


Fig. 4. b) Radiation pattern at 8.075 GHz ($=0^{o}$ and 90^{o})

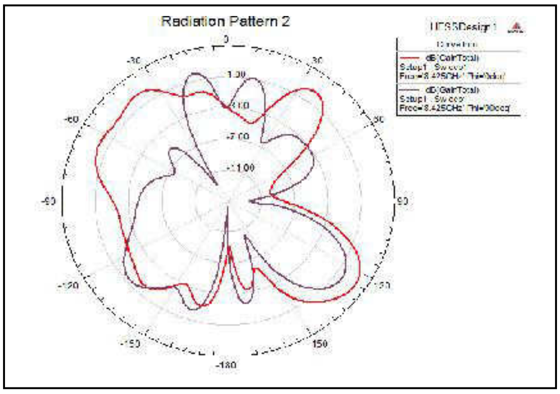


Fig. 4. c) Radiation pattern at 8.425 GHz ($=0^{o}$ and 90^{o})

IV. CONCLUSION

A multiband and broadband Equilateral triangular microstrip patch antenna has been designed and analyzed using finite element based HFSS simulator. It is possessing good return loss at three frequency bands while third frequency band is also resulting in broad impedance bandwidth of 1275 MHz. The gain and radiation pattern at these frequency bands provide good balance with return loss for the anticipated design. This antenna can be made to operate at different frequency bands by adjusting the position and dimensions of slotted perturbations in the design.

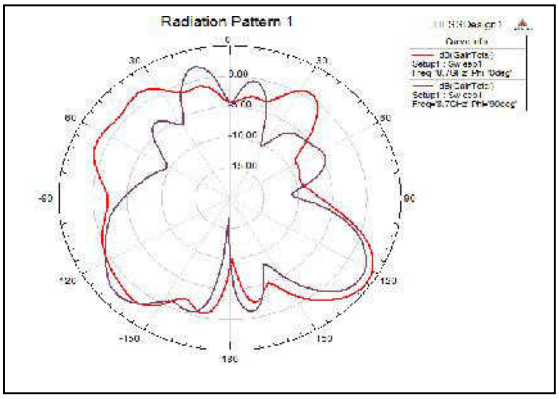


Fig. 4. d) Radiation pattern at 8.7 GHz ($=0^{o}$ and 90^{o})

REFERENCES

[1] D.M. Pozar, “Microstrip Antennas,” Proceeding of the IEEE, Vol. 40, No. 1, January 1992.

[2] J.R. James, P.S. Hall, and C. Wood, “Microstrip antenna: Theory and Design,” IEE Electromagnetic waves, Series 12, London: Peter Peregrinus, 1986.

[3] J.F. Zucher, and F.E. Gardiol, Broadband Patch antennas, Artech House, 1995.

[4] H.R Hassani and D. Mirshekar-Syahkal, “Analysis of triangular patch antennas including radome effects,” IEE Proceedings on Microwaves, Antennas and Propagation, vol. 139, pp. 251 – 256, 1992.

[5] J. Helszajz and D. S. James, “Planar triangular resonators with magnetic walls,” IEEE Trans. Microwave Theory and Tech., vol. MTT-26, pp. 95-100, 1978.

[6] A.A. Deshmukh, S. Nagarbowdi, and N. Phatak, “Broadband Gap Coupled Variations of Equilateral Triangular Microstrip Antennas,” International Journal of Computer Applications, Vol. 88, No. 8, pp. 34-40, 2014.

[7] G.P. Gautheir, A. Courtay, and G.M. Rebiez, “Microstrip antennas on synthesized low dielectric constant substrate,” IEEE Transactions on Antenna Propagation, 1997.

[8] R. Garg, and S.A. Long, “An Improved Formula for the Resonant Frequencies of the Triangular Microstrip Patch Antenna,” IEEE Trans. on Antenna and Propagation, vol. 36, Issue: 4, 1988.

[9] J. S. Dahele, and K. F. Lee, “On the Resonant Frequencies of the Triangular Patch Antenna,” IEEE Trans. Antennas and Propagation, vol. -35, Issue. 1, pp. 100-101, 1987.

[10] X. Gang, “On the Resonant Frequencies of Microstrip Antennas,” IEEE Trans. on Antenna and Propagation, vol.-27, No. 2, pp. 245-247, 1989.

Multi-Slotted Microstrip Patch Antenna for Multiband Communication

Punit
Dept. ECE
CTIEMT, Shahpur
Jalandhar, India
punitksanyal@gmail.co

Utkarsh
Dept. ECE
CTIEMT, Shahpur
Jalandhar, India
chandelutkarsh1991@gmail.com

Nancy Gupta
Dept. ECE
CTIEMT, Shahpur
Jalandhar, India
Nancy.gupta24@gmail.com

Navneet Gill
Dept. ECE
CTIEMT, Shahpur
Jalandhar, India
pallnavneet5@gmail.com

***Abstract*—A Multi-slotted Microstrip Patch Antenna has been designed in order to meet the multiband requirements in S band of electromagnetic spectrum. The proposed design has been simulated and analyzed using HFSS simulator based on Finite-Element Method. It possess good return loss (-24.24 dB, -13.6 dB, -12.91 dB), gain of 6.03 dB and good directional characteristics in broadside to proposed patch antenna. This design has added advantage of being small in size and simple in design.**

Keywords—multi-slotted patch antennas, multiband, return loss

I. INTRODUCTION

In today's era, there has been increasing need of mobility and huge pre-requisite for people to communicate with each other and to have timely access to information regardless of the location of the individuals or the information. Such as a phone call placed from a distant place may begin an important corporate deal or a remote access to medical records by a medical practitioner may play an incredible role in saving a life.

The engineering challenges posed by increasing mobile communications can be met only with an efficient, reliable wireless communication networks for which antennas play an important role. Due to various advantages such as low profile, conformable to planar and non-planar surfaces, simple and inexpensive manufacturing using modern printed-circuit technology, mechanically robust when mounted on rigid surfaces, compatibility with MMIC designs, versatility in terms of resonant frequency, polarization, pattern and impedance, Microstrip patch antennas are very prevalent these days[1].

Although patch antennas were initially proposed by Georges Deschamps in 1953 in USA, the first practical patch antenna was fabricated in 1970's by Byron [2,3]. A Microstrip Patch Antenna (MPA) consists of a flat sheet or "patch" of metal, mounted over a larger sheet of conductor called a ground plane separated by substrate of specific height and dielectric constant. The substrate height and dielectric constant has significant impact on antenna performance. The metal patch can acquire multiple shapes such as rectangle, square, ring, elliptical, circular disc, triangle, pentagon, and hexagonal antennas [4,5] .

In today's era, there is huge requirement of designing patch antennas that allows for multiple frequency band operation. Generally, the Microstrip antennas are having major drawback of narrow operational bandwidth; thereby designing of multiband patch antennas is an area of wide research for many wireless applications. Introducing multiple slots in antenna is a very prevalent practice that allows microstrip patch antennas to resonate at multiple frequencies as these slots introduce the modes near the fundamental mode of resonant frequency, resulting in multiband operation [6,7]. There has been huge work done in this area by several researchers. Prachi et. al. presented different slotted rectangular antenna structures to achieve dual band operation in frequency range 0.7 to 3.5 GHz using HFSS simulator [8]. K.F. Lee et. al. presented experimental Study of the U-Shaped Slot rectangular antenna along with the effects of patch size, slot size and feed position on the performance of antenna. The researchers also discussed two element array of U slot patches [9]. M. H. Ullah et. al. discussed a dual band slotted patch antenna in the frequency range 14.3[illegible]2GHz and 17.4 GHz to 18.9 GHz using HFSS electromagnetic solver and later on, proposed antenna is printed on epoxy polymer resin material substrate for experimental testing [10]. M. R. Zaman et. al. proposed multiband patch antenna for femtocell applications in LTE(Long Term Evaluation) frequency range. The proposed antenna composed of monopole structure with rectangular slot in ground plane [11].

II. PROPOSED ANTENNA DESIGN

The slotted rectangular microstrip patch antenna discussed in this paper, has been designed and simulated using HFSS simulation software that results in tri-band of resonant frequencies that lie between S-band of electromagnetic spectrum. The proposed slotted patch antenna is shown in *fig. 1*.

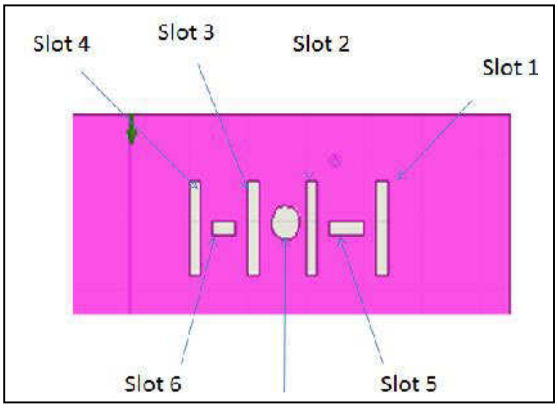


Fig. 1. Proposed slotted patch antenna.

The first resonant frequency is at 2.26 GHz ultra-high frequency band. The next two resonant frequencies are at 2.75 and 3 GHz respectively. The substrate material used is FR4_epoxy with the height of 1.6 mm. The dielectric constant of substrate material used is 4.4. The anticipated antenna is rectangular patch antenna with multiple slots introduced in it. The proposed geometry includes six rectangular slots and one circular slot whose position and dimensions are given in table 1 and table 2.

TABLE I. RECTANGULAR SLOT DIMENSIONS

Slot no.	Length (mm)	Width (mm)	Position
1	14	2	14,10,1.6
2	14	2	14,20,1.6
3	14	2	14,30,1.6
4	14	2	14,40,1.6
5	2	4	14,12,1.6
6	2	4	14,32,1.6

TABLE II. CIRCULAR SLOT DIMENSIONS

Slot no.	Radius(mm)	Position
7	2.5	20.1, 26.5,1.6

One of most popular feeding techniques, coaxial probe feeding is used to provide signal to the antenna. The ground size is 70X80 mm while patch size is 30X75 mm. The air box considered is of 120X120 mm. The complete geometry along with substrate, ground and air box is shown in fig 2.

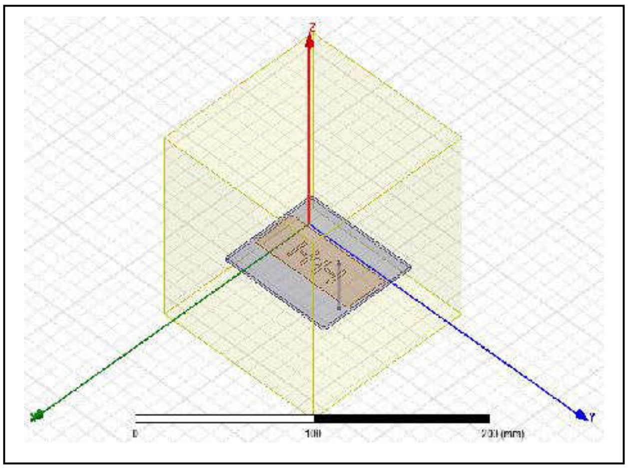


Fig. 2. Slotted MPA in HFSS simulator.

III. RESULTS AND DISCUSSIONS

The return loss is shown in *fig. 3* which shows that proposed slotted antenna can be used at three resonant frequencies. It has been observed from *fig. 3* that return loss is below -10 dB for the three frequencies i.e. 2.26 GHz, 2.75 GHz and 3 GHz. The coaxial feed technique is used due to advantages such as good impedance matching and ease of fabrication.

The proposed antenna is simulated for various feed locations using error and trial method so as to achieve proper impedance matching at selected feed position. Providing signal at proper feed location plays an important role in antenna performance. The proper impedance matching at selected feed location will result in minimum return loss and there will maximum coupling of power from feeding line to the antenna. Thus, antenna will be acting as a good radiator of electromagnetic waves.

Fig. 4 and *fig. 5* shows the gain and radiation pattern plot for slotted antenna. The radiation Pattern shows graphical representation of power radiated by antenna. The radiation pattern is plotted for azimuthal angle (phi) at 0 and 90 degrees for co-polarization and cross-polarization results. The plot indicates that maximum power is radiated in broadside (perpendicular) direction to patch antenna. The radiation pattern shown is for E plane of an antenna. The *fig. 5* indicates that proposed antenna is providing maximum gain of 6.03 dB in perpendicular direction (broadside) to patch antenna surface. The plot indicates that the gain is quite uniform in broadside direction. The proposed antenna is showing great radiation characteristics in pattern of *fig. 4.*

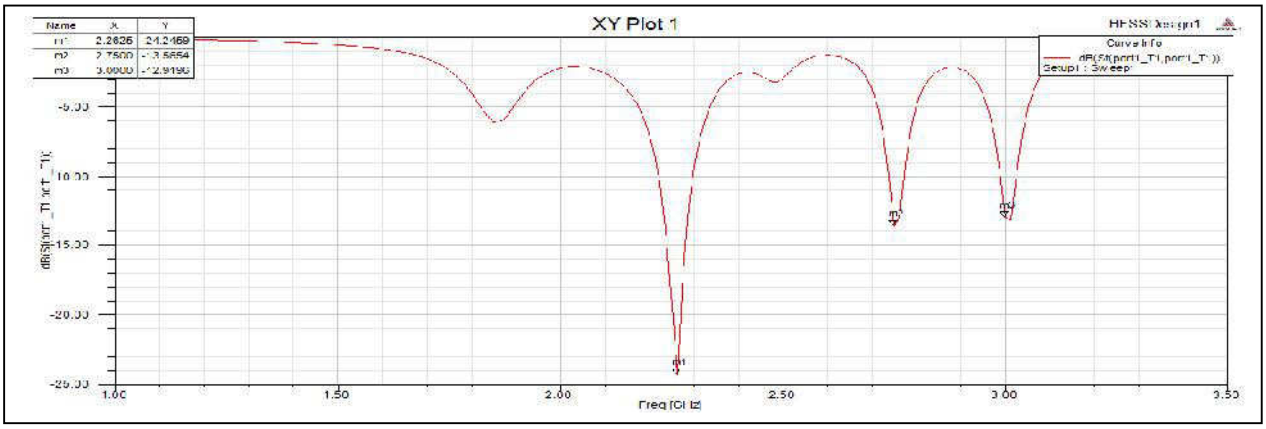
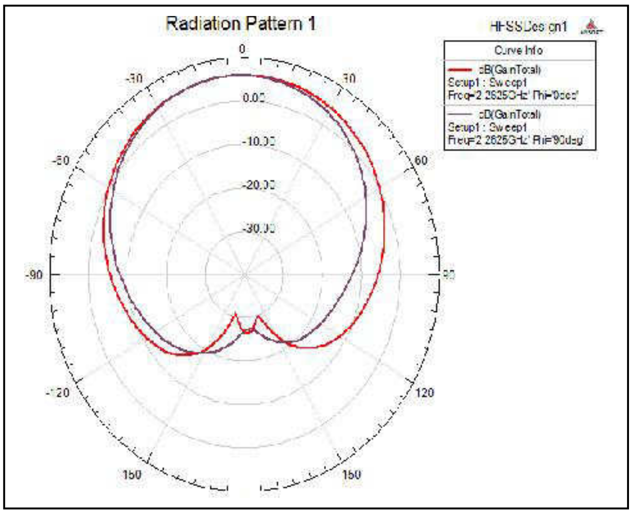


Fig. 3. Return loss for proposed antenna

Fig. 7. Axial ratio plot of multi-slotted antenna



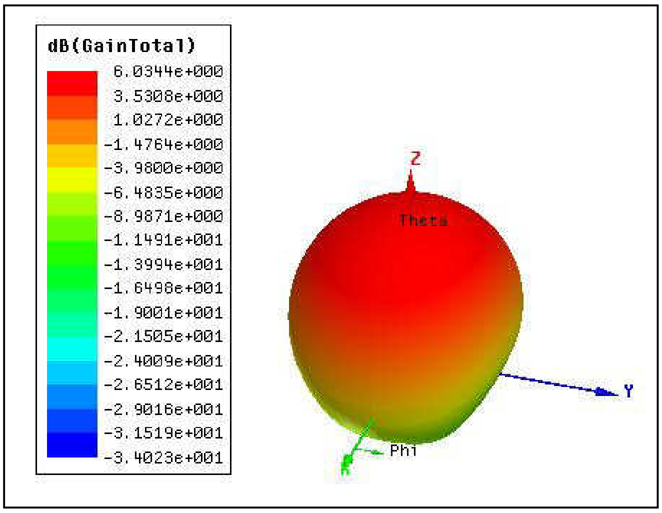


Fig. 4. E-plane radiation pattern of slotted patch antenna

Fig. 5. 3-D polar plot of slotted patch antenna representing gain

The *fig. 6* shows the electric field distribution of proposed slotted patch antenna surface at resonant frequency of 2.26 GHz. The *fig. 7* indicates the axial ratio plot in direction of maximum radiation (theta 0 degree) which shows that axial ratio is far above the 2 dB criteria for the required resonant frequencies. It can be inferred from the graph that the anticipated antenna possess nearly linearly polarization. Thus, making this antenna ideal as a directional antenna.

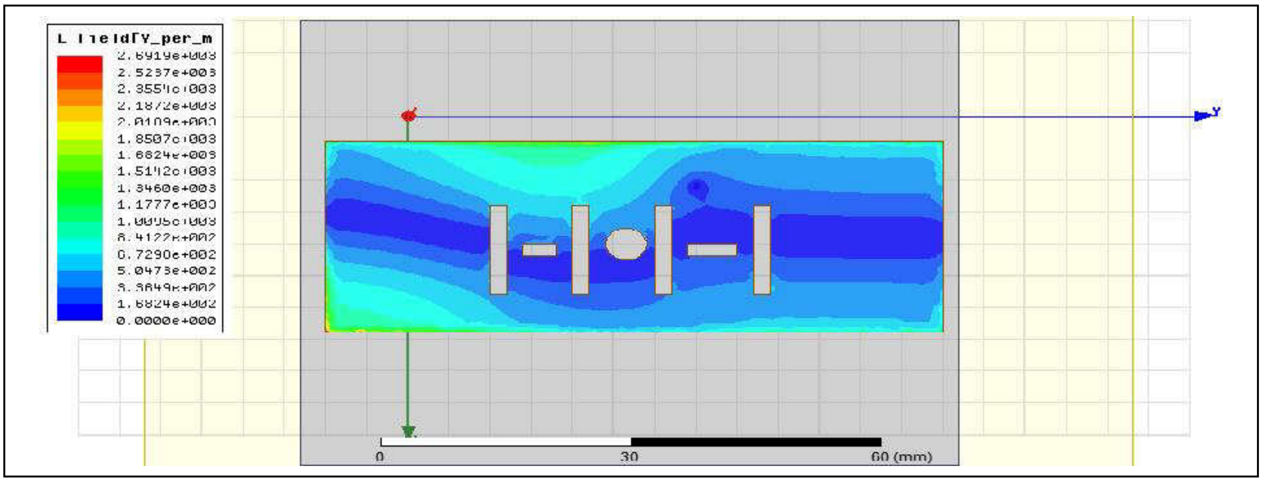
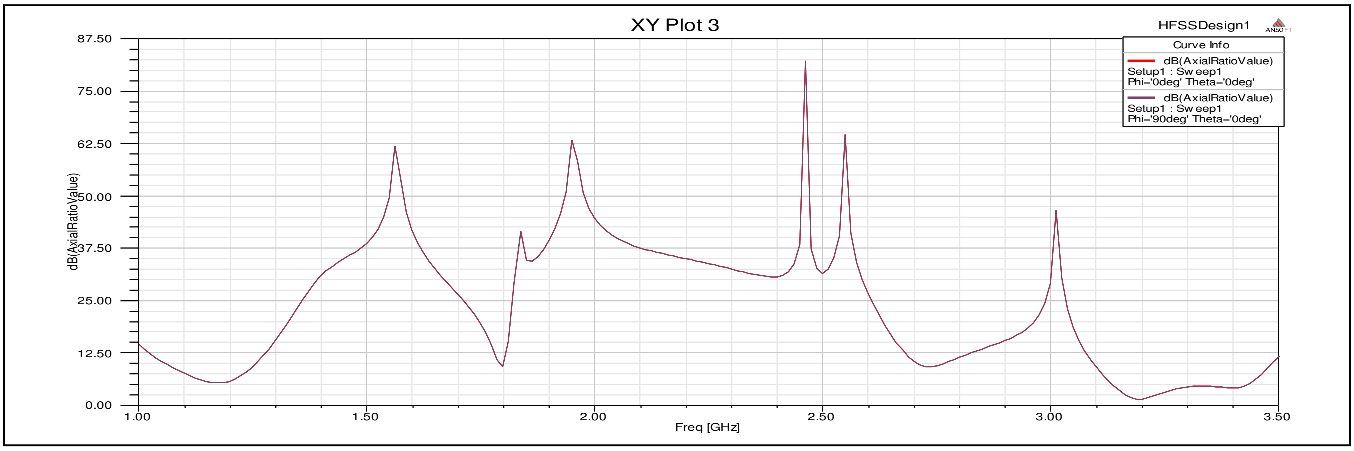


Fig. 6. Electric field distribution for slotted antenna



IV. CONCLUSION

A simple linearly polarized multi-slotted patch antenna has been proposed to provide multi-band characteristics along with good gain of 6.03 dB and is analyzed using finite-element based simulator HFSS. The design is simple in designing and easy to fabricate. Moreover, this antenna has low profile, thus making it good choice for many microwave applications. Also, it has been observed that by changing the locations and dimensions of multiple slots, the antenna can be made to resonate at any other desired resonant frequency.

REFERENCES

[1] Constantine A. Balanis, Antenna Theory & Analysis Design, 3rd ed., John Wiley & Sons, Inc. 811, 2005.

[2] G.A. Deschamps, "Microstrip Microwave Antennas," 3rd USAF Symposium on Antennas, 1953.

[3] E. V. Byron, "A new flush-mounted antenna element for phased array application," in Proceedings of the Phased Array Antenna Symposium, pp. 187–192, 1970.

[4] L. Liu, S. W. Cheung, R. Azim, and M. T. Islam, "A compact circular-ring antenna for ultra-wideband applications," Microwave and Optical Technology Letters, vol. 53, no. 10, pp. 2283–2288, 2011.

[5] S. K. Rajgopal, and S. K. Sharma, "Investigations on ultrawideband pentagon shape microstrip slot antenna for wireless communications," IEEE Transactions on Antennas and Propagation, vol. 57, no. 5, pp. 1353–1359, 2009.

[6] G. Kumar, P.K. Keshri, and B. Sunil, "Wideband Patch Antenna Design for Wi-MAX and WLAN Application with Modified Ground Plane", International Journal of Computer Applications, Vol. 92, No.1, April 2014.

[7] H. Nornikman, F. Malek, N. Saudin, M.M. Shukor, N.A. Zainuddin, M.Z.A.A. Aziz, B.H. Ahmad, and M.A. Othman "Design of Rectangular Stacked Patch Antenna with Four L-Shaped Slots and CPW-Fed for WiMAX Application," 3rd International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME) Bandung, November 7-8, 2013.

[8] Prachi, B. Chauhan, and M. Jaiswal, "Performance Analysis of Different Slotted U, Half U, Pi and Half Pi Microstrip Patch Antennas for LTE Application," International Journal of Computer Applications, Vol. 86, No 18, January 2014.

[9] K.F. Lee, K.M. Luk, K.F. Tong, and Y.L.Yung, "Experimental study of the rectangular patch with a U-shaped slot," International Symposium on Antennas and Propagation Society, Vol. 1, pp. 10-13, Baltimore, USA, 1996.

[10] M.H. Ullah, M.T. Islam, M.R. Ahsan, J.S. Mandeep, and N. Misran, "A Dual Band Slotted Patch Antenna on Dielectric Material Substrate," International Journal of Antennas and Propagation, Vol. 2014, Article ID 258682, 7 pages, 2014. doi:10.1155/2014/258682.

[11] M.R. Zaman, A.U. Ahmed, R. Azim, H. Abdullah, M.T. Ali, and M.T. Islam, "Multiband Patch Antenna for Femtocell Application," International Journal of Antennas and Propagation, vol. 2014, Article ID 352763, 6 pages, 2014. doi:10.1155/2014/352763.

Design and Development of Low Cost, Portable Digital Temperature Sensor

S. Ramasamy
Dept. of ECE
Manipal Institute of Technology
Manipal, Karnataka.
ramasamy.sprm@manipal.edu

***Abstract*—The main objective of this paper is to build a cost effective and accurate temperature sensor with added advantage of being portable which can be placed anywhere to measure the temperature. Incorporating the use of microcontrollers for building a low cost digital temperature sensor makes the device user friendly. The accuracy of the sensor is ensured by the usage of DS1620 which is a digital thermometer with thermostat capabilities. The DS1620 measures temperature using a bandgap based temperature sensor, The use of microcontrollers 89c51 makes displaying the temperature cheap and reliable.**

Keywords—Temperature sensor, Accuracy , Low cost design, Portable, Microcontroller.

I. INTRODUCTION

The EMG signal is a biomedical signal that measures electrical currents generated in muscles during its contraction representing neuromuscular activities. The EMG signal ranges few millivolts depending on desired contraction intensity.The EMG signal ranges between 0 – 10mV depending on desired contraction intensity. The signal itself is a biomedical signal that measures electrical currents generated in muscles during its contraction representing neuromuscular activities. The EMG signal is a complicated signal. Because it acquires noise when it travels through the different tissues in the body. Hence it is so important to ensure the signal which is being analyzed is only the EMG signal itself.

II. MATERIALS AND METHODS

Raw EMG signal offers us valuable information in a particularly useless form. This information is useful only if it can be quantified. The mentioned functions are applied on raw EMG to achieve the actual information of the signal. The development of algorithms in Matlab will be made using online resources for dealing with the different functions needed in writing the signal analysis algorithm. Therefore the online databank physionet provides raw EMG signals of a healthy patient, a patient with neuropathy and a patient with myopathy. Neuropathy is the term for damage to the nervous system, which may be caused by diseases or trauma to the nerve or side-effects of systemic illness. Myopathy is a muscular disease where the muscle fibers do not function for one of many reasons, resulting in muscular weakness. The raw EMG data from: 1) a 44 year old man without history of neuromuscular disease; 2) a 62 year old man with chronic low back pain and neuropathy due to a right L5 radiculopathy; and 3) a 57 year old man with myopathy due to longstanding history of polymyositis, treated effectively with steroids and low-dose methotrexate are considered for analysis.

III. SIGNAL PROCESSING ALGORITHM

The fast Fourier transform is the most common method to determine the frequency spectrum of EMG. It helps to provide efficient and effective ways of understanding the signal and its nature. The signal analysis and characterisation algorithms in Matlab includes the representation of the raw EMG signal as well as the Fast Fourier Transform, the power spectral density (PSD) and the median frequency (MF) of the EMG signal. Generally the Fast Fourier Transform is used to get the spectral distribution of the raw signal. The power spectral density of the EMG signal ranges from 0 to 400 Hz for most muscles. Above this frequency, the frequency components of the EMG signal have amplitude less than 1 uV rms and are no longer distinguishable from the noise of the detection and recording system. Consequently, it is recommended that the general representation of the frequency spectrum be considered to range up to 500 Hz.

One of the parameter of the power density spectrum is the median frequency. Median frequency is the frequency that divides the power density spectrum in two regions having the same amount of power. It has been suggested that the root mean square value of suitably filtered EMG is proportional to muscle tension.

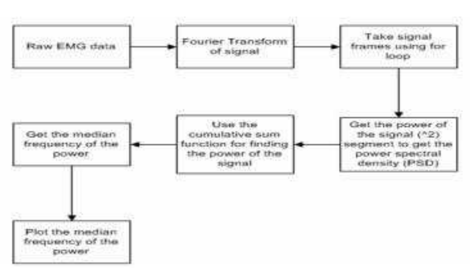


Fig.1.Block diagram of system

The Fourier transform of the raw EMG signal was acquired so as to bring the signal into the frequency domain. The signal appeared as shown in figure 2. The next step was to loop around the signal getting smaller segments or “frames” of whatever number of samples was desired. A simple for loop was used to fulfil this step, with variables used for setting the frames size and the samples size. After the loop was set up, the power spectral density of the signal needed to be found. This was done by simply getting the square of the signal in its frequency form. After this the cumulative sum function was used to get the signal into power form. Then with the signal in its power form the median frequency was got and then plotted using each frame analyzed. This gives a strong visual as to how the power of the muscle is staying strong, or else fatiguing under prolonged use. The final step was to do signal overlapping, so as to get a more detailed look at what was occurring in the muscle.

IV. RESULTS AND DISCUSSION

The first step was taking the Fourier transform of the signal so as to bring the signal into the frequency domain. The power spectral density of the signal needed to be found. This was done by simply getting the square of the signal in its frequency form. After this the cumulative sum function was used to get the signal into power form. Then with the signal in its power form the median frequency was got and then plotted using each frame analyzed. This gives a strong visual as to how the power of the muscle is staying strong, or else fatiguing under prolonged use. The final step was to do signal overlapping, so as to get a more detailed look at what was occurring in the muscle.

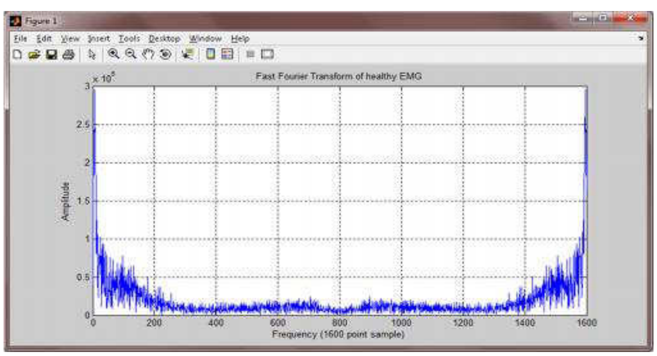


Fig. 2. Fast Fourier Transform of Healthy EMG Signal

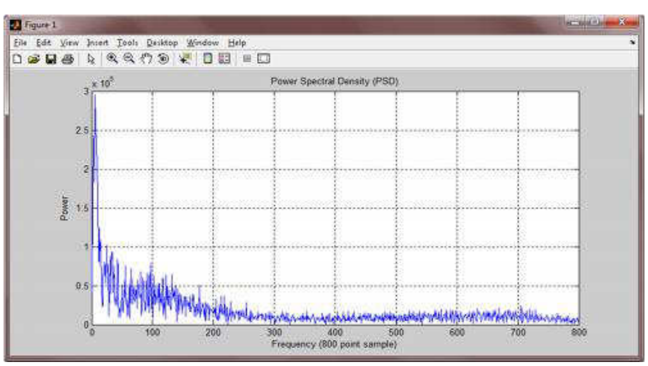


Fig. 3. Power Spectral Density of a Healthy EMG Signal

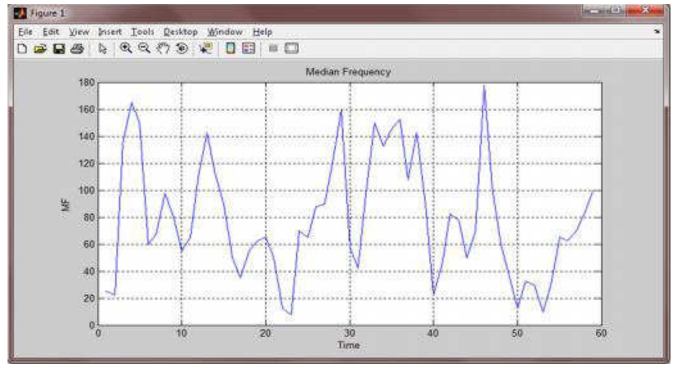


Fig. 4.Median Frequency of Healthy EMG Signal

V. CONCLUSION

The key features of this analysis are the computation and presentation of significant digital signal processing functions - Fast fourier transform, power spectral density and median frequency. Raw EMG offers us valuable information in a particularly useless form. This information is useful only of it can be quantified. Thus they are useful for EMG signal analysis to provide efficient and effective ways of understanding the signal and its nature.

REFERENCES

[1] Meng H., Abel EW., Forster A., Singularity analysis of the EMG interference pattern, 2nd European Medical & Biological Engineering Conference, Vienna, Austria, December 04-08, 2002.

[2] Abel EW., Arikidis NA., Forster A., Inter-scale wavelet maximum-a fine to coarse algorithm for wavelet analysis of the EMG interference pattern. IEEE Transactions on Biomedical Engineering, 49:337-344, 2002.

[3] Abel EW., Arikidis NA., Forster A., Inter-scale local maxima-a new technique for wavelet analysis of EMG signals. 20th Annual Conference of the IEEE Engineering in Medicine and Biology Society, Hong Kong, October 29-November, 1998.

[4] Forster A., Arikidis NA., Abel EW., Wavelet analysis of the interference pattern. Annual Conference of the British Society of Clinical Neurophysiology, Nottingham, 1998.

[5] Abel EW., Zacharia PC., Forster A., Farrow TL., Neural network analysis of the EMG interference pattern, Med. Eng. Phys, 18, 12-17, 1996.

[6] Lakshmi Shaw et.al., Online EMG Signal Analysis for Diagnosis of Neuromuscular Diseases by using PCA and PNN,International Journal of Engineering Science and Technology, Vol 4 ,2012.

Analogue Study of Double Gate SOI FinFET and Tri-gate Bulk MOSFET Structures

Garima Bakshi
Dept. of Electronics and Communication
SET, Ansal University, Gurgaon

Dalveer Kaur
Dept. of Electronics and Communication,
Punjab Technical University, Jalandhar

Vikrant Sharma
Dept. of Electronics and Communication,
CT Institute of Technology, Jalandhar

Amulay Nidhi Puranik
Dept. of EC
Global Centre of Education Excellence, Singapore
garima.bakshi31@yahoo.com

***Abstract*--In this paper we briefly present the impact of systematic variations on transistor performance for the tri gate Bulk MOSFET and double gate SOI Fin FET. This paper compares the variations in threshold voltage, sub threshold swing and drain induced barrier lowering (DIBL) by varying physical dimensions (gate oxide thickness, channel width and channel length) of the devices. It compares the Temperature profile of the devices along the channel .the result obtained from the tri gate bulk MOSFET is more scalable than double gate SOI Fin FET and can be used for better yield and reliability.**

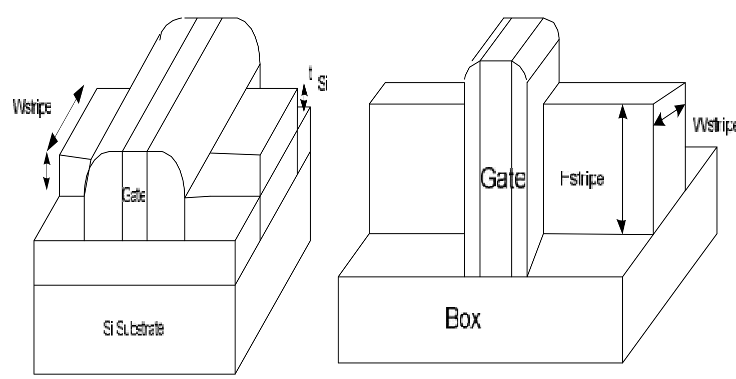
Keywords--DIBL, Reliability, Sub-threshold swing, TCAD, Threshold voltage.

I. INTRODUCTION

The invention of CMOS has revolutionized the VLSI industry and has become a fundamental building block in electronics industry. Due to increased use of these devices has not only boomed the growth of electronic industry but has also gained the attention of researchers to develop more and more compact size devices with better performance. Hence the microelectronics industry has significantly invested in new and better technologies aiming at the fabrication of devices with extremely reduced dimensions. This has led to the design of different types of MOSFET with the aim to improve the characteristics of the device [1]. Over the past four decades the main concern of VLSI has been to combine the increasing number of devices in a single chip without degrading the performance and results. This increase in number has been possible due to scaling down or shrinking of device dimension. This leads to higher packing density and also the manufacturing price lowers down. With the scaling of dimensions the gate oxide thickness and channel length reduces. This reduces the delay in the device at the cost of leakage power. Also an undesirable phenomenon takes place known as short channel effects. The name is given because these effects are seen in devices having channel length (below 100nm) and hence short channel. The predominant features are lack of pinch off and saturation (due to large drain conductance) and large leakage current (a shift in threshold voltage, and therefore VT dependence on drain voltage, due to drain induced barrier lowering (DIBL), and hot carrier effect) [2], [3]. These effects arises occurs because gate loses its ability to control the operation of channel due to proximity of source and drain regions.

Threshold voltage roll off is the decrease in the threshold voltage of the device with the decrease in channel length.DIBL is the reduction in the threshold voltage due to the electric field from the drain towards the source region. To minimize the short channel effects many structures have been proposed till now. The conventional MOSFET with one gate at the top was replaced by dual gate i.e. two gate terminals to control the proper operation of channel. But with the decrease in channel length it was found that double gate MOSFET was also not able to control the operation and changes in width or length changed the threshold voltage considerably. So a better device with trigate i.e. gate on the three sides of channel have been used to counter effect the increase in short channel effect with the decrease in channel length. The double gate SOI FinFET and trigate bulk MOSFET have been compared as candidates to extend transistor scaling to its ultimate limit [4]-[6].

II. DEVICE STRUCTRURE AND SPECIFICATION



(a) Trigate Bulk (b) Double Gate SOI Fin FET

Fig.1. Three dimensional view of MOSFET's

The transistor structures are illustrated in Fig. 1. Sentaurus 3-D device simulations [7] using advanced physical models are performed for the following n-channel MOSET nominal designs: physical gate length L_G=20nm, equivalent gate oxide thickness t_{OX}=9Å, and supply voltage V_{DD}=0.7V. For the trigate bulk MOSFET design, silicon stripe width W_{STRIPE}= L_G and t_{si}=HSTRIPE=0.6 * LG are chosen. For the double gate SOI FinFET WSTRIPE= 0.6 * LG and HSTRIPE =1.1* LG are chosen [8]. For the double gate SOI FinFET the two gates at the side of the channel are operational. The thickness of gate oxide at the top surface is intentionally kept high so that the device behaves as double gate only instead of behaving as a trigate. The optimal design parameters for the double gate SOI FinFET and trigate bulk MOSFET have been illustrated in Table1.

Table I

	Parameter	Double Gate SOI FinFET	Trigate Bulk MOSFET
Design	tSi / LG	1.1	0.6
	LG (nm)	20	20
	tOX (Å)	9	9
	Gate work function(eV)	4.57	4.27
	W/ LG	0.6	1
Performance (VDD=0.7V)	ION (μA) (for IOFF=13nA)	127	85
	DIBL(mV)	62.319	57.038
	Subthreshold Swing(mV/dec)	70.875	70.654

In the tri-gate bulk MOSFET designing, to improve the electrostatic integrity [9] retrograde well doping is used. N-type doping at a concentration of 10^{20} atoms/cm^3 is used for source & drain regions. . The source and drain extension regions are also included to avoid the deeper penetration of depletion region in the channel. The ID-VG characteristics of the optimal designs are shown in Fig. 2.

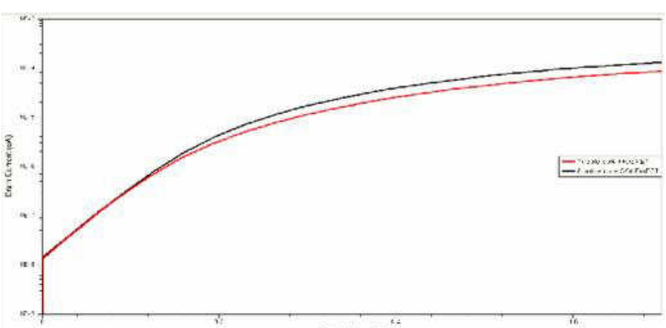


Fig 2. Simulated MOSFET transfer characteristic (I_D-V_G)

The result shows that SOI FinFET has higher current (ION) as compared to the trigate bulk MOSFET.

III. RESULTS AND DISCUSSION

Here in both the devices variation of different parameters with change in dimensions is shown. Also the temperature variation of along the channel of device is shown.

A. Variation in Threshold Voltage (V_T)

Although the threshold voltage of a MOSFET is not a figure of merit for device/circuit performance, it is the most important parameter for MOS device modeling and circuit design. This parameter, which represents the onset of significant drain current flow, has been given several definitions [10]-[12]. Here the threshold voltage has been extracted using the constant current method [13]. The change in threshold voltage with the change in gate thickness oxide is shown in fig 3. For the trigate V_T increases with increasing t_{ox} .Because the gate electrode wraps around the channel region in the trigate design, the amount of depletion charge per unit channel width is smaller. The SOI FinFET design shows different dependence, i.e., as t_{ox} increases V_T decrease due to worse electrostatic integrity. Fig 4 shows the variation of V_T with respect to W_{STRIPE}.

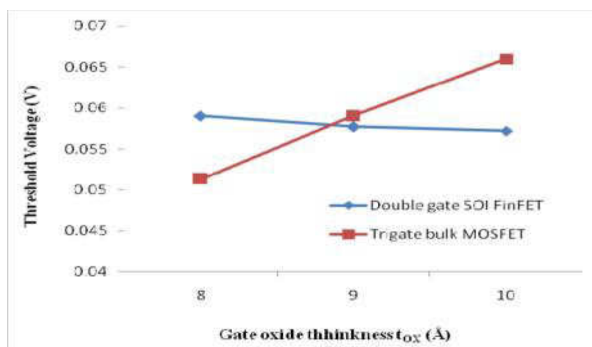


Fig 3. V_T sensitivity to t_{ox} variation

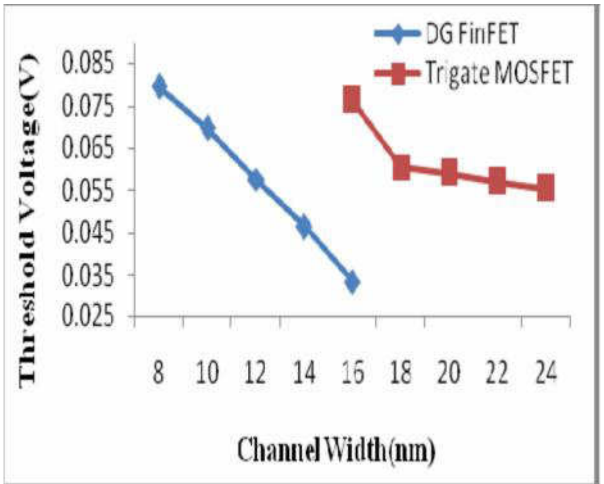


Fig 4. V_T sensitivity to $W_{STRIPE\ variation}$

The graph shows that *VT* changes significantly in SOI FinFET as compared to trigate MOSFET. This happens, because in SOI FinFET as width increases the distance between the two gates increases and the ability to control the channel operation is reduced, while in trigate the gate above channel also helps in operation of channel thereby decreases the sensitivity of *VT* with respect to WSTRIPE. *VT* roll-off phenomenon can be seen in Fig. 5. The result shows that trigate MOSFET is less sensitive to variation *VT* with respect LG when compared to SOI FinFET. Also the threshold voltage is higher in trigate when channel length is scaled down therefore making it a good choice at smaller channel length.

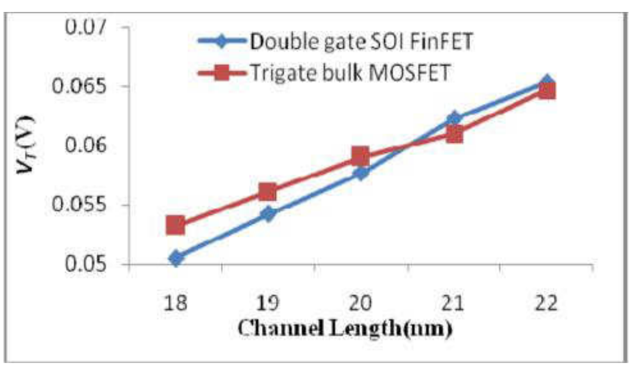


Fig 5. V_T sensitivity to $L_{G\ variation}$

B. Variation in Subthreshold Swing (SS)

Subthreshold swing is defined below threshold voltage as the voltage change required changing the current decade times. Fig. 6 shows the SS change with respect to gate oxide thickness. The results show that subthreshold swing is smaller in trigate bulk MOSFET as compared to double gate SOI FinFET thus making former a better device. Similarly Fig. 7 and Fig. 8 show the change in subthreshold swing with respect to channel width and channel length respectively. Trigate bulk MOSFET shows better subthreshold swing compared to double gate SOI FinFET because the SOI FinFET relies on good gate control(i.e. a combination of ultrathin t_{ox} and narrow W_{STRIPE}) to suppress short channel effect, whereas the trigate

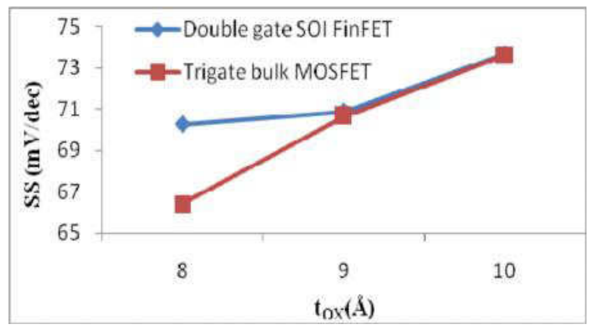


Fig 6. Sub threshold swing sensitivity to t_{OX} variation

bulk design employ retrograde channel doping to help improve electrostatic integrity. The reason for decrease in SS with increase in channel length is that with increase in channel length the gate control over the channel increases hence the leakage current decreases. Also the sub threshold swing is lower in tri gate bulk MOSFET as compared to SOI Fin FET which clearly means that leakage in former device is less.

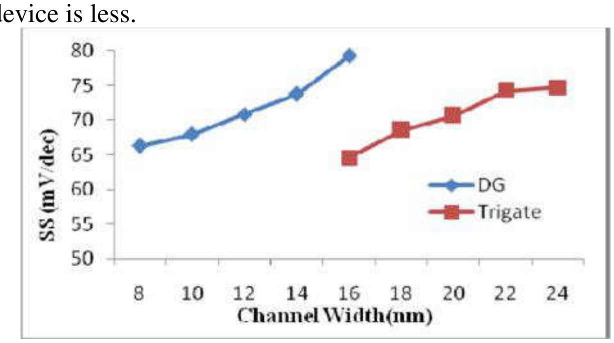


Fig 7. Sub threshold swing sensitivity to W_{STRIPE} variation

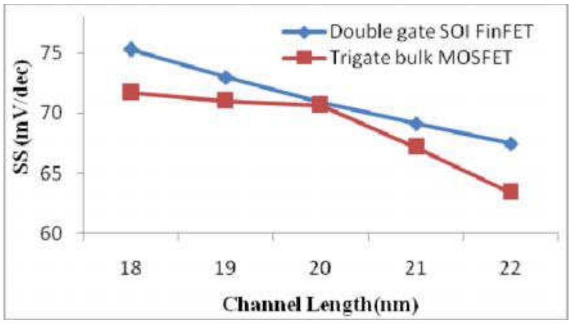


Fig 8. Sub threshold swing sensitivity to L_G variation

C. Variation in Drain Induced Barrier Lowering (DIBL)

DIBL is caused by the encroachment of the depletion region from the drain into the channel. At high drain bias, the depletion region strongly affects the channel potential. As a result, the threshold condition can be reached at a lower gate voltage since the drain has already created a large portion of the depletion region. Strong DIBL is an indication of poor short channel behaviour. However, since the transistor is no operated at a lower drain bias, DIBL itself is not a parameter that is directly related to circuit operation, but rather it is an indication of the degraded device characteristics. As channel length is reduced, the effects of DIBL in the sub threshold region (weak inversion) show up initially as a simple translation of the sub threshold current vs. gate bias curve with change in drain-voltage, which can be modeled as a simple change in threshold voltage with drain bias. However, at shorter lengths the slope of the current vs. gate bias curve is reduced, that is, it requires a larger change in gate bias to effect the same change in drain current[14]. Fig. 9 shows the variation of DIBL with change in t_{OX} variation.

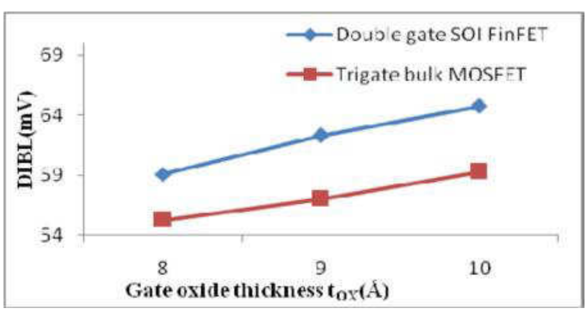


Fig 9. DIBL sensitivity to t_{OX} variation

DIBL variation with respect to t_{OX} is almost the same in both the devices. But the DIBL is lower in tri gate bulk MOSFET when compared to double gate SOI FinFET thereby making the former device more suitable. Fig. 10 and Fig. 11 show the DIBL variation in both the devices with change in channel width and channel length respectively.

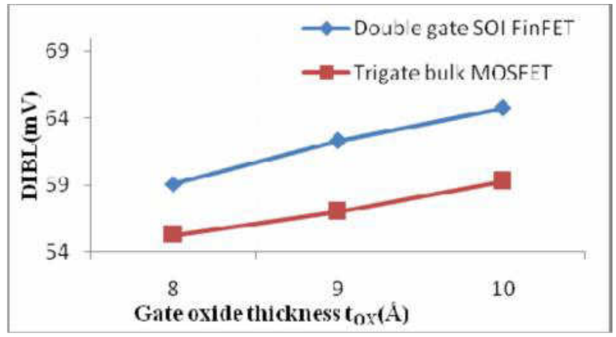


Fig 10. DIBL sensitivity to W_{STRIPE} variation

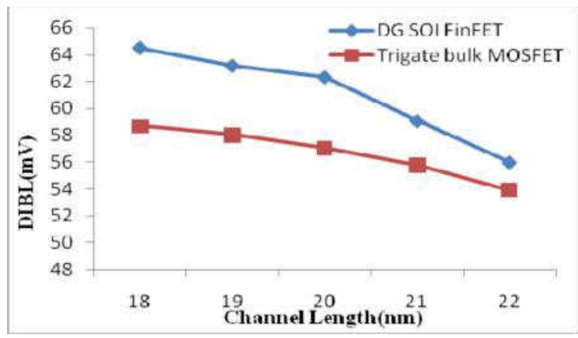


Fig 11. DIBL sensitivity to L_G variation

With increase in channel width the gate loses its ability to control the carrier movement in the channel. This phenomenon occurs more strongly in double gate SOI FinFET when compared to tri gate bulk MOSFET. Thus it can be clearly seen that DIBL variation with respect to WSTRIPE is more sensitive in SOI FinFET and hence makes it vulnerable. As the channel length is increased the distance between drain and source region increases. Thus with the increase in channel length the effect of electric field from the drain region is decreased and hence a reduction in DIBL is observed. At longer channels the DIBL value is nearby in both the devices but show remarkable difference in value when seen at smaller channel length.

D. Temperature Variation

The temperature profile along the channel of both the device is shown in Fig. 12. The peak temperature in SOI device is 448.5K while the peak temperature in bulk device is 426K. This shows that the bulk device is thermally efficient than the SOI device.

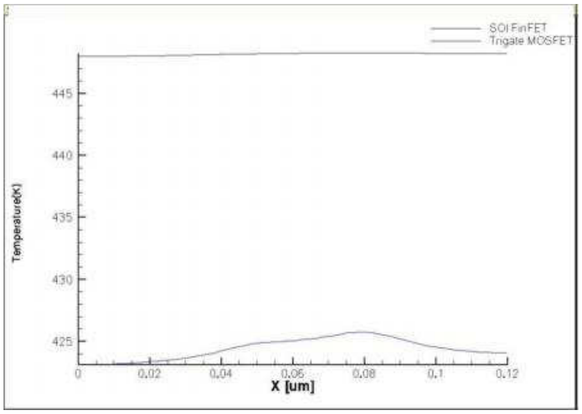


Fig 12. Temperature profile along the channel

The thermal efficiency of the bulk device can be attributed to the presence of window under the channel [15]. Since the temperature is reduced the reliability and overall performance of the bulk device improves.

IV. CONCLUSION

The impact of process induced systematic variation and temperature dependence has been compared for double gate SOI Fin FET and tri-gate bulk MOSFET. Results show that tri-gate bulk MOSFET have higher threshold voltage ,lower sub threshold swing ,drain induced barrier lowering and temperature profile along the channel .thus it appears to be promising device architecture for transistor scaling to the end of the technology roadmap.

REFERENCES

[1] J. P. Colinge "FinFETs and other Multi-Gate Transistors." Springer Science, Bussiness Media, LLC, 2008.

[2] Y. Awano, M. Kosngi, K. Kosemura, T. Mimura, and M. Abe, "Shortchannel effects in subquarter-micrometer-gate HEMT's: Simulation and experiment," *IEEE Trans. Electron Devices*, vol. 36, p.2260, Oct, 1989.

[3] U. K. Mishra, A. S. Brown, and S. E. Rosenbaum, "DC and RF performance of 0.1-mm gate length Al0.48As/Ga0.47In0.53As psedomorphic HEMT," in IEDM Tech. Dig., 1988, p. 180.

[4] D. J. Frank, Y. Taur, and H. S. P. Wong, "Genaralized scale length for two dimensional effects in MOSFETs," *IEEE Electron Device Lett.*, vol 19, no 10, pp. 385-387, Oct, 1988.

[5] R. H. Yan, A. Ourmazd and K F. Lee, "Scaling the Si MOSFET: From bulk to SOI to bulk," *IEEE Trans. Electron Devices*, vol. 39, n. 7, pp. 1704-1710, July 1992.

[6] X. Sun and T.J. King Liu, "Scale length assessment of the tri-gate bulk MOSFET design," *IEEE Trans. Electron Devices*, vol. 56, no. 11, pp. 2840-2842, Nov. 2009.

[7] Synopsys, Inc. Mountain View, CA, Sentaurus User's Manual 2009.06, 2009.

[8] X. Sun, Victor Moroz, Nattapol Damrongplasit, Changhwan Shin, and Tsu-jae King Liu, "Variation Study of the planar ground plane bulk MOSFET, SOI FinFET, and Trigate bulk MOSFET designs," *IEEE Trans. Electron Devices*, Vol. 58, No. 10, Oct., 2011.

[9] A. Hokazano, H. Itokawa, N. kusunoki, I. Mizushima, S. Inaba, S. Kawanaka, and Y. Toyoshima, "Steep Channel & halo profiles utilizing boron-diffusion-barrier layers (Si:C) for 32nm mode and beyond," in VLSI Symp. Tech. Dig., 2008, pp. 112-113.

[10] Ortiz-Conde A, Rodriguez J, Garcia Sanchez FJ, Liou JJ. "An improved definition for modeling the threshold voltage of MOSFETS." Solid- State Electron 1998; 42: 1743-6.

[11] Salcedo JA, Ortiz-Conde A, Garcia Sanchez FJ,Muci J, Liou JJ, Yue Y. "New approach for defining the threshold voltage of MOSFETs." *IEEE Trans Electron Device*s 2001; 48:809-13.

[12] Benson J, D'Hallweyn NV, Redman-White W, Easson CA, Uren MJ, Faynot O, Pelloie JL, "A physically based relation between extracted threshold voltage and surface potential flat band voltage for MOSFET compact modeling." *IEEE Trans Electron Devices* 2001; 48: 1019-21.

[13] Ortiz-Conde A, Garcia Sanchez FJ, Liou JJ, Cerdeira A, Estrada M, Yue Y. "A review of recent MOSFET threshold voltage extraction methods." Microelectronics Reliability 42(2002) 583-596.

[14] Yannis Tsividis "Operational Modeling of the MOS Transistor", Second Edition, New York: Mc: Graw-Hill., 1999, p-268.

[15] Sajad A. Loan, S. Qureshi, S. Sunder Kumar Iyer, "A Novel Partial-Ground-Plane-Based MOSFET on Selective Buried Oxide: 2-D Simulation Study." *IEEE Trans Electron Devices*, Vol. 57, No. 3, March 2010.

Analysis of Two Elements Triangular Patch Antenna Array

Juhi Rai
Dept. of ECE, SLIET Longowal
juhirai78@gmail.com

Anupma Marwaha
Dept. of ECE, SLIET Longowal
marwaha_anupma@yahoo.co.in

***Abstract* – The single element antennas are unable to meet the gain or radiation pattern requirement. Combining single elements of antenna to form an array can be used to resolve this issue. The performance of array is largely dependent on the element spacing and feed methods. In this papertwo element triangular patch array using corporate feeding is optimally designed by varying the element spacing and feed dimensions. The designed antenna resonates at 10 GHz suitable for X band and wireless applications. The simulations were performed on FEM based commercial software HFSS. The gain obtained for /2 spacing is 8.93dB and return loss is of value -13dB whereas for spacing maximum gain is 10.39dB and return loss is -24.5dB. By keeping the substrate thickness constant, the results obtained show the effects of changing the spacing between the elements on gain, directivity and other parameters of the triangular patch antenna.**

Keywords—Equilateral triangular patch antenna, Edge feed, Finite element method, Antenna array, Element spacing, HFSS.

I. INTRODUCTION

Now days various applications where size, weight, cost, performance,robustness, ease of installation (in aerodynamic structures) areof consideration, microstrip antennas are used. The single element patch antennas however have limitations of low power, low gain, poor efficiency, narrow bandwidth.

To overcome these shortcomings antenna size can be increased as larger area results in higher gain. Another way of increasing the dimension of the antenna without enlarging the size of single element, is to gather radiating elements in an geometrical and electrical configuration to form an array. Using array radiation pattern in a specified direction can be obtained by varying the progressive phase shift [1]. Array configurations can be used for variety of applications where high directivity and gain are requirement for example tracking radar, search radar, remote sensing, satellite applications, and many more.

The arrays of patches on a single substrate can be designedin different configurations of microstrip array for achieving high gain, wide bandwidth and improved efficiency. Feeding networks are responsible for voltage distribution among elements of array [2]. For series and corporate feeding array configuration proper impedance matching provides high efficiency antenna array[3]. Corporate feeding configuration modifies power distribution among elements. Steering beam can be obtained by introducing phase change.[4]

Zhaohui Huang, Xing Jiang(2010)designed a directional ultra wideband printed circular monopole antenna array operating in 3.1-10.6GHz.The two elements of the array are fedby a broadband T-junction power divider.The gains are 5.25dB at 3GHz, 7.56dB at 6GHz and 8.6dB at 9GHz[5].

AadeshAryaet al. (2011) described the design of microstrip antenna array with EBG structure to improve return loss, directivity and gain at 5.85GHz resonant frequency.Return loss has been increased in rectangular patch antenna array from -7.8dB to -19dB when EBG structure was applied and in circular patch antennait increases from -7.8dB to -17dB. Directivity has been also improved after applying EBG structure in both antenna from 8.3dB to 8.9dB in case of circular and from 8.26dB to 8.5dB in case of rectangular. The maximum gain has been increased in only circular antenna from 4.7dB to 5.4dB[6]. Among the most common shapes of microstrip patches, rectangular and circular shapes are used because of their simplicity in design whereas triangular shapes patches have the advantage of being physically smaller for operating at a given resonant frequency in comparison to other shapes. The triangular patch shape is considered here for the design of two element array. The performance enhancement is facilitated by optimally selecting the element spacing and the feed width for corporate feeding.

II. MATERIALS AND METHODS

The design is specified with dimension of a single microstrip for triangular patch antenna. The triangular shape patch is designed here to operate at the desired frequency 10 GHz. The corresponding patch antenna model, suspended in air through substrate material Rogers RO 4003, having dielectric constant of $\varepsilon_r = 3.4$ and of substrate is h=100mil shows stable electrical properties at high frequency. The side of triangle is calculated using equation 1 and 2 [7,8,9]. To calculate the wavelength for providing spacing between elements equation 3 is used.The lowest order mode dominant resonant frequency is given by

$$f_r = \frac{2c}{3a\sqrt{\varepsilon_r}} \quad \text{... (1)}$$

Considering the resonant frequency to be 10 GHz, the side of equilateral triangular patch, "a" can be evaluated by

$$a = \frac{2c}{3\sqrt{\varepsilon_r}f_r} \quad \text{...(2)}$$

Where c= velocity of light =3×10^8 m/s

$$a = \frac{2\times 3\times 10^8}{3\times\sqrt{3.4}\times 10\times 10^9} = 1.084\ cm$$

Calculating wavelength for spacing

$$\lambda = {}^{c}/_{f_r} = 3\text{cm} \quad \text{... (3)}$$

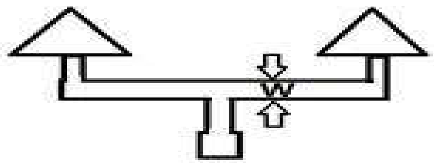


Fig. 1. Two element triangular patch array for /2 and spacing between elements and varying feed width W (i)0.2 cm(ii)0.15cm (iii)0.1 cm

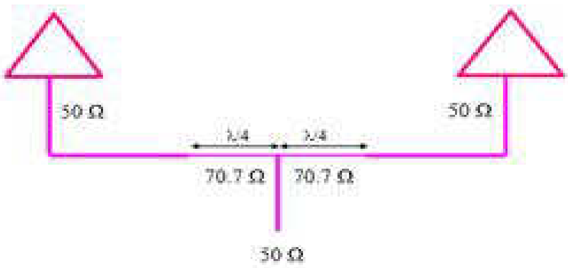


Fig.2. Layout of reactive T-junction used in feed networks

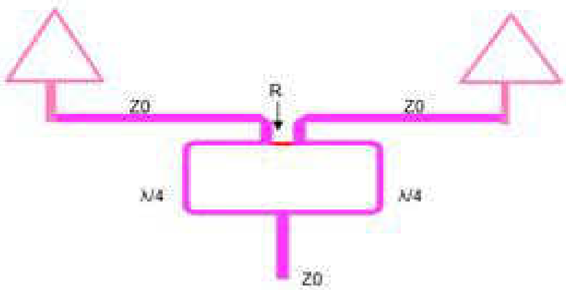


Fig.3. Layout of Wilkinson power divider used in feeding networks

The feed design in the network is not only used for power division. Moreoverbroadening the bandwidth match the impedance of antenna array. Two elements patch are fed by a symmetric dividers and connecting transmission lines (corporate feed network).Feeding networkcan be implemented as part of the matching transmission resonator circuits,this allows double tuning of patch elements without giving up valued space on the radiating surface on the substrate. However, the choice of positions in the feed network for the matching transmission resonator is not unique, the properties of the transmission lines and of their discontinuities need to be simulated very accurately in order to design the matching circuit effectively. The characteristic impedance Z_0 is 50 , so the impedance of the output lines should be 100 . However, the desired output impedance is also 50 , but quarter-wave transformers are used to get the 100 impedances back to the characteristic impedanceZ_0. The width of the lines is inversely proportional to their impedanceand the impedance of the quarter wave lines is therefore 2×Z0, which is 70.7 as given in Fig.2.

III. DESIGN OF PATCH ANTENNA ARRAY

This section presents the numerical results with finite element method (FEM), simulations being carried out by using commercial software Ansoft HFSS v.13 for microstrip line edge fed equilateral triangular shaped patch array.For 3D volumetric passive device modeling it is a high-performance electromagnetic (EM) field simulator. In this paper, microstrip line feed techniqueis used because of ease of fabrication, robustnessand simplicity in modeling as well as impedance matching which makes it suitable for severe environmental conditions such as outer space. The width of microstrip feed is smaller as compared to the patch andcan be etched on the same substrate to provide a planar structure.

The corporate-feed is used to provide power splitting of two element network. This can be done by using either tapered lines or using quarter wavelength impedance transformers. In this paperthe patch elements are connected by using the quarter wavelength impedance transformer. Corporate-feed arrays are general and versatile. This method has more control of the feed of each element and is ideal for scanning phased arrays and multiband arrays. It improves directivity as well as radiation efficiency and reducesthe beam fluctuations over a range of frequencies.Theamplitude can be varied using either amplifiers or attenuatorswhereas phase of elements can be adjusted by using phase shifters.

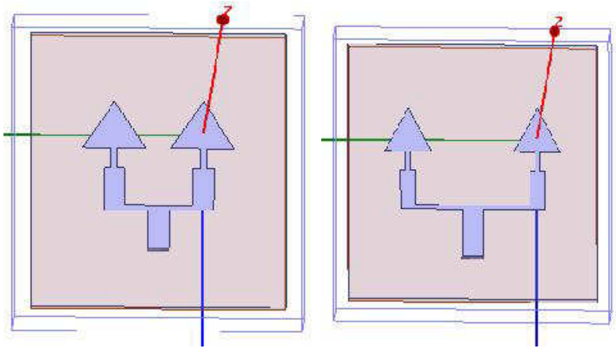


Fig. 4. HFSS model of triangular patch array for(a) /2 spacing, (b) spacing.

IV. RESULTS AND DISCUSSION

Analysis for /2 two elements of array spacing between keeping the other parameters constant and varying the feed width is done,different values of S_{11} are obtained and the return loss curves are plotted in Fig. 5. Among all the feed widths, the perfect matching condition is obtained at W=0.15cm and the resulting value of gain is at peak maxima 8.9dB is shown in fig. 6 and return loss plot obtained has a value of S_{11}-13.09dB at resonating frequency 8.45GHz whereas other feeding widths give relatively smaller valuesof gain.

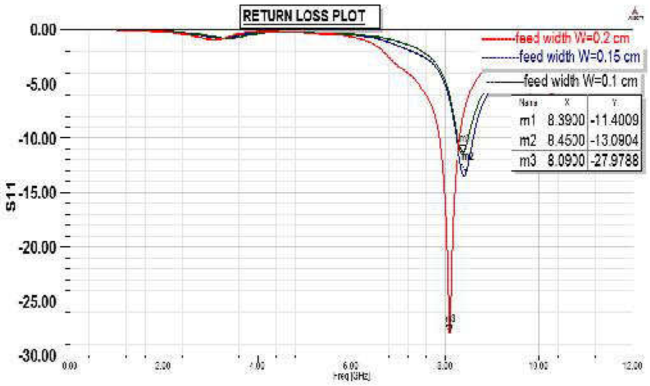


Fig. 5.Return loss plot for /2 spacing between elements and varying feed widths 1)0.2 cm(2)0.15cm(3)0.1 cm

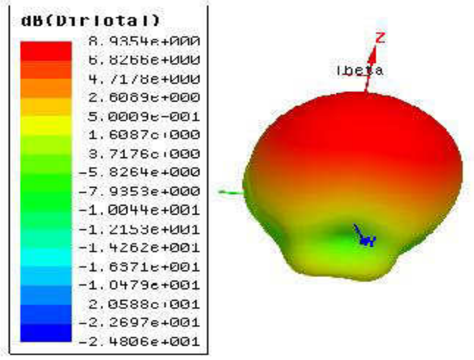


Fig. 6.Directivity plot for /2 spacing at feed width 0.15 cm

Table.1 shows the relative comparison with various feed widths.Array configurations improve gain, bandwidth and radiation efficiency. The resonating frequency is 8.45GHz at 0.15 cm feed width.

Table I. The comparative analysis of different feed widths for triangular patch antenna for /2 spacing.

Feed width	Gain	Return loss	Resonant frequency
0.2 cm	7.3 dB	-22.5 dB	8.09 GHz
0.15cm	8.935 dB	-13.5 dB	8.45 GHz
0.10cm	8.6987 dB	-11.5 dB	8.39 GHz

The spacing between two elements in array should be greater than /2 and less than . As we increasethe spacing between elements then results show the higher directivity and reduced cross polarization or mutual coupling. Two element spacing array can be optimized by varying the feed width of corporate feed network.As feedingnetwork can be implemented as part of the matching transmission resonator circuits.

Feed width	Gain	Return loss	Resonant frequency
0.2 cm	8.6164 dB	-27 dB	9.1 GHz
0.15cm	9.586 dB	-25.5 dB	10.1 GHz
0.10cm	10.329 dB	-24.5 dB	10.3 GHz

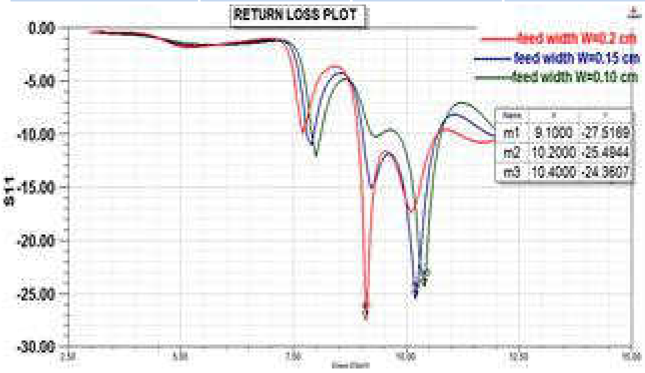


Fig. 7.Return loss plot for spacing between elements and varying feed widths 1)0.2 cm(2)0.15cm(3)0.1 cm

For spacing between two elements of array keeping other parameters fixed and varying the feed width, different values of $S_{11}$are obtained and the return loss curves are plotted.Among all the feed widths, the fig. 7 shows the return loss, S11 plot for spacing between elements shoots down to a maxima of -24.5dB for W=0.1cm at resonant frequency of 10.3 GHz indicating excellent matching conditions and the resulting value of gain is at peak maxima 10.32 dB is shown in fig. 8. The bandwidth is almost 1 GHz, other feeding widths give relatively smaller values gain.

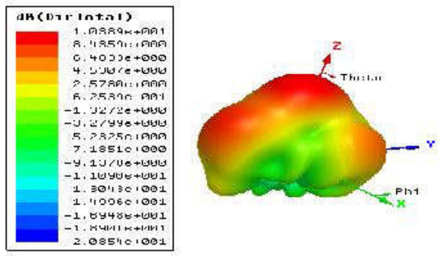


Fig. 8.Directivity plot for spacing at feed width 0.1 cm

The element spacing of the array is initially selected i.e. and the optimization of the patch antenna array is then done by varying the feed width and keeping the dimensions of quarter wave transformer to a calculated fixed value. Table.2 shows the relative comparison of varying feed width.

Optimization two element array configuration can be performed by varying the element spacing and secondly by varying the feed width of the matching network.

Table II. The comparative analysis of different feed widths for triangular patch antenna for spacing

IV. CONCLUSION

FEM based HFSS software has been used here for designing, simulation and analysing the performance of microstrip fed triangular patch antenna array for varying element spacing and different feed widths. The microstrip patch antennas are the most popular antennas for Integrated Radio Frequency systems as they are compatible with microwave integrated circuits. Among all the feed widths and element spacing, the perfect matching condition is obtained at W=0.15cm and for /2 spacing between elements. The maximum gain in this case is 8.9 dB and the return loss of -13.5dB. For spacing the perfect matching condition is obtained at W=0.1 cm and the value of maximum gain is 10.3dB, return loss is -24.5 dB, with a bandwidth of 1 GHz is achieved at optimized feed width. It can be concluded that selection of proper feed width and element spacing are the key features in achieving impedance matching of the circuit, desired efficiency, gain, directivity.

REFERENCES

[1] Constantine A.Balanis, "Antenna Theory- Analysis and Design", 3rd ed., John Willey & Son, Inc, USA, (2005).

[2] H. Cheng-Chi, et al., "An aperture-coupled linear microstrip leaky-wave antenna array with two-dimensional dual-beam scanning capability," Antennas and Propagation, IEEE Transactions on, vol. 48, pp. 909-913, 2000.

[3] K. Gi-Cho, et al., "Ku-band high efficiency antenna with corporate-series-fed microstrip array," in Antennas and Propagation Society International Symposium, 2003. IEEE, 2003, pp. 690-693 vol.4.

[4] A. Abbaspour-Tamijani and K. Sarabandi, "An affordable millimeter-wave beam-steerable antenna using interleaved planar subarrays," Antennas and Propagation, IEEE Transactions on, vol. 51, pp. 2193-2202, 2003.

[5] Zhaohui Huang, et al., "Design of a Directional Ultra-Wideband Planar Antenna Array''6/10/2010 IEEE

[6] Aadesh Arya, et al., "Performance Comparison between Rectangular & Circular Patch Antenna Array with EBG Structure", International Journal of Applied Engineering Research, Vol.7, No.11, 2012 .

[7] R. Garg, et al., "An improved formula for the resonant frequency of triangular microstrip patch antenna", IEEE Transaction on Antennas Propagation, Vol. AP-36, p. 570, Apr. 1988.

[8] J. S. Dahele, et al., "On the resonant frequencies of the triangular patch antenna", IEEE Transaction on Antennas Propagation, Vol. AP-35, pp.100- 101, January 1987.

[9] K. Guney, ''Resonant Frequency of a triangular Microstrip Antenna" Microwaveopt.Technol.Lett. , Vol6.No9. PP 555-557, 1993

Published by

CT EDUCATIONAL SOCIETY

CT INSTITUTE OF ENGINEERING MANAGEMENT & TECHNOLOGY

SHAHPUR CAMPUS

UE II-Pratappura Road, Shahpur,
Jalandhar-144020 I Ph. : +91-181-5055127-28
Cell : +91-99145-04904

MAQSUDAN CAMPUS

Greater Kailash, G.T. Road, Maqsudan,
Jalandhar-144008 Ph. : +91-181-5009605-595
Cell : +91-98146-82308

NEW DELHI OFFICE

74, World Trade Centre, Barakhamba Lane,
New Delhi- 110001 Tel. 011-41527941
Cell : +91-70421-22004

CORPORATE OFFICE

Basement Seiyu Complex, Near Model
Town Gurudwara, Jalandhar-144003
Ph. : +91-181-2440713 Fax: +91-181-2440726
Cell : +91-99140-04901

ISBN : 978-81-929077-0-3

9 788190 719667