ISBN : 978-81-924867-0-3

AMRITSAR COLLEGE OF ENGINEERING & TECHNOLOGY
12 Km Stone,Amritsar-Jalandhar G.T Road,
Amritsar, Punjab. INDIA
Phone : 0183-5010423, Fax : 0183-5069535
e-mail : contact@acetamritsar.org • Website : www.acetamritsar.org

National Conference on Research and Innovations in Engineering and Technology (RIET-12)

Editors
DR. V.K. BANGA
PROF. SANJEEV KUMAR

NATIONAL CONFERENCE

On

RESEARCH AND INNOVATIONS IN ENGINEERING AND TECHNOLOGY

17th – 18th August 2012

(RIET-12)

Organized By
Deptt. of Electronics and Communication Engineering
& Electrical and Electronics Engineering

PROCEEDINGS

Editors
DR. V.K. BANGA
PROF. SANJEEV KUMAR

AMRITSAR COLLEGE OF ENGINEERING & TECHNOLOGY
(NBA & NAAC Accredited)

Sponsored By

Department of Science & Technology, New Delhi

Punjab Technical University, Jalandhar

National Conference

on

Research and Innovations in Engineering and Technology (RIET-12)

17th-18th August, 2012

Editors

Dr. V. K. Banga

Prof. Sanjeev Kumar

Organized by:

Department of Electronics and Communication Engineering

&

Department of Electrical and Electronics Engineering

Amritsar College of Engineering and Technology,

Amritsar, Punjab.

National Conference

on

Research and Innovations in Engineering and Technology (RIET-12)

17^{th}-18^{th} August, 2012

Chief Patrons

Sh. H. L. Sharma

Sh. Amit Sharma

Patron

Sh. M. S. Aujla

Conference Chair

Dr. V. K. Banga

Organizing Secretary

Er. Sanjeev Kumar

Organizing Committee

Er. Sandeep Kaushal

Er. Narinder Sharma

Er. Bimal Kumar

Er. Sanjeev Soni

Er. Namrata Kad

ADVISORY COMMITTEE MEMBERS

Dr. Rajneesh Arora	**VC PTU Jalandhar Chairman Advisory Committee**
Dr. G. K. Singh	**IIT Roorkee**
Dr. K. K. Pant	**IIT Delhi**
Dr. S. N. Singh	**IIT Kanpur**
Dr. S.S. Pattnaik	**NITTTR Chandigarh**
Dr. P. S. Bhimbhra	**TU, Patiala**
Dr. Yaduvir Singh	**NIET, Greater Noida**
Dr. Rajesh Kumar	**TU, Patiala**
Dr. Dilbag Singh	**NIT Jalandhar**
Dr. M.L. Singh	**GNDU Amritsar**
Dr. Onkar Singh	**HBTI Kanpur**
Dr. Sandeep Yadav	**IIT Rajasthan**
Dr. R. N. Yadav	**MANIT Bhopal**
Dr. Ajay K Saxena	**DEI Agra**

CONFERENCE ORGANIZING COMMITTEES

RIET-12

Editorial Committee

Prof. Namarta Kad

Er. Neha Sharma

Er. Simardeep Kaur

Er. Gurjit Singh

Ms. Meenu Sharma

Publication Committee

Prof. Sandeep kaushal

Er. Keshav Sood

Er. Gaurav Soni

Registration and Reception Committee

Prof. Sanjeev Soni

Er. Atul Mahajan

Er. Guneet Kaur

Er. Simarjeet Kaur

Hospitality Committee

Prof. Bimal Kumar

Er. Neelakshi Gupta

Er. Simardeep Kaur

Er. Sunny Malhotra

Er. Ajay Pal Singh

Media, Publicity and Prize distribution Committee

Prof. Narinder Sharma

Er. Pooja Vij

Er. Damanbir Singh

Er. Jagdeep Singh

Er. Reetu Pathania

Mr.Ashwani Kumar

Tours and Transportation Committee

Er. Raminder Pal Singh

Er. Arvinderjit Singh

Mr. Gurpartap Singh

Mr. Paramjit Singh

MESSAGE

I am glad to know that a National Level Conference on “Research and Innovations in Engineering and Technology (RIET-12)” is being organized by Electronics and Communication Engineering and Electrical and Electronics Engineering Department of Amritsar College of Engineering and Technology on 17^{th}- 18^{th} August, 2012.

This is indicative of the achievements of the team concerned not only in academics but also in various other fields such as research and developments, extracurricular activities, project developments and the list is much more long.

Technology in world is going through a tremendous flux. The Conference aims at bringing together Researchers, Scientists, Academicians and many others to interact and exchange their experiences, which in turn would help the students and faculty to gain immensely at the Professional front. I believe that the conference will serve its objectives and will provide the necessary impetus in setting forth a dynamic process of continuous interaction amongst the researchers, teachers and professionals of the country.

I also take the opportunity to thank the Electronics and Communication Engineering and Electrical and Electronics Engineering Department for organizing this event and convey my greetings to the organizers and wish a grand success in the endeavor for the event.

H. L. Sharma
(Sr. Advocate)
Chairman
ACET Amritsar

MESSAGE

It gives me immense pleasure to note that a National Level Conference on “Research and Innovations in Engineering and Technology (RIET-12)” is being organized by Electronics and Communication Engineering Department of Amritsar College of Engineering and Technology on 17^{th}- 18^{th} August, 2012. It is a timely step in sharing new innovations and transmitting the same to its consumers at a very fast pace. Learning is a continuous and unending process and such conferences unveil the treasure of knowledge. It creates the talent, creativeness and professional skills of an individual.

This Conference will spread the light of awareness about the latest and upcoming fields and research areas in Engineering and Technology. While praying for the institute to accomplish its mission, I send my best wishes and congratulate the students and the staff of the department for organizing this event.

I am very confident that, we all are going to gain a highly rewarding and learning experience through this conference. I further extend warm wishes and good luck to this endeavor.

I on behalf of ACET, Amritsar wish all academicians and professionals for fruitful interaction and future direction in this significant field. I congratulate the Electronics and Communication Engineering and Electrical and Electronics Engineering Department for a organizing this Conference.

Amit Sharma
(Sr. Advocate)
Managing Director
ACET Amritsar

MESSAGE

On Behalf of Amritsar College of Engineering and Technology, I welcome all the participants to National Level Conference on "Research and Innovations in Engineering and Technology (RIET-12)".

The main aim of our Institute is to educate young minds to be equipped for the future Challenges. Our mission is to groom students in accordance with latest Technologies. I am proud to say that our Institute is working tirelessly to make students ready to march towards these challenges. I wish that this conference would serve our agenda for improvements in our teaching curriculum so that students successfully match the market requirements in the global scenario.

I am sure that this conference will pave the way for providing a forum to the researchers, academicians and students to express their innovative and creative research skills. This event will spread the light of awareness about the latest and upcoming fields and research areas in Engineering and Technology.

I pray for the institute to accomplish its mission. I send my best wishes and congratulate the students and the staff of the department for organizing this event. I also wish that this type of conference should be a routine feature.

Dr. M. S. Aujla
Principal

MESSAGE

It is a matter of great pleasure for me that Department of Electronics & Communication Engineering and Electrical and Electronics Engineering of our college is organizing a National Conference on “Research and Innovations in Engineering and Technology (RIET-12)” on 17-18th Aug. 2012.

It is outcome of indeterminable efforts put by the organizers in planned manner within a stipulated period. This conference is an effort to meet the surging technological challenges in topical themes areas of Communication Technologies, Computational Techniques, Security Advances & other related areas, which is very important for national development.

The Conference will provide the ideal forum to stimulate ideas and establish collaborations and to initiate intense discussions about latest developments in the area of Science and Technology.

It is not an easy job for the organizers of this national conference of such a magnitude, and I would like to thank all the members of the organizing committee for their hard work.

I wish the conference a great success.

Dr. S. K. Aggarwal
Dean Academic Affairs’

MESSAGE

I feel Proud that the Department of Electronics and Communication Engineering and Electrical and Electronics Engineering department are organizing a National Level Conference on “Research and Innovations in Engineering and Technology (RIET-12)” on 17^{th}-18^{th} August, 2012.

Electronics is now very vast fields in which new developments are occurring almost on daily basis. These developments are changing the shape of the society. They are useful to all people of the world. Considering the fast pace of globalization and dynamic nature of Trends in Engineering and Technology, the need to hold such conferences from time to time is fully justified and confident that the conference will provide a platform where experts on the subject will share their innovative ideas.

The conference received an overwhelming response from various professionals across the country. We had received more than 130 papers covering various areas like Communication Technologies, Computational Techniques, Security Advances and other related topics. The papers received in this conference have been reviewed by reviewer committee and editorial board depending upon the subject matter of the paper. After the review process, the submitted papers were selected on the basis of originality, significance, clarity for the objective of the conference.

I am confident that the deliberations during the Conference will enrich all the participants in particular and institute in general. This will become an occasion for academician, professionals, participants and students to acquire latest knowledge in their respective fields.

I hope that the sincere efforts, zeal and vigor of the members of Electronics and Communication Engineering and Electrical and Electronics Engineering Department of ACET would be prolific enough in making this event a grand accomplishment.

I wish this Conference a grand success.

Dr. V. K. Banga
Conference Chair
RIET-12

MESSAGE

It is a matter of great pleasure for me that our department is organizing a National Level Conference on "Research and Innovations in Engineering and Technology (RIET-12)" on 17th-18th August, 2012. This Conference is going to be an important event in the process of unfolding many new ideas.

This Conference aims at exploiting the ideas through enlightened deliberations and their fruitful implementation both in industries and academics curricular activities. Experts from various centers will deliver keynote lectures during the conference for value addition to the professional skills of conference delegates.

I believe that organizing such events is a must for enabling professionals, scientists, scholars and educators to analyze the future needs and to keep themselves updated with the latest advances in the field of science and technology. This event too is an effort to meet the surging technological challenges in Research & Development areas, which is very important for National development.

I also take the opportunity to thank the Management, Principal and staff of ACET for their encouragement and continuous support without which this event would not have taken shape.

I extend my personal kudos to the all organizers of this National Conference and wish them to make this Conference and event, very hard to forget and long to cherish.

With Best Wishes.

Sanjeev Kumar
Organizing Secretary
RIET-12

INDEX

National Level Conference on Research and Innovations in Engineering & Technology (RIET-12)
August 17-18, 2012, Amritsar College of Engineering & Technology, Amritsar, Punjab

A. Communication Technologies

B. COMPUTATIONAL TECHNIQUES

C. SECURITY ADVANCES

Performance Analysis of Space Time Block Code

Reetu[1] and Sanjeev Kumar[2]
[1]Lecturer, Department of Electronics & Communication Engineering, ACET, Amritsar, India
[2]Associate Professor, Department of Electronics & Communication Engineering, ACET, Amritsar, India
Email: reetu.gaggel1204@gmail.com, sanjeev_be@indiatimes.com

***Abstract* — To combat multipath fading, Space Time Block Code is widely used in wireless communication systems using multiple transmit and receive antennas. Data is encoded using a space time block code, and the encoded data is split into n streams which are simultaneously transmitted using n transmit antennas. The received signal at each receive antenna is a linear superposition of the n transmitted signals perturbed by noise. In this paper we will analyze the performance of Space Time Block Code. The effect of channel estimation is also shown in this paper.**

***Index Terms* – STBC, MIMO, OFDM.**

I. INTRODUCTION

Today wireless system requires high voice quality and high bit rate data services. Also the communication units are remote and light weighted, and they have to be operating in various environments reliably. So in other words the wireless systems are supposed to have better quality and coverage, be more power and bandwidth efficient, and be deployed in diverse environments. But in most situations, the wireless channel suffers from attenuation due to destructive and constructive addition of multipath in the propagation media and to interference from other users.

The channel statistic is significantly often Rayleigh which makes it difficult for the receiver to reliably determine the transmitted signal unless some less attenuated replica of the signal is provided to the receiver. This technique is called diversity, which can be provided using temporal, frequency, polarization, and spatial resources. This means that severe attenuation in a multipath wireless environment makes it extremely difficult for the receiver to determine the transmitted signal unless the receiver is provided with some form of diversity, i.e., some less-attenuated replica of the transmitted signal is provided to the receiver. The receiver diversity plays a crucial role in wireless communication. Time and frequency diversity are commonly used diversity schemes. Time interleaving with error correction coding can provide diversity improvement. However, time interleaving results in larger delays when the channel is slowly varying. In many scattering environment, antenna diversity is used to reduce the effects of multipath fading. The classical approach is to use multiple antennas at the receiver and perform combining or selection and switching in order to improve the quality of the received signal. The major problem with using the receive diversity approach is the cost, size, and power of the remote units. The use of multiple antennas and radio frequency (RF) chains (or selection and switching circuits) makes the remote units larger and more expensive.

In some applications, the only practical means of achieving diversity is deployment of antenna arrays at the transmitter and/or the receiver. However, considering the fact that receivers are typically required to be small as it is not easy to establish multiple antennas at the remote stations. This leads to the transmit diversity. Transmit diversity has been use as a method of combating fading in various wireless communication channels. This transmit diversity technique improves the signal quality at the receiver on one side of the link by simple processing across two transmit antennas on the opposite side. This transmit diversity technique improves the error performance, data rate of wireless communication systems. Alamouti discovered remarkable scheme for transmission using two antennas [1]. Space–time block coding, introduced in [2], generalizes the transmission scheme discovered by Alamouti to an arbitrary number of transmit antennas and is able to achieve the full diversity promised by the transmit and receive antennas. These codes retain the property of having a very simple maximum likelihood decoding algorithm based only on linear processing at the receiver [2]. Space-time block codes can be constructed for any number of transmitted antennas. For specific case of 3 and four antennas, these diversity schemes were improved to provide ¾ of possible transmission rate.

II. SPACE TIME BLOCK CODING (STBC)

It is assumed that there are $\boldsymbol{N}$ transmit antennas and M receive antennas in a wireless communication system in which STBC is used. The input source data bits are firstly modulated, and then carried into a space-time block encoder. Mapping from the modulated symbols to a transmission matrix, which is completed by the STBC encoder, is a key step in STBC systems. The input symbols of the encoder are divided into groups of several symbols. The number of symbols in a group is according to the number of transmit antennas and mapping rule. A PXN transmission matrix means there are N transmitting antennas and P time slots. Different symbol columns are transmitted through different antennas separately and different symbol rows in different time slots. For example, the encoded symbol of column i and row f should be transmitted through the ith antenna in the fth time slot.

A block of m binary symbols Ct= Ct1, Ct2.....Ctm is fed into ST block. The ST block maps the block of m binary symbols into nt modulation symbols from signal set of M=2m points Xt=(Xt1,Xt2,...... Xtnt). The nt parallel outputs are simultaneously transmitted by different antennas.

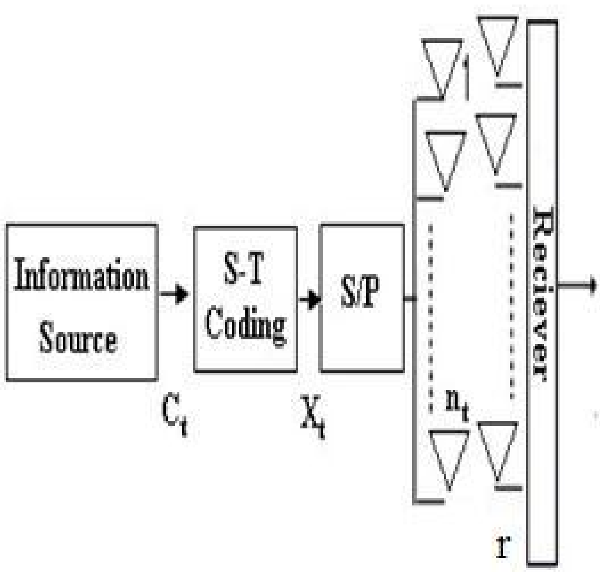


Fig. 1: Wireless System with STBC

A simple space time codes was suggested by Siavash M. Alamouti in his landmark paper [1]. This paper offers simple methods to achieve spatial diversity with two transmit antennas and one receive antenna. The encoding of signal is done in space and time (space–time coding). The encoding, however, may also be done in space and frequency. Instead of two adjacent symbol periods, two adjacent carriers may be used (space–frequency coding).

Consider we have a transmission sequence e.g. x_1, x_2, x3.....x_n. In normal transmission, we send x_1 in first time slot, x_2 in second time slot and so on. However, Alamouti suggested that group the symbols into the group of two. In the first time slot send x1 and x2 from first and second antenna and –x2* and x1* from first and second antenna at second time slot. In third time slot x3 and x4 from first and second antenna and there conjugates in fourth time slot and so on.

Now the signal is transmitted through various channels. The channel experience by each transmit antenna is independent from the channel experienced by another antennas. For ith transmitted antenna, the transmitted symbol is multiplied by hi i.e. Rayleigh channel coefficient. The channel experienced between each transmit to receive antenna is randomly varying in time. However the channel is assumed to remain constant over two time slots. G2 represents a code which utilizes two transmit antennas and is defined by:

$$\mathcal{G}_2 = \begin{pmatrix} x_1 & x_2 \\ -x_2^* & x_1^* \end{pmatrix}$$

RECEIVER WITH ALAMOUTI STBC

In the first time slot, the received signal is

y1=h1x1 + h2 x2 + n0

In second time slot the received signal is

y2=-h1 x2* + h2 x1*+n1

COMBINING SCHEME

x1’=h1* y1+h2*y2

x2’=h2*y1+h1*y2

These combined signals are then sent to maximum likelihood detector. There are many applications where higher order of diversity is needed and multiple receive antennas at the remote unit is feasible. In such cases, it is possible to provide a diversity order of 2M with two transmit and M receive antenna.

III. PERFORMANCE ANALYSIS

(A) Performance of Almouti code

Alamouti shows the BER performance of encoded coherent BPSK for MRRC and the new transmit diversity scheme in Rayleigh fading.

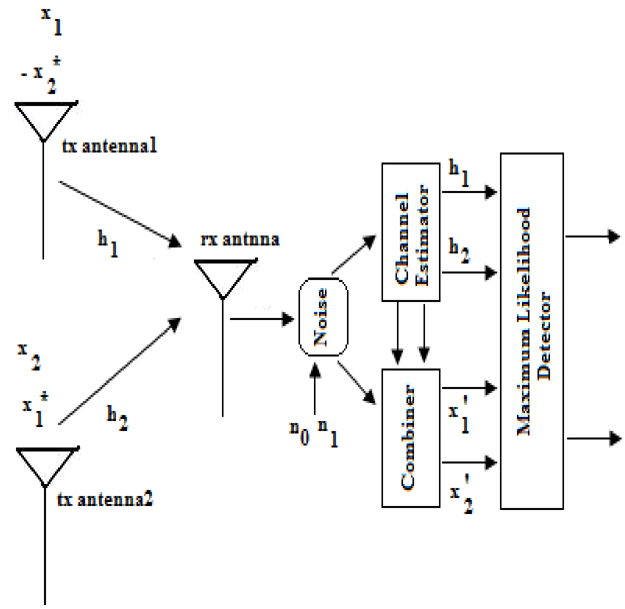


Fig 2: Simple 2Tx diversity scheme with 1Rx

It is assumed that the total transmit power from the two antennas for the new scheme is the same as the transmit power from the single transmit antenna for MRRC. We assume that the receiver has perfect knowledge of the channel. Alamouti shows that the new scheme with two transmitters and a single receiver is 3 dB worse than two branches MRRC. This 3dB penalty is incurred because this scheme requires the simultaneous transmission of two different symbols from two antennas. As the antenna radiates limited power and in order to have the same total radiated power from two transmit antennas the energy allocated to each symbol should be halved. This results in the 3dB penalty in error performance. But if there is no power limitation problem then the total radiated power may be doubled then no performance penalty is incurred.

(B) *Space Time Block Code with channel estimation*

The most effective technique to mitigate multipath fading in a wireless channel is knowledge of channel condition. If channel experienced by the receiver on one side of link are known at the transmitter on the other side, the transmitter can predictor the signal in order to overcome the effect of the channel at the receiver. In this performance analysis, we study the performance of Alamouti scheme with two receive antennas with and without channel estimation. In the realistic scenario where the channel state information is not known at the receiver, this has to be extracted from the received signal. We assume that the channel estimator performs this using orthogonal pilot signals that are pretended to every packet [3]. It is assumed that the channel remains unchanged for the length of the packet (i.e., it undergoes slow fading).

Figure 3 shows that with 8 pilot symbols for each 100 symbols of data, channel estimation causes about 1dB degradation in performance for the selected Eb/No range. This improves with an increase in the number of pilot symbols per frame shown in fig. 4. In this comparison; we keep the transmitted SNR per symbol to be the same in both cases.

(C) Orthogonal Space-Time Block Coding

Tarokh [2] develop the theory for orthogonal space-time block coding (OSTBC) for an arbitrary number of elements, allowing for extremely simple decoding with almost no growth in complexity. The development of OSTBC is based on the theory of orthogonal designs for real symbols unearthed by Tarokh and his co-authors in [2]. The authors extend these designs to include complex symbols. The discussion summarizes some of the key results that lead to OSTBC. The theory of OSTBC starts with real designs, assuming g real data. A real orthogonal design is a N × N matrix of "indeterminate" made of N variables xn or -xn. Given a block of N symbols, the indeterminate are replaced with the corresponding symbols. At time instant l, the lth row is transmitted over the N antenna elements. For example, a 2× 2 orthogonal design is:

$$\mathcal{O}_2 = \begin{pmatrix} x_1 & x_2 \\ -x_2 & x_1 \end{pmatrix}$$

Given a block of two symbols s1 and s2, at the first time instant, s1 is transmitted from the first element and s2 from the second element. In the second instant, −s2 is transmitted from the first element and s2 from the second element. Complex orthogonal design also shown in paper [2]. In the figure 5, we present some performance results for orthogonal space- time block coding using four transmit antennas (4x1 system) using a half-rate code, G4, as per [3]. We expect the system to offer a diversity order of 4 and will compare it with 1x4 and 2x2 systems, which have the same diversity order also. To allow for a fair comparison, we use quaternary PSK with the half-rate G4 code to achieve the same transmission rate of 1 bit/sec/Hz.

As expected, the similar slopes of the BER curves for the 4x1, 2x1 and 1x2 systems indicate an identical diversit y order for each system. Also observe the 3 dB penalty for the 4x1 system that can be attributed to the same total transmitted power assumption made for each of the three systems. If we calibrate the transmitted power such that the received power for each of these systems is the same, then the three systems would perform identically. Again, the theoretical performance matches the simulation performance of the 4x1 system as the total power is normalized across the diversity branches. In this performance analysis two transmitters are used maximum number of 3000 packets. Frame length of 100 is used. Numbers of pilot symbols per frame, used for channel estimation are 8 for fig.3 and 16 for fig.4. And bit energy to noise power spectral density varies to 12 dB. It is clear from the fig. 3 and 4 that there is radual improvement in the performance as we keep on increasing the value of pilot symbol per frame, which means as we append more pilot symbols in front of frame then the channel can be estimated more precisely and hence improvement in result. OSTBC is using four transmit antennas.

Maximum numbers of packets to be transmitting are 1500. Frame length of 150 is used. Bit energy to noise power spectral density varies to 20 dB. In this we compare the OSTBC result with Alamouti scheme, MRC technique and without diversity. MRC is showing best result in the figure. But if we increase the transmitted power or apply same power to each transmitter in OSTBC, the performance of OSTBC matches to the MRC technique.

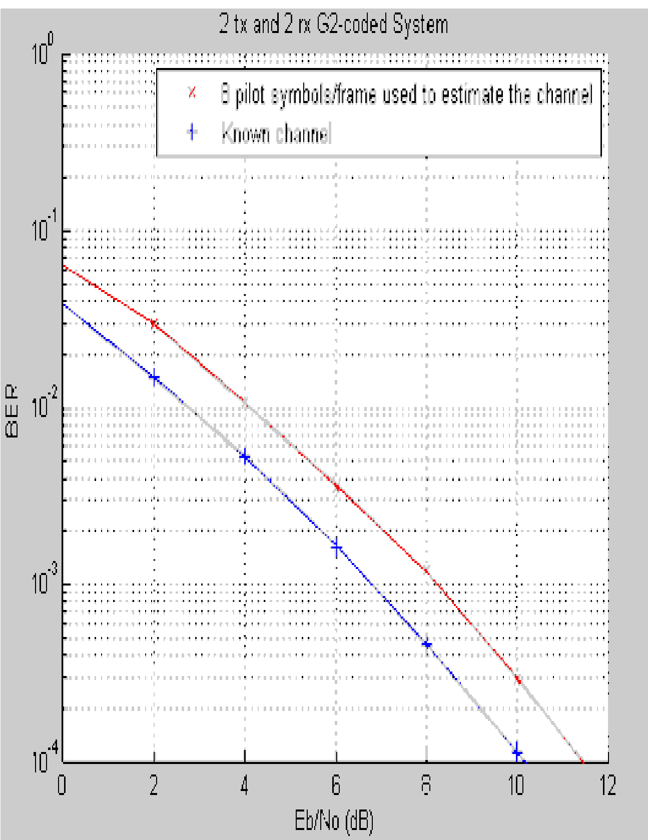


Fig 3: BER with and without channel estimation with 8 pilot

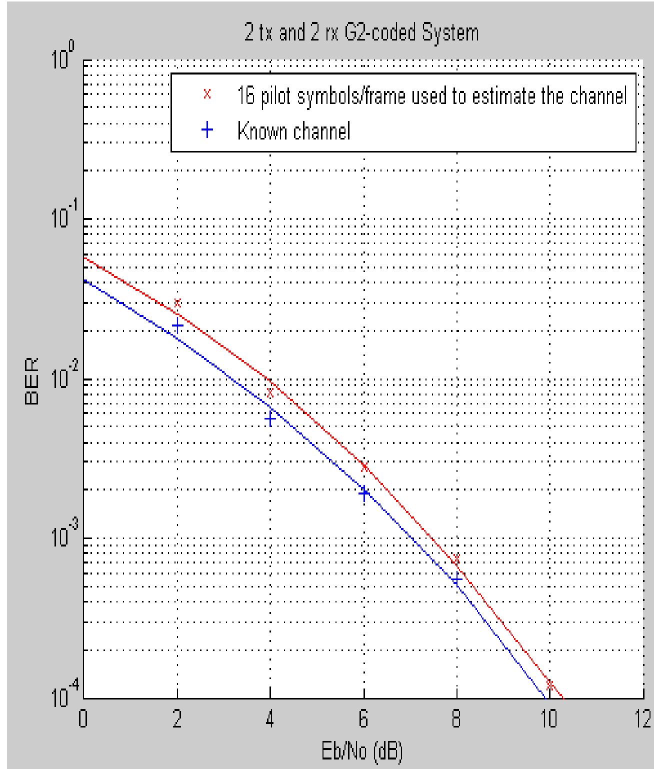


Fig 3: BER with and without channel estimation with 16 pilot

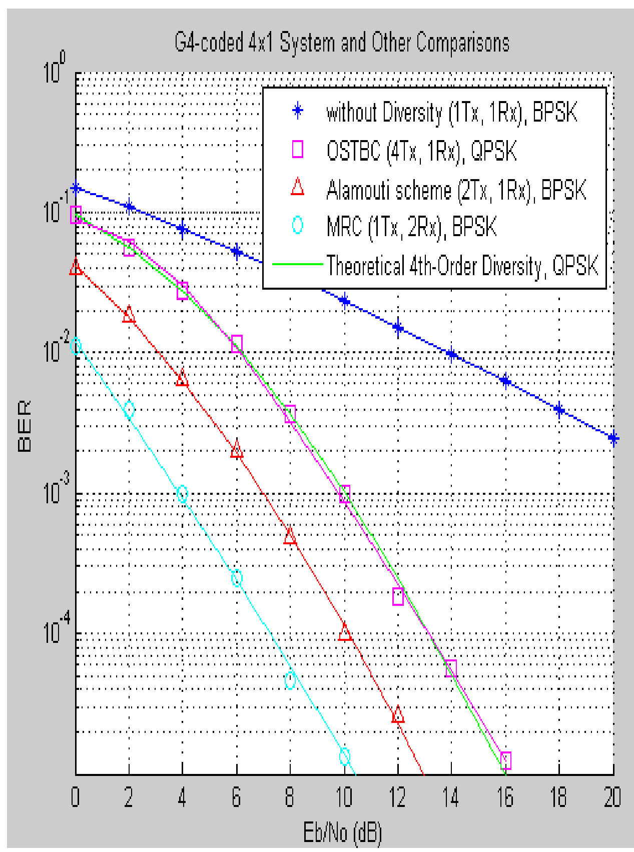


Fig 5: Othogonal STBC

IV. CONCLUSION

In this paper wehave analyze the Alamouti technique for MIMO wireless system. Alamouti STBC technique provides diversity improvement at all the remote units in a wireless system. Diversity is further improved by using orthogonal design. Also when the Pilot Symbol value increases, there is an improvement in BER as channel is precisely estimated. We also compare the OSTBC with various other techniques like MRC, Alamouti using STBC. By applying same power to each transmitter of OSTBC technique, as applied to single transmitter in MRC technique, the performance of OSTBC matches to MRC.

REFERENCES

[1] S. M. Alamouti, "A simple transmit diversity technique for wireless communications", IEEE(R) Journal on Selected Areas in Communications, Vol.16, No. 8, Oct. 1998, pp. 1451-1458.

[2] A.F. Naguib, V. Tarokh, N. Seshadri, and A.R Calderbank, "Space-time codes for high data rate wireless communication: Mismatch analysis", Proceedings of IEEE International Conf.. Communications, pp. 309-313, June 1997.

[3] V. Tarokh, H. Jafarkhami, and A.R.Calderbank, "Space-time block codes from orthogonal designs", IEEE Transaction on Information Theory, Vol. 45, No. 5, Jul. 1999, pp. 1456-1467.

[4] Fernando Gregorio, "Space Time code for MIMO Systems" Helsinki University of Technology.

[5] S. M. Alamouti, V. Tarokh, and P. Poon, "Trellis coded modulation and transmit diversity: Design criteria and performance evaluation," in Proc. IEEE ICUPC 98, Oct. 1998, pp. 703–707.

[6] "The application of orthogonal designs to wireless communication, "in Proc. IEEE Information Theory Workshop, Killarney, Ireland, June 1998, pp. 46–47.

[7] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communications: Performance criteria in the presence of channel estimation errors, mobility and multiple paths," IEEE Trans. Commun., vol. 47, pp. 199–207, Feb. 1999.

[8] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space–time codes for high data rate wireless communication: Performance analysis and code construction," IEEE Trans. Inform. Theory vol. 44, no. 2, pp. 744–765, Mar1998.

[9] G. Ganesan and P. Stoica, "Space-time block codes: a maximum SNR approach," IEEE Trans. Inform. Theory, vol. 47, pp. 1650-1656, May 2001.

[10]W. Su and X.-G. Xia, "On space-time block codes from complex orthogonal designs," Wireless Pers.Comm., vol. 25, pp. 1-26, Apr. 2003.

[11]W. Su, X.-G. Xia, and K. J. R. Liu, "Asystematic design of high-rate complex orthogonal space time block code,"IEEE commun." pp .380-382 june 2004.

Face Recognition using Adaptive Estimation of Principal components and Neural Network

Harpreet Kaur [1] and Sanjeev Kumar[2]
[1]Lecturer, Department of Electronics & Communication Engineering, MKCET, Amritsar, India
[2]Associate Professor, Department of Electronics & Communication Engineering, ACET, Amritsar, India
Email: er.harpreetbhullar@gmail.com, sanjeev_be@indiatimes.com

Abstract-- **Face recognition is one of biometric methods, to identify given face image using main features of face. Automatic face recognition systems try to find the identity of a given face image according to their memory. The memory of a face recognizer is simulated by a training set. Training set consists of the features extracted from known face images of different persons. Thus, the task of the face recognizer is to find the most similar feature vector among the training set to the feature vector of a given test image. In this paper, face recognition system is implemented using the Principal Component Analysis (PCA) algorithm and then , a neural based algorithm is presented, to detect frontal views of faces. The dimensionality of face image is reduced by the Principal component analysis (PCA) and the recognition is done by the Back propagation Neural Network (BPNN) and then the results are compared .The parameters Acceptance rate, false acceptance rate and false rejection ratio is compared.**

Keywords--PCA, BPN Network, AR, FAR and FRR.

I. INTRODUCTION

A face recognition system is a computer vision and it automatically identifies a human face from database images. The face recognition problem is challenging as it needs to account for all possible appearance variation caused by change in illumination, facial features, occlusions, etc. This paper gives a PCA based algorithm for face recognition. Holistic approach, feature-based approach and hybrid approach are some of the approaches for face recognition. Here, a holistic approach is used in which the whole face region is taken into account as input data. This is based on principal component-analysis (PCA) technique, which is used to simplify a dataset into lower dimension while retaining the characteristics of dataset.

Pre-processing and Principal component analyses are the major implementations of this paper. Pre-processing is done for two purposes

1) To reduce noise and possible convolute effects of interfering system.
2) To transform the image into a different space where classification may prove easier by exploitation of certain features.

PCA is a common statistical technique for finding the patterns in high dimensional data's. Feature extraction, also called Dimensionality Reduction, is done by PCA for a three main purposes like

1) To reduce dimension of the data to more tractable limits.
2) To capture salient class-specific features of the data.
3) To eliminate redundancy.

Here recognition is performed by PCA. The Algorithm for Face recognition using PCA is as follows:

1) Pre-processing stage –Images are made zero-mean and unit-variance.
2) Dimensionality Reduction stage: PCA - Input data is reduced to a lower dimension to facilitate classification.
3) Classification stage – Calculated the minimum Euclidean distance.

This paper describes about Principal component analysis, Back Propagation Neural Networks, demonstrates experimentation and results.

II. PCA ALGORITHM

Principal component analysis (PCA) involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. PCA is a popular technique, to derive a set of features for both face recognition.

Any particular face can be

1) Economically represented along the Eigen pictures coordinate space.
2) Approximately reconstructed using a small collection of Eigen pictures.

To do this, a face image is projected to several face templates called eigenfaces which can be considered as a set of features that characterize the variation between face images. Once a set of eigenfaces is computed, a face image can be approximately reconstructed using a weighted combination of the eigenfaces. The projection weights form a feature vector for face representation and recognition. When a new test image is given, the weights are computed by projecting the image onto the Eigen-face vectors. The classification is then carried out by comparing the distances between the weight vectors of the test image and the images from the database. Conversely, using all of the eigenfaces extracted from the original images, one can reconstruct the original image from the eigenfaces so that it matches the original image exactly.

III. STEPS FOR PCA ALGORITHM

The algorithm used for principal component analysis is as follows.

1) Acquire an initial set of M face images (the training set) & Calculate the Eigen-faces from the training set, keeping only M' eigenfaces that correspond to the highest eigenvalue.

2) Calculate the corresponding distribution in M'-dimensional weight space for each known individual, and calculate a set of weights based on the input image.

3) Classify the weight pattern as either a known person or as unknown, according to its distance to the closest weight vector of a known person.

Let the training set of images be by

$\Gamma 1, \Gamma 2 \ldots \Gamma M$.

The average face of the set is defined

$$\Psi = \frac{1}{M}\sum_{n=1}^{M}\Gamma_n \quad \text{--------------------------------} \ (1)$$

Each face differs from the average by vector

$$\Phi_i = \Gamma_i - \Psi \quad \text{-------------------------------------} \ (2)$$

The co- variance matrix is formed by

$$C = \frac{1}{M}\sum_{n=1}^{M}\Phi_n . \Phi_n^T = A.A^T \quad \text{--------------------} \ (3)$$

Where the matrix A = [Φ1, Φ 2,....., Φ M].

This set of large vectors is then subject to principal component analysis, which seeks a set of M orthonormal vectors *u* 1...*u M*

To obtain a weight vector of contributions Ω of individual Eigen-faces to a facial image Γ, the face image is transformed into its Eigen-face components projected onto the face space by a simple operation

$$\omega_k = u_k^T (\Gamma - \Psi) \quad \text{---------------------} \ (4)$$

For k=1,.., M', where M' ≤ M is the number of eigen-faces used for the recognition. The weights form vector

Ω = [ω1 ,ω2 ,......,ω M ']

that describes the contribution of each Eigen-face in representing the face image Γ, treating the eigen-faces as a basis set for face images. The simplest method for determining which face provides the best description of an unknown input facial image is to find the image k that minimizes the Euclidean distance ε *k*.

$$\varepsilon_k = \| (\Omega - \Omega_k) \|^2 \quad \text{---------------------} \ (5)$$

where Ω k is a weight vector describing the k face from the training set. A face is classified as belonging to person k when the 'ε k' is below some chosen threshold Θε. otherwise; the face is classified as unknown.

The algorithm functions by projecting face images onto a feature space that spans the significant variations among known face images. The projection operation characterizes an individual face by a weighted sum of eigenfaces features, so to recognize a particular face, it is necessary only to compare these weights to those of known individuals. The input image is matched to the subject from the training set whose feature vector is the closest within acceptable thresholds.

Eigen faces have advantages over the other techniques available, such as speed and efficiency. For the system to work well in PCA, the faces must be seen from a frontal view under similar lighting.

IV. EXPERIMENTAL RESULTS

In this paper for experimentation, 60 images are taken for training. One of the images as shown in fig 1 is taken as the Input image. The mean image and reconstructed output image by PCA, is as shown in fig 2 and 3.

Fig 1 Input Image

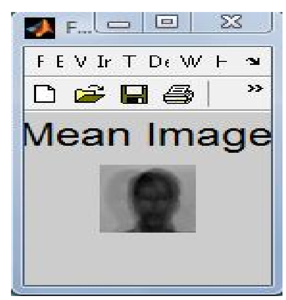


Fig 2 Mean Image

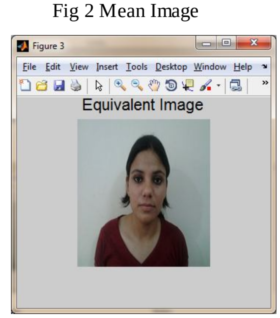


Fig 3 Recognized Image by PCA method

V. Back Propagation Neural Networks Algorithm

Neural networks are composed of simple elements operating in parallel. These elements are inspired by biological nervous systems. As in nature, the connections between elements largely determine the network function. You can train a neural network to perform a particular function by adjusting the values of the connections (weights) between elements. Typically, neural networks are adjusted, or trained, so that a particular input leads to a specific target output. The next figure illustrates such a situation. There, the network is adjusted, based on a comparison of the output and the target, until the network output matches the target. Typically, many such input/target pairs are needed to train a network. Neural networks have been trained to perform complex functions in various fields, including pattern recognition, identification, classification, speech, vision, and control systems. Neural networks can also be trained to solve problems that are difficult for conventional computers or human beings. The toolbox emphasizes the use of neural network paradigms that build up to--or are themselves used in-- engineering, financial, and other practical applications. Back propagation is the generalization of the Widrow-Hoff learning rule to multiple-layer networks and nonlinear differentiable transfer functions. Input vectors and the corresponding target vectors are used to train a network until it can approximate a function, associate input vectors with specific output vectors, or classify input vectors in an appropriate way as defined by you.

VI. Neural Networks and Back Propagation Algorithm

We can use back propagation neural network as follows: Choose what you would like to classify and the classes you want to sort them into. They must be in the form of separable units that can each be encoded. Prepare a training set. This consists of a list of inputs with correct outputs to train the network. Initialize the network. Choose the number of input nodes, output nodes, number of hidden layers and the stopping criterion. The number of input nodes is the number of elements in your input. Start the training phase. This will use the training set to reorganize the network until the stopping criterion is met. When this is met, the network will be saved and it will no longer be reorganized when an input is given. Test the network on an input not included in the training set. If the success rate is low, then try training a network with a different training set and stopping criterion.

VII. Supervised Learning - How it works

1) Send in an input x.
2) Run it through the network to generate an output y.
3) Tell the machine what the "right" answer for x actually is (we have a large number of these known Input-Output exemplars).
4) Comparing the right answer with y gives an error quantity.
5) Use this error to modify the weights and thresholds so that next time x is sent in it will produce an answer nearer to the correct one.
6) The trick is that at the same time as adjusting the network to make it give a better answer for x, we are adjusting the weights and thresholds to make it give better answers for other inputs. These adjustments may interfere with each other!

A typical back propagation network with Multi-layer, feed-forward supervised learning is as shown in the figure. 4. Here learning process in Back propagation requires pairs of input and target vectors. The output vector 'o' is compared with target vector 't'. In case of difference of 'o' and 't' vectors, the weights are adjusted to minimize the difference. Initially random weights and thresholds are assigned to the network. These weights are updated every iteration in order to minimize the mean square error between the output vector and the target vector.

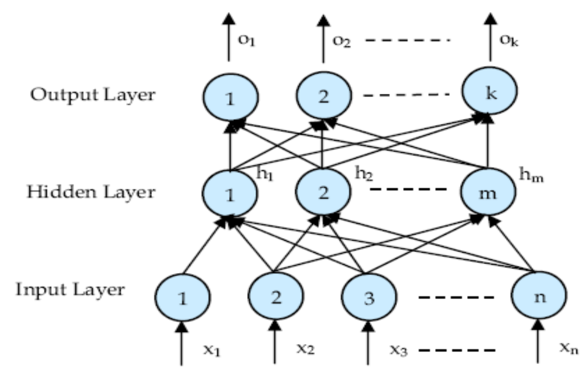


Fig 4 Basic Block of Back propagation neural network

Input for hidden layer is given by

$$net_m = \sum_{z=1}^{n} x_z w_{mz} \quad \text{---------------------------- (6)}$$

The units of output vector of hidden layer after passing through the activation function are given by

$$h_m = \frac{1}{1+\exp(-net_m)} \quad \text{-------------------- (7)}$$

In same manner, input for output layer is given by

$$net_k = \sum_{z=1}^{m} h_z w_{kz} \quad \text{---------------------------- (8)}$$

and the units of output vector of output layer are given by

$$o_k = \frac{1}{1+\exp(-net_k)} \quad \text{--------------------- (9)}$$

For updating the weights, we need to calculate the error. This can be done by

$$E = \frac{1}{2}\sum_{i=l}^{k}(o_i - t_i)^2 \quad \text{------------------- (10)}$$

o_i and t_i represents the real output and target output at neuron i in the output layer respectively. If the error is minimum than a predefined limit, training process will stop; otherwise weights need to be updated. For weights between hidden layer and output layer, the change in weights is given by

$$\Delta w_{ij} = \alpha \delta_i h_j \quad \text{--------------------------- (11)}$$

where α is a training rate coefficient that is restricted to the range [0.01,1.0], h_{aj} is the output of neuron j in the hidden layer, and δ_i can be obtained by

$$\delta_i = (t_i - o_i) o_i (l - o_i) \quad \text{--------------- (12)}$$

Similarly, the change of the weights between hidden layer and output layer, is given by

$$\Delta w_{ij} = \beta \delta_{Hi} x_j \quad \text{------------------------- (13)}$$

Where β is a training rate coefficient that is restricted to the range [0.01,1.0], x_j is the output of neuron j in the input layer, and δ_{Hi} can be obtained by

$$\delta_{Hi} = x_i (l - x_i) \sum_{j=1}^{k} \delta_j w_{ij} \quad \text{-------------- (14)}$$

x_i is the output at neuron i in the input layer, and summation term represents the weighted sum of all δ_j values corresponding to neurons in output layer that obtained in equation. After calculating the weight change in all layers, the weights can simply updated by

$$w_{ij}(new) = w_{ij}(old) + \Delta w_{ij} \quad \text{---- (15)}$$

This process is repeated, until the error reaches a minimum value.

VIII. Experimentation and Results

For the efficient operation of the back propagation network it is necessary for the appropriate selection of the parameters used for training.

Initial Weights:
This initial weight will influence whether the net reaches a global or local minima of the error and if so how rapidly it converges. To get the best result the initial weights are set to random numbers between -1 and 1.

Training a Net:
The motivation for applying back propagation net is to achieve a balance between memorization and generalization; it is not necessarily advantageous to continue training until the error reaches a minimum value. The weight adjustments are based on the training patterns. As long as error the for validation decreases training continues. Whenever the error begins to increase, the net is starting to memorize the training patterns. At this point training is terminated.

Number of Hidden Units:
If the activation function can vary with the function, then it can be seen that a n-input, m-output function requires at most 2n+1 hidden units. If more number of hidden layers are present, then the calculation for the δ's are repeated for each additional hidden layer present, summing all the δ's for units present in the previous layer that is fed into the current layer for which δ is being calculated.

Learning rate:
In BPN, the weight change is in a direction that is a combination of current gradient and the previous gradient. A small learning rate is used to avoid major disruption of the direction of learning when very unusual pair of training patterns is presented.

IX. CONCLUSION

Face recognition has received substantial attention from researches in biometrics, pattern recognition field and computer vision communities. In this paper, Face recognition using Eigen faces and Neural network has been done. It is concluded that for eigen face method the Acceptance ratio is about 90 % .In case of neural network using backpropagation the Acceptance ratio is about 93.33 %.so the result for neural network is better.

TABLE I
RESULTS OF PCA AND NEURAL

No. of Test Images = 30			
S No.	**Parameters**	**PCA (in % age)**	**NEURAL (in % age)**
1	Acceptance Rate	90	93.33
2	False Acceptance Rate	6.67	3.37
3	False Rejection Rate	3.33	3.33

REFERENCES

[1] B.K.Gunturk,A.U.Batur, and Y.Altunbasak,(2003) "Eigenface-domain super-resolution for face recognition," IEEE Transactions of . Image Processing. vol.12, no.5.pp. 597-606.

[2] M.A.Turk and A.P.Petland, (1991) "Eigenfaces for Recognition," Journal of Cognitive a. Neuroscience. vol. 3, pp.71-86.

[3] [T.Yahagi and H.Takano,(1994) "Face Recognition using neural networks with multiple combinations of categories," International Journal of Electronics Information and Communication Engineering., vol.J77-D-II, no.11, pp.2151-2159.

[4] S.Lawrence, C.L.Giles, A.C.Tsoi, and A.d.Back, (1993) "IEEE Transactions of Neural a. Networks. vol.8, no.1, pp.98-113.

[5] http://en.wikipedia.org/facerecognitionsystem Wikipedia, Face Recognition System.

[6] Kailash J. Karande Sanjay N. Talbar "Independent Component Analysis of Edge a. Information for Face Recognition" International Journal of Image Processing Volume (3) : Issue (3) pp: 120-131.

Apple Grading Method Based on Color using LABVIEW

Swati[1], Surinder Kumar[2]
[1]Malout Institute of Management and Information Technology /Electronics and Communication,Malout India.
[2]Malout Institute of Management and Information Technology /Electronics and Communication,Malout India.
Email: groversavi88@gmail.com, skmimit@gmail.com

***Abstract*— Quality of fruits and vegetables depends a lot on its surface color. Distribution of color on the surface indicates parameters like ripeness, defects, etc. Qualitative sorting is usually performed by trained inspectors which is rather expensive and is determined by operators' inconsistency and subjectivity. Machine vision technology offers objective solutions for all these problems and it is considered to be a promise for replacing the traditional human inspection methods. This paper introduces and discusses several techniques of analyzing the color and quality of fruit, based on image processing. Classification and grading has been done on apples and results indicate accuracy of the results up to 95%.**

Keywords- Colour Models, Fruit Grading, Hue Histogram, Graphical User Interface (GUI).

I. Introduction

Commercial value of fruits and vegetables depends on its surface color and its overall appearance. Since long, human inspectors have been deployed to do quality sorting of fruits and vegetables and which is rather expensive and found to be suffering from operator's inconsistency and subjectivity [1]. Only the color indicates parameters like defects and any brows or blemish in apples etc. The quality decisions vary among the graders due to lack of adaptation of human eye to small changes in color and the effect of the background on the perceived color and color intensity are the main sources of error. Hence, it is hard to provide precise guideline for visual inspection of fruits based on colour. Automated visual inspections techniques involving color based image processing have been reported since long to overcome human errors.

Mayer (1988) discussed advantages and disadvantages of employing human operators in the process of produce classification and grading [2]. In his paper, in 1994, Deck (1994) has presented pros and cons of semi-automatic sorting systems [3]. In 2000, P. Sudhakara Rao [4] discussed color analysis of fruits using machine Vision systems for automatic sorting and grading.

Machine vision techniques have been tried in this paper to grade apples based on their color and homogeneity of color on its surface.

II. Methodology

There are the mainly three color models used for image processing: RGB (Red, Green, Blue), CMY (Cyan, Magenta, Yellow), YIQ (Luminance, In-phase and Quadrature) and HSI (Hue, Saturation, Intensity). Red, Green and Blue are the primary colour components. They are additive, when adding colored lights and subtractive when adding paint pigments. Although the process followed by human brain in perceiving colour is a psychological phenomenon that is not yet fully understood, the physical nature of colour can be expressed on a formal basis supported by experimental and theoretical results. A particular color model is adopted depending on a specific application; for example, the RGB model is commonly used for color monitors, color video cameras etc., the CMY model is used for colour printers, YIQ [Luminance, In phase, Quadrature] model is used for TV broadcast. It has been experimentally found that HSI model is most suitable for finding out the ripeness of fruits, vegetables, colour matching samples, etc. Hue is a colour attribute that describes a pure colour where as Saturation gives a measure of the degree to which a pure colour is diluted by a white light. The Intensity is decoupled from the colour information of the image and Hue & saturation components are intimately related to the way in which human beings perceive color.

HSI model for grading of fruits by colour because of some advantages like:

- HSI model is an ideal tool for developing image processing algorithms based on colour descriptors that are natural and intitutive to humans.
- HSI is quite inexpensive in terms of time and memory space than RGB.
- Value of Hue is invariant of any intensity alterations.
- Classification algorithm is highly simplified.

.

The objective here is to grade the fruit based on surface color distribution estimated through color histograms. In RGB colour space, every individual colour component, namely Red, Green and Blue has its histogram.

Then, the percentage composition of every individual color component, which a fruit possesses, is estimated and this helps in classifying the apples. For a particular variety of Red Apples, a higher percentage composition of the red component was assigned a superior grade, the next lower composition the second grade & likewise the descending grades were assigned. Similarly for golden apples, higher percentage composition of yellow or golden colour was assigned a superior grade the next lower composition the second grade & likewise the descending grades.

Figure 1 illustrates a block diagram of the process followed in this automated visual inspection and grading of apples.

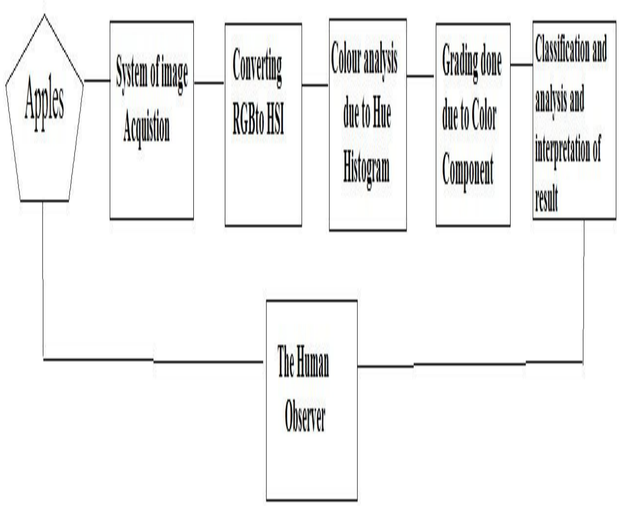


Figure1.Block diagram of the process

Images of sample apples were taken using a RGB camera and then read this image into LABVIEW software. RGB data is first converted into HSI data.

Various colour features can be calculated from the RGB components by using linear or non-linear transformation.Hue and Saturation are calculated from RGB values by

$$\mathrm{H} = \cos^{-1}\{\frac{\{0.5[(r-g)+(r-b)}{[(r-g)^2+(r-b)(g-b)]^1/2]}$$

H□ [0, π] for b ≤ g (1)

$$\mathrm{H} = 2\prod - \cos^{-1}\{\frac{\{0.5[(r-g)+(r-b)}{[(r-g)^2+(r-b)(g-b)]^1/2]} \ldots$$

H□ [0, π] for b > g (2)

S = 1-3.min(r, g, b) s □ [0, 1] (3)

I=(R+G+B) / 3.255 (4)

With image representation in the HSI domain, the colour analysis was based on primarily the Hue value. The three dimensional RGB space is reduced to a one-dimensional 'H' Space for colour analysis.

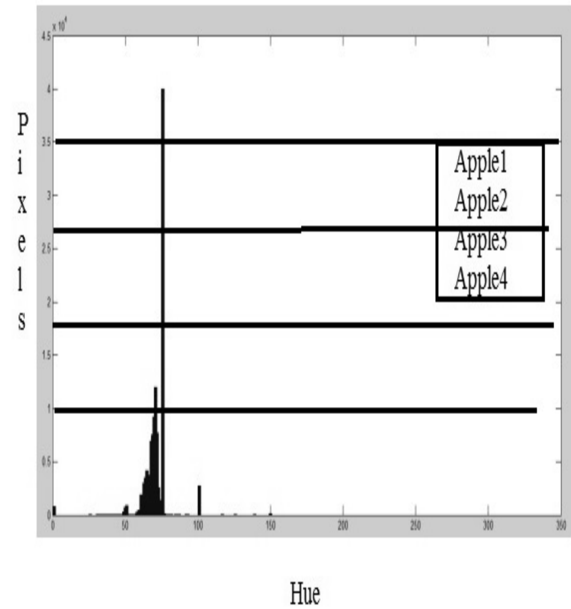


Figure 2. Typical Hue Histogram of apple

The Hue histogram is used to do the color analysis and grading the fruit. The probability density of the Hue pertains to surface colour is maximum at the median density pixel value. Hence the median density Hue value represents the overall distribution of surface colour of the apples is taken as the basis for segregating the fruits belonging to that grade. During the training process 20 Apple samples belonging to a known GRADE were taken and median density of the combined histogram is calculated, thus obtaining a distinct median density value for each of the GRADE to build the reference table

III. EXPERIMENTAL SET UP

Following steps were done to acquire the images and analysis further;

- Image acquisition
- Select ROI (Region of Interest)
- RGB to HSI
- Take the HUE component for extracting the colour present.
- Auto analysis of the colour based on Hue Histogram which gives numerical value for grading.
- Auto grade based on the analysis.
- Displays the result.

The images taken and processed ones are listed below in various figures;

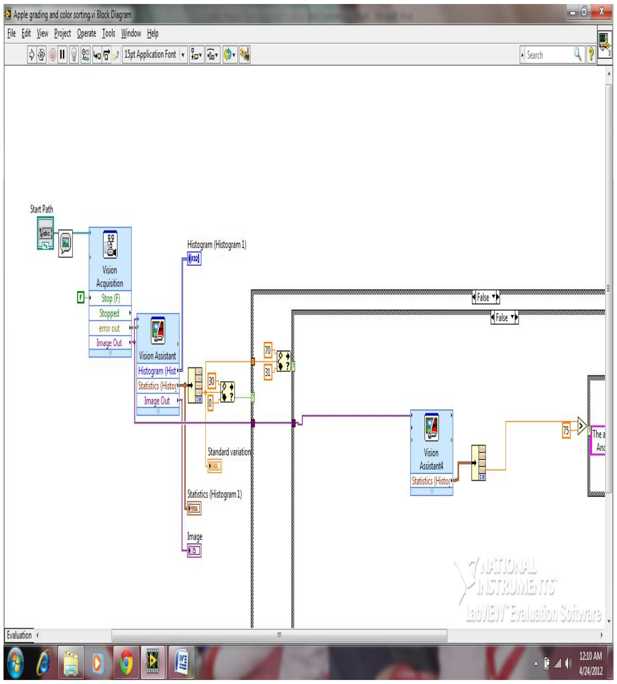


Figure3. Block Diagram of Color Sorting and Grading Through Histogram Statistics

Figure 3 shows block diagram of color sorting and grading and figure 4 is vision assistant in which image can be loaded into labview.

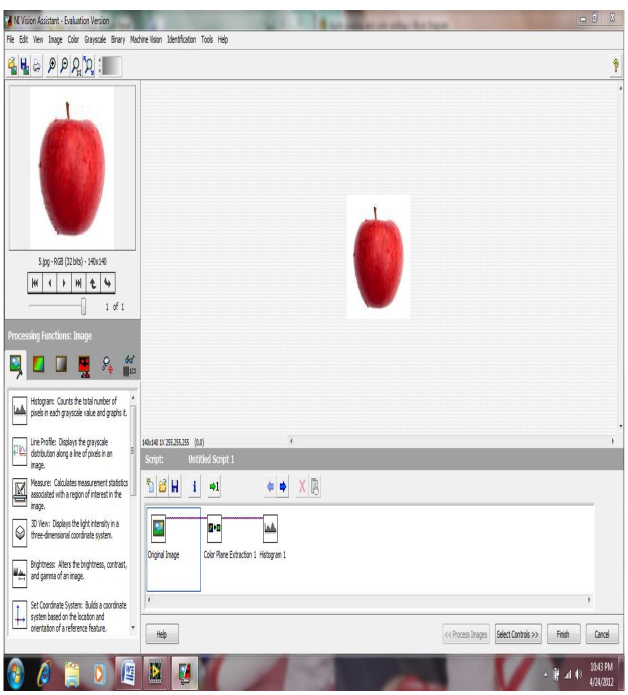


Figure4. Vision Assistant

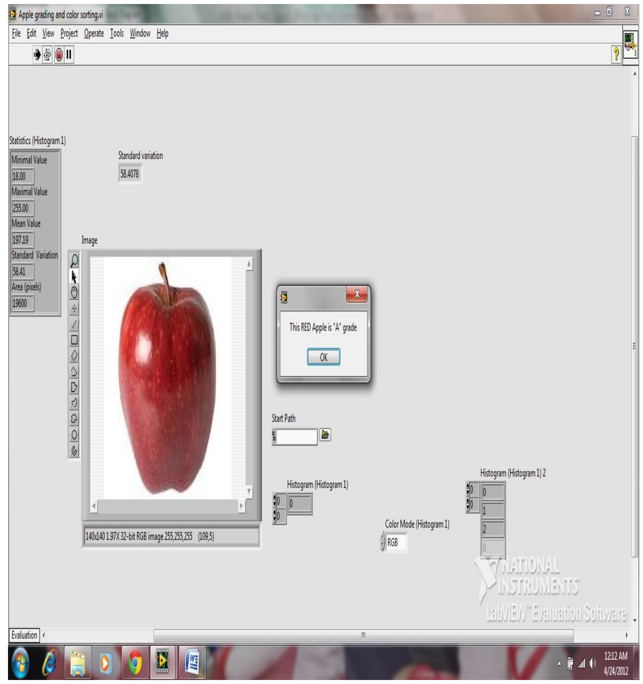


Figure 5. Front Panel of Color Sorting (RED APPLE)

Extract Hue compont of the image as depicted in figure 5 and show it as a red apple, green apple, yellow apple.

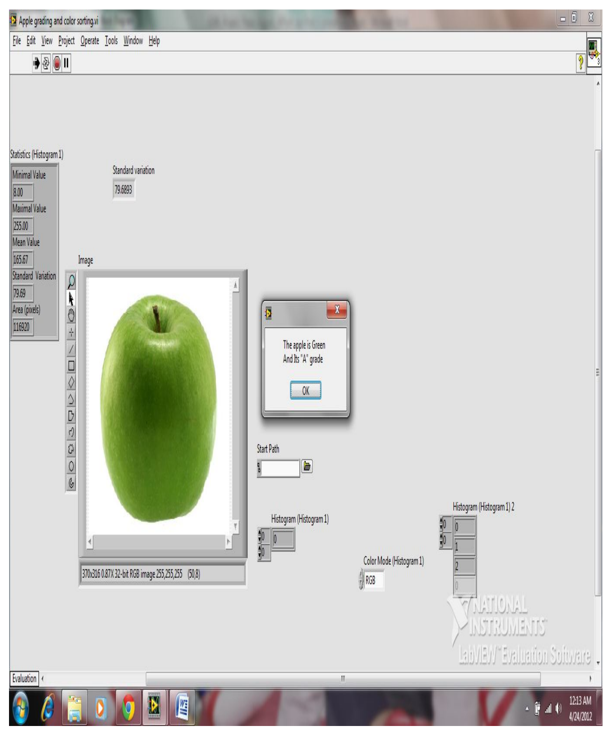


Figure 6. Front Panel of Color Sorting (GREEN APPLE)

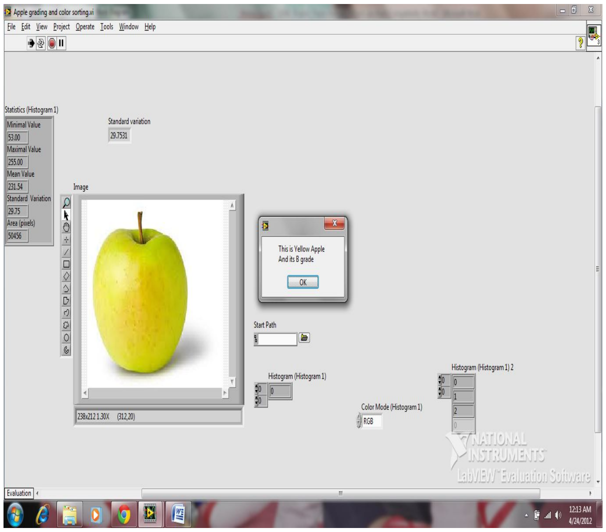


Figure 7. Front Panel of Color Sorting (YELLOW APPLE)

IV: RESULTS AND DISCUSSION

Training of the system on known grade of fruit was done and decision grading was tabulated as below Table.1, 2 and 3. The gradation values have been experimentally determined to have a span depending on Hue values. Hue value 0 represents 100% Red and as we go away from 0 the percentage of red colour occurrence is decreasing. If there is any defect such as any blemish in the red apple its hue component increases beyond 20% and the apple is graded as poor or reject grade. Similarly grading table has been generated for golden apples and green apples in which golden or yellow apple colour vary from 35 to 60 in which 35 represents the 100% golden and it goes away from 35 it golden colour occurrence is decreasing and also for green colour apples vary from 70 to 120. Following reference Table 1, 2, and 3 are used for the Apple classification.

Table1. Reference Table for color classification of red apples

GRADE	Classification based on their Hue Value
A	0-1
B	2-5
C	5-10
D	10-15
E	15-20

Table2. Reference Table for color classification of golden (yellow) Apples

GRADE	Classification based on their Hue Value
A	35-40
B	40-45
C	45-50
D	50-55
E	55-60

Table3. Reference Table for color classification of green Apples

GRADE	Classification based on their Hue Value
A	70-80
B	80-90
C	90-100
D	100-110
E	110-120

The system has been trained on 100 Apples, 5 sets of 20 Apples of each variety. It has been observed that changes in light intensity level during inspection affected the results of classification, although conceptually Hue should be independent of intensity of the lighting level. Experiments revealed that the Hue histogram of red apple shifted towards the green direction if the intensity is increased and vice-versa. The Hue histograms grading results achieved were about 95 % accurate as compared to manual grading system

CONCLUSION

In order to overcome the classification errors and inconsistencies involved in fruit grading done through human inspectors, an automated computerized inspection systems has been discussed in this paper. Use of color analysis using Hue histograms has been illustrated to demonstrate the grading of apples. Accuracy to the level of 95% has been achieved in experimental results. 5% Error present may be due to system calibrations vis-à-vis human grading and/or variation of Hue parameter with respect to light variations which should not be there ideally. This inaccuracy leaves a scope for further investigation.

REFERENCES

[1] Newmann T.S, Jain A.K, “A survey of automated visual inspection”, Computer Vision and Image Understanding Vol: 61, No. 2, pp.231-262, 1995.

[2] Meyers, J, B. Jr., “Improving dynamic visual inspection performance,” M.Sc. thesis, Univ. of Georgia, Athens, 1988

[3] Deck S.H., “Evaluations of semi-automated vegetable sorting concept,”Transactions of the ASAE, 1994.

[4] P. Sudhakara Rao, A. Gopal, R. Revathy and K. Meenakshi CEERI Chennai” Colour analyze fruits using machineVision system for automatic sorting And grading “
CSIR Madras Complex, Taramani, Chennai - 600 011

[5] Marius Buzera, voicu groza, gabriela prostean and octavian prostean “Techniques of analyzing the colour of produces for automatic classification” Blvd.V.Parvan, no.2, 300223, Timisoara, Romania (978-1-4244-2083-4/08/$25.00 ©2008 IEEE), Feburary, 2008.

[6] Liyanage C De Silva, Anton Pereira, and Amal Punchihewa” Food Classification using Color Imaging” Institute of Information Sciences and Technology, Massey University, New Zealand,2004

Review Study of Short Range Wireless Protocols: Bluetooth, ZigBee, UWB, and Wi-Fi

Tarun Singh[1], Gurbhej Singh[2] and Sumit Sharma[3]
[1, 2 ,3]Lecturers, Department of Electronics & Communication Engineering, BKSCET, Amritsar, India
Email: tarun23.singh@yahoo.co.in, gurbhej86@gmail.com, ersumitsharma@gmail.com

***Abstract*—In today's world there are number of short range wireless range wireless communication protocols i.e. Bluetooth, ZigBee, UWB,Wi-Fi which consume low power . Basically ZigBee is used for reliable wireless control and monitoring system, Bluetooth is used for cordless mouse , handfree headset and keyboard etc. , Wi-Fi is designed as an extension or substitution of cable network for computer to computer communication where as UWB oriented to high bandwidth multimedia links. In this paper, we provide a review of these popular short range wireless communication protocol.**

Index Terms- **Wireless protocols, Bluetooth, Ultra-wideband (UWB), ZigBee, Wi-Fi**.

I. INTRODUCTION

From last few years, automation has been developed worldwide into very attractive research areas. It associates with different technologies including communication, information, control, and sensor. Industrial communication is very important in factory automation [1]. Sensors, controllers and heterogeneous machines are combined for interconnection purpose with factory automation. Many different network types have been promoted for use on a shop floor, including control area network (CAN), Process field bus (Profibus), Modbus, and so on. However, how to select a suitable network standard for a particular application is a critical issue to the industrial engineers. Lain et al. [2] evaluated the Ethernet (carrier sense multiple access with collision detection, CSMA/CD bus), ControlNet (token-passing bus), and DeviceNet (CSMA with arbitration on message priority, CSMA/AMP bus) for networked control applications. After a detailed discussion of the medium access control (MAC) sublayer protocol for each network, they studied the key parameters of the corresponding network when used in a control situation, including network utilization and time delays.

Wireless communication is fast growing technology for accessing network and services which provide flexibility and mobility [3]. Eliminating cables is one of the advantages of wireless network with respect to cable devices. It also provides dynamic network formation, easy deployment and having low cost. Basically short range wireless system is currently based on four protocols: the Bluetooth (IEEE802.15.1), and UWB (IEEE 802.15.3), ZigBee (IEEE 802.15.4), and Wi-Fi (802.11a/b/g). The action range of these wireless technologies around to 10-100 meters. . For Bluetooth and Wi-Fi, Ferro and Potorti [4] compared their main features and behaviors in terms of various metrics, including capacity, network topology, security, quality of service support, and power consumption. In [5], Wang et al. compared the MAC of IEEE 802.11e and IEEE 802.15.3. Their results showed that the throughput difference between them is quite small. In addition, the power management of 802.15.3 is easier than that of 802.11e. ZigBee have wider variety of real industrial needs over Bluetooth due to its long term battery operation, greater useful range and flexibility.

This paper provides an overview on short range wireless protocols, comparison of these protocols and study there characteristics like transmission time, data coding efficiency and protocol complexity.

II. SHORT RANGE WIRELESS PROTOCOLS

This section introduces the Bluetooth, UWB, ZigBee, and Wi-Fi protocols, which corresponds to the IEEE 802.15.1, 802.15.3, 802.15.4, and 802.11a/b/g standards, respectively. The IEEE defines only the PHY and MAC layers in its standards. For each protocol, separate alliances of companies worked to develop specifications covering the network, security and application profile layers so that the commercial potential of the standards could be realized. The major goal of this paper is not to contribute to research in the area of wireless standards, but to present a comparison of the four main short-range wireless networks.

A. Bluetooth over IEEE 802.15.1

Bluetooth, also known as the IEEE 802.15.1 standard is based on a wireless radio system designed for short-range and cheap devices to replace cables for computer peripherals, such as mice, keyboards, joysticks, and printers. This range of applications is known as wireless personal area network (WPAN). Two connectivity topologies are defined in Bluetooth: the piconet and scatternet. A piconet is a WPAN formed by a Bluetooth device serving as a master in the piconet and one or more Bluetooth devices serving as slaves. A frequency-hopping channel based on the address of the master defines each piconet. All devices participating in communications in a given piconet are synchronized using the clock of the master. Slaves communicate only with their master in a point-to-point fashion under the control of the master. The master's transmissions may be either point-to-point or point-to-multipoint. Also, besides in an active mode, a slave device can be in the parked or standby modes so as to reduce power consumptions.

A scatternet is a collection of operational Bluetooth piconets overlapping in time and space. Two piconets can be connected to form a scatternet. A Bluetooth device may participate in several piconets at the same time, thus allowing for the possibility that information could flow beyond the coverage area of the single piconet. A device in a scatternet could be a slave in several piconets, but master in only one of them.

B. UWB over IEEE 802.15.3

UWB has recently attracted much attention as an indoor short-range high-speed wireless communication. [7]. One of the most exciting characteristics of UWB is that its bandwidth is over 110 Mbps (up to 480 Mbps) which can satisfy most of the multimedia applications such as audio and video delivery in home networking and it can also act as a wireless cable replacement of high speed serial bus such as USB 2.0 and IEEE 1394. Following the United States and the Federal Communications Commission (FCC) frequency allocation for UWB in February 2002, the Electronic Communications Committee (ECC TG3) is progressing in the elaboration of a regulation for the UWB technology in Europe. From an implementation point of view, several solutions have been developed in order to use the UWB technology in compliance with the FCC's regulatory requirements. Among the existing PHY solutions, in IEEE 802.15 Task Group 3a (TG3a), multi band orthogonal frequency-division multiplexing (MB-OFDM), a carrier-based system dividing UWB bandwidth to sub-bands, and direct-sequence UWB (DS-UWB), an impulse-based system that multiplies an input bit with the spreading code and transmits the data by modulating the element of the symbol with a short pulse have been proposed by the WiMedia Alliance and the UWB Forum, respectively. The TG3a was established in January 2003 to define an alternative PHY layer of 802.15.3. However, after three years of a jammed process in IEEE 802.15.3a, supporters of both proposals, MB-OFDM and DS-UWB, supported the shut down of the IEEE 802.15.3a task group without conclusion in January 2006. On the other hand, IEEE 802.15.3b, the amendment to the 802.15.3 MAC sublayer has been approved and released in March 2006.

C. ZigBee over IEEE 802.15.4

ZigBee over IEEE 802.15.4, defines specifications for low-rate WPAN (LR-WPAN) for supporting simple devices that consume minimal power and typically operate in the personal operating space (POS) of 10m. ZigBee provides self-organized, multi-hop, and reliable mesh networking with long battery lifetime [8-9]. Two different device types can participate in an LR-WPAN network: a full-function device (FFD) and a reduced-function device (RFD). The FFD can operate in three modes serving as a PAN coordinator, a coordinator, or a device. An FFD can talk to RFDs or other FFDs, while an RFD can talk only to an FFD. An RFD is intended for applications that are extremely simple, such as a light switch or a passive infrared sensor. They do not have the need to send large amounts of data and may only associate with a single FFD at a time.

Consequently, the RFD can be implemented using minimal resources and memory capacity. After an FFD is activated for the first time, it may establish its own network and become the PAN coordinator. All star networks operate independently from all other star networks currently in operation. This is achieved by choosing a PAN identifier, which is not currently used by any other network within the radio sphere of influence.

Once the PAN identifier is chosen, the PAN coordinator can allow other devices to join its network. An RFD may connect to a cluster tree network as a leave node at the end of a branch, because it may only associate with one FFD at a time. Any of the FFDs may act as a coordinator and provide synchronization services to other devices or other coordinators. Only one of these coordinators can be the overall PAN coordinator, which may have greater computational resources than any other device in the PAN.

D. Wi-Fi over IEEE 802.11a/b/g

Wireless fidelity (Wi-Fi) includes IEEE 802.11a/b/g standards for wireless local area networks (WLAN). It allows users to surf the Internet at broadband speeds when connected to an access point (AP) or in ad hoc mode. The IEEE 802.11 architecture consists of several components that interact to provide a wireless LAN that supports station mobility transparently to upper layers. The basic cell of an IEEE 802.11 LAN is called a basic service set (BSS), which is a set of mobile or fixed stations. If a station moves out of its BSS, it can no longer directly communicate with other members of the BSS. Based on the BSS, IEEE 802.11 employs the independent basic service set (IBSS) and extended service set (ESS) 47network configurations. As shown in Fig. 1, the IBSS operation is possible when IEEE 802.11 stations are able to communicate directly without any AP. Because this type of IEEE 802.11 LAN is often formed without pre-planning, for only as long as the LAN is needed, this type of operation is often referred to as an ad hoc network. Instead of existing independently, a BSS may also form a component of an extended form of network that is built with multiple BSSs. The architectural component used to interconnect BSSs is the distribution system (DS). The DS with APs allow IEEE 802.11 to create an ESS network of arbitrary size and complexity. This type of operation is often referred to as an infrastructure network.

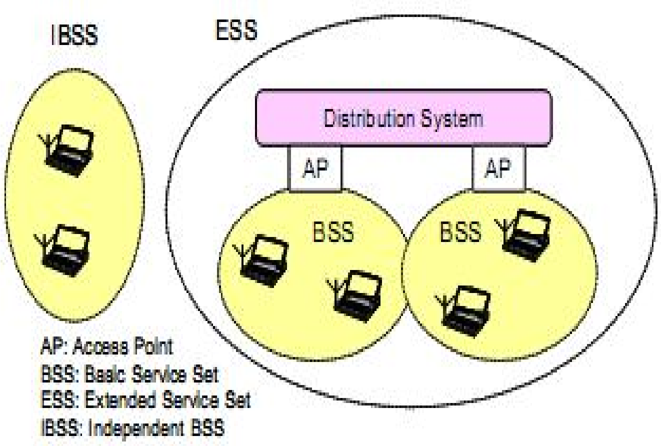


Fig. 1. IBSS and ESS configurations of Wi-Fi networks.

III. COMPARATIVE STUDY

Table 1 summarizes the main differences among the four protocols. Each protocol is based on an IEEE standard. UWB and Wi-Fi provide a higher data rates, while Bluetooth and ZigBee give a low data rates. UWB, Bluetooth and ZigBee are designed for WPAN communication (about 10 m), while Wi-Fi is designed for WLAN (about 100m).Zigbee can also reach 100 m in some applications.

TABLE I
COMPARISON OF THE BLUETOOTH, UWB, ZIGBEE, AND WI-FI PROTOCOLS

Standard	Bluetooth	UWB	ZigBee	Wi-Fi
IEEE spec.	802.15.1	802.15.3a *	802.15.4	802.11a/b/g
Frequency band	2.4 GHz	3.1-10.6 GHz	868/915 MHz; 2.4 GHz	2.4 GHz; 5 GHz
Max signal rate	1 Mb/s	110 Mb/s	250 Kb/s	54 Mb/s
Nominal range	10 m	10 m	10 - 100 m	100 m
Nominal TX power	0 - 10 dBm	-41.3 dBm/MHz	(-25) - 0 dBm	15 - 20 dBm
Number of RF channels	79	(1-15)	1/10; 16	14 (2.4 GHz)
Channel bandwidth	1 MHz	500 MHz - 7.5 GHz	0.3/0.6 MHz; 2 MHz	22 MHz
Modulation type	GFSK	BPSK, QPSK	BPSK (+ ASK), O-QPSK	BPSK, QPSK COFDM, CCK, M-QAM
Spreading	FHSS	DS-UWB, MB-OFDM	DSSS	DSSS, CCK, OFDM
Coexistence mechanism	Adaptive freq. hopping	Adaptive freq. hopping	Dynamic freq. selection	Dynamic freq. selection, transmit power control (802.11h)
Basic cell	Piconet	Piconet	Star	BSS
Extension of the basic cell	Scatternet	Peer-to-peer	Cluster tree, Mesh	ESS
Max number of cell nodes	8	8	> 65000	2007
Encryption	E0 stream cipher	AES block cipher (CTR, counter mode)	AES block cipher (CTR, counter mode)	RC4 stream cipher (WEP), AES block cipher
Authentication	Shared secret	CBC-MAC (CCM)	CBC-MAC (ext. of CCM)	WPA2 (802.11i)
Data protection	16-bit CRC	32-bit CRC	16-bit CRC	32-bit CRC

* Unapproved draft

The typical parameters of the four wireless protocols used for transmission time evaluation are listed in Table II

TABLE II
TYPICAL SYSTEM PARAMETERS OF THE WIRELESS PROTOCOLS

Standard	Bluetooth	UWB	ZigBee	Wi-Fi
IEEE Spec.	802.15.1	802.15.3	802.15.4	802.11a/b/g
Max data rate (Mbit/s)	0.72	110*	0.25	54
Bit time (μ s)	1.39	0.009	4	0.0185
Max data payload (bytes)	339 (DH5)	2044	102	2312
Max overhead (bytes)	158/8	42	31	58
Coding efficiency† (%)	94.41	97.94	76.52	97.18
* Unapproved 802.15.3a.	† Where the data is 10K bytes.			

IV. CONCLUSION

This paper has presented a broad overview of the four most popular wireless standards, Bluetooth, UWB, ZigBee and Wi-Fi with quantitative evolution in terms of frequency band, channel Bandwidth, data rate and power consumption and data protection. This paper is not to draw any conclusion regarding which one is superior since the suitability of network protocols is greatly influenced by practical applications, of which many other factors such as the network reliability, roaming capability, recovery mechanism, chipset price, and installation cost need to be considered in the future.

REFERENCES

[1] R. Zurawski, "Guest editorial of special section on factory communication systems," IEEE Trans. Ind. Electron., vol. 49, no. 6, pp. 1186-1188, Dec. 2002.

[2] F. L. Lian, J .R. Moyne, and D. M. Tilbury, "Performance evaluation of control networks: Ethernet, ControlNet, and DeviceNet," IEEE Contr. Syst. Mag. vol. 22, no. 1, pp. 66-83, Feb. 2001.

[3] A. Willig, "An architecture for wireless extension of Profibus," in Proc. IEEE Int. Conf. Ind. Electron. (IECON'03), Roanoke, VA, Nov. 2003, pp. 2369-2375.

[4] E. Ferro and F. Potorti, "Bluetooth and Wi-Fi wireless protocols: A survey and a comparison," IEEE Wireless Commun., vol. 12, no. 1, pp. 12-16, Feb. 2005.

[5] X. Wang, Y. Ren, J. Zhao, Z. Guo, and R. Yao, "Comparison of IEEE 802.11e and IEEE 802.15.3 MAC," in Proc. IEEE CAS Symp. Emerging Technologies: Mobile & Wireless Commun, Shanghai, China, May, 2004, pp. 675-680.

[6] Baker, N. "ZigBee and Bluetooth: Strengths and weaknesses for industrial applications," IEE Computing & Control Engineering, vol. 16, no. 2, pp 20-25, April/May 2005.

[7] D. Porcino and W. Hirt, "Ultra-wideband radio technology: Potential and challenges ahead," IEEE Commun. Mag., vol. 41, no. 7, pp. 66-74, July 2003.

[8] J. S. Lee, "Performance evaluation of IEEE 802.15.4 for low-rate wireless personal area networks," IEEE Trans. Consumer Electron., vol. 52, no. 3, pp. 742-749, Aug. 2006.

[9] J. S. Lee and Y. C. Huang, "ITRI ZBnode: A ZigBee/IEEE 802.15.4 platform for wireless sensor networks," in Proc. IEEE Int. Conf. Systems, Man & Cybernetics, Taipei, Taiwan, Oct. 2006, pp. 1462-1467.

DWT BasedText Information Extraction From Color Images

Neha Gupta[1] and Dr. V.K Banga[2]
[1]Department of Electronics & Communication Engineering, ACET, Amritsar, India
[2]Professor, Department of Electronics & Communication Engineering, ACET, Amritsar, India
Email: nehajandialaug05@gmail.com, vijaykumar.banga@gmail.com

Abstract— **This paper presents a method for extracting text from different kinds of images such as scene text images, document images, caption images with an unified framework. However, variation of text due to differences in size, style, orientation and alignment, as well as low image contrast and complex background make the problem of automatic text extraction extremely challenging. The text extraction algorithm system is designed to extract text from all the three kinds of images and eliminates the need to devise separate method for various kinds of images. Firstly, the color image is converted into grayscale image. After that, Haar Discrete Wavelet Transform (DWT) is employed. Haar DWT decompose image into four sub image coefficients, one is average and other three are detail. Now, Sobel edge detector is applied on three detail components, the resultant edges so obtained to form edge map. Finally, morphological operations are performed on processed edge map and further thresholding is applied to improve the performance.**

Keywords- Image segmentation, Discrete Wavelet Transform, Haar wavelet, Sobel edge detecto, Text information extraction.

I. INTRODUCTION

An image contains variety of data from low to high level information. Among them, text is the most informative types. Text detection has been used in a number of applications such as [1]: document analysis, content based image retrieval, video classification, sign recognition, camera based text detection, license plate extraction etc. Image content can be divided into two main categories: perceptual content and semantic content [2]. Perceptual content includes attributes such as color, intensity, shape, texture and their temporal changes, where as semantic content means objects, events and their relations. Text extraction can be done from three kinds of images namely:

a) Document image
b) Scene text image
c) Caption text image

The diagram showing three types of images:

a. Document image

b. Scene text image

c. Caption text image

Figure1: Document, Scene and Caption Image

Document images may be in the form of scanned book covers, CD covers or video images. Text in images or videos is classified as scene text and caption text. Scene text is also called as graphics text. Natural images that contain text are called scene text. The name of caption text is artificial text and it is one in which text is inserted or superimposed in the image [7].

II. WHAT IS TEXT INFORMATION EXTRACTION?

The problem of text information extraction needs to be defined more precisely before proceeding further. The architecture of Text information extraction is shown in figure 2. Text information extraction system receives an input in the form of a still image or a sequence of images.

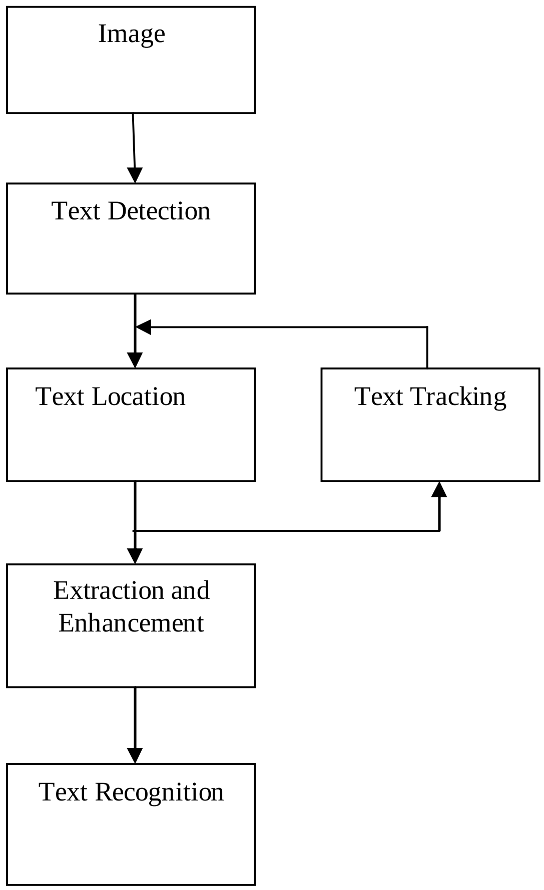


Figure 2: Architecture of Text Information Extraction

The images can be in gray scale or color, compressed or un-compressed and the text in the image may or may not move. The text information extraction problem can be divided into the following sub problems:

A. Text Detection

There is no prior information on whether or not the input image contains any text. Various approaches assume that the certain type of video frame or image contain text. However, in case of video, the number of frames containing text is much smaller than the number of frames without text. As text usually has a higher intensity than the background. They counted the number of pixels that are lighter than predefined threshold value and exhibited a significant color difference relative to their neighborhood, and regarded a frame with a large number of such pixels as a text frame. This method is extremely simple and fast. But, problems are also associated with this approach.

B. Text localization

The process of localization involves further enhancing the text regions by eliminating non-text regions. One of the properties of text is that usually all characters appear close to each other in the image and thus forming a cluster.

Text localization method can be categorized into two types:

1) Region based methods : This approach uses the properties of the color or gray scale in a text region or their differences regarding the background. This method is basically divided in two sub categories: edge based and connected component (CC) based methods. The edge based method is mainly focus on the high contrast between text and background. In this method, firstly text edges are identified in an image and are merged. Finally, some heuristic rules are applied to discard non-text regions. Connected component based method considers text as a set of separate connected components, each having distinct intensity and colour distributions. The edge based methods are robust to low contrast and different text size where as CC based methods are somewhat simpler to implement, but they fail to localize text in images with complex background [1]. Region based approach include region growing and split and merge algorithm, exploit spatial information to group character pixels efficiently.

2) Texture based methods: As we know that, text in images has distinct textural properties which can be used to differentiate them from the background or other non text regions [3]. This method is based on the concept of textural properties. In this method, Fourier transforms. Discrete cosine transform and wavelet decomposition are generally used. The main drawback of this method is that it is highly complex in nature but, in other hand, it is more robust than the CC based methods in dealing with complex background.

C. Tracking

Text tracking could be performed in a shorter time than text detection and localization; this would speed up the overall system. Every block is checked as to whether its fill factor is above a given threshold value. For each block that meets the required fill factor, a block matching algorithm is performed but when a block has an equivalent in a subsequent frame and the gray scale difference between the blocks is less than a threshold value, the block is considered as a text component.

D. Extraction and Enhancement

Although most text with simple background and high contrast can be correctly localized and extracted, poor quality text can be difficult to extract. Text enhancement techniques can be divided into two categories: single frame or multi frame base. Many threshold techniques have been developed for still images. However, these methods do not work well for video sequences.

E. OCR (Recognition)

The common OCR systems available require the input image to be such that the character can be easily parsed and recognized. The text and background should be monochrome and background to text contrast should be high.

III. TEXT EXTRACTION ALGORITHM

The block diagram of the text extraction algorithm is shown in figure3. The input image may be a color or gray scale image. If the image is color image then, preprocessing operation is applied on the image as shown in the flowchart. In our algorithm, input data is a color image which is entered to the system and the segmented text on a clear background is the output [9].

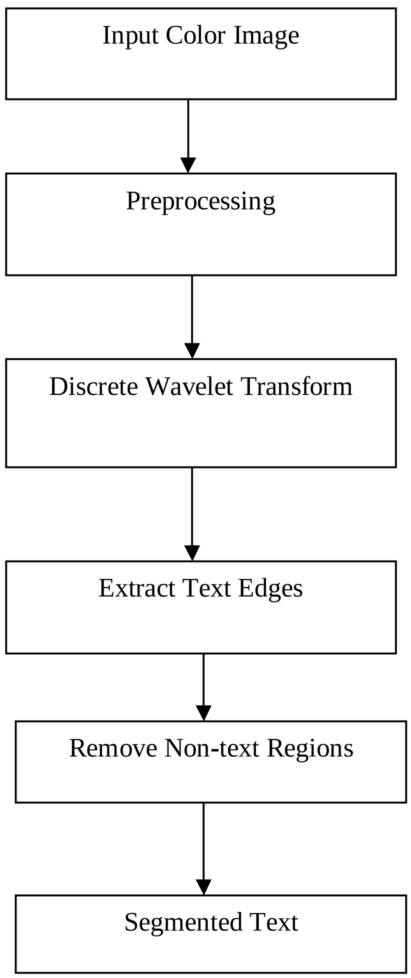


Figure 3 Block Diagram of Text Extraction Algorithm

A. Preprocessing

In this step, if the input image is a gray-scale image, then image is processed directly starting at discrete wavelet transform. If the input image is colored , then its RGB components are combined to give an intensity image. Usually, color images are normally captured by the digital cameras. The pictures are often in the Red-Green-Blue color space. Intensity image Y is given by:

$$Y = 0.299R + 0.587G + 0.114B$$

Image Y is then processed with 2-d discrete wavelet transform.

B. Discrete wavelet transform

In this paper, we are using Haar discrete wavelet transform which provides a powerful tool for modeling the characteristics of textured images. Most textured images are well characterized by their contained edges. It can decompose signal into different components in the frequency domain [12]. We are using 2-d DWT in which it decomposes input image into four components or sub-bands, one average component(LL) and three detail components(LL, HL, HH) as shown in figure 4 [9] .The three detail component sub-bands are used to detect candidate text edges in the original image. Using Haar wavelet, the illumination components are transformed to the wavelet domain. This stage results in the four LL, HL, LH and HH sub-image coefficients. The traditional edge detection filters can provide the similar result as well but it cannot detect three kinds of edges at a time. Therefore, processing time of the traditional edge detection filters is slower than 2-d DWT. The reason we choose Haar DWT because it is simpler than that of any other wavelets.

LL	HL
LH	HH

Figure 4 Result of 2-D DWT Decomposition

Some of the following advantages are as follows:

a. Haar wavelets are real, orthogonal and symmetric.

b. Its coefficients are either 1 or -1.

c. Haar wavelets are real, orthogonal and symmetric.

d. Its coefficients are either 1 or -1.

e. It is the only wavelet that allows perfect localization in the transform domain

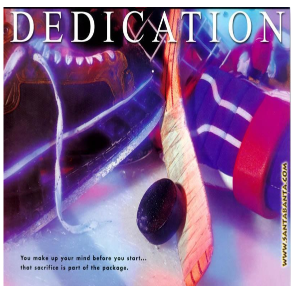
C. Extraction of text edges

In this extraction process, three detail sub-components are used to detect edges of the text blocks. By finding the edges in the three sub images namely horizontal sub image, vertical sub image and diagonal sub image, fusing the edges contained in each sub image. In this way candidate text regions can be found. In this algorithm, we use Sobel edge detector because it is efficient to extract strong edges that are needed in this application. The next step is to form an edge map using weighted 'OR' operator. To get binary edge map, thresholding is to be applied. After that we perform morphological dilation operation on binary edge map. Basically the function of dilation is to fill the gaps inside the obtained text regions.

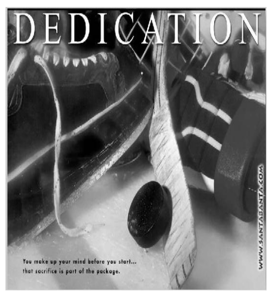
D. Removing non-text regions

To improve the efficiency of the system, non text regions are removed using projection profile method. The projection profile is used to separate text blocks into single line text. There are two types of projection profile. One is horizontal profile and another one is vertical profile. A horizontal profile is defined as vector of the sum of the pixel intensities over each column and a vertical profile is defined as vector of the sum of the pixel intensities over each row. The horizontal and vertical projections of the binary edge map are found. The average value of maximum and minimum of the vertical projection is taken as threshold and in the same way, maximum and the minimum value of horizontal projection is taken as threshold.

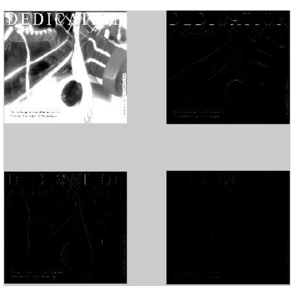
In case of vertical projection, the rows whose sum of pixel intensities above the threshold are taken. Similarly, in case of horizontal projection only the columns whose sum of pixel intensities above the threshold is taken [8]. In this way there is proper localization of text regions in the image is found. Finally, a threshold is applied which result in the segmented text in a black background as shown in figure 5.



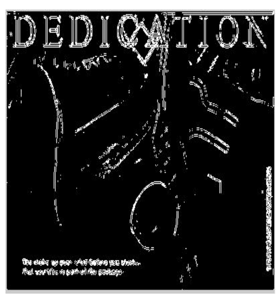
a) Original Image



b) Gray Scale Image



c) DWT Image



d) Binary Edge Map

e)Text Localization

Figure 5. Result from Text Extraction Process

IV. CONCLUSION

In this paper, we present a relatively simple and effective algorithm for text detection and extraction by applying DWT to the images [13]. As it requires less processing time which is essential for real time applications [11]. Mostly all the previous methods fail when the characters are not aligned well or when the characters are too small. They also result in some missing characters when the characters have very poor contrast with respect to the background [12]. But this algorithm is not sensitive to image color or intensity, uneven illumination and reflection effects. This algorithm can be used in large variety of application fields such as vehicle license plate detection to detect number plate of vehicle, mobile robot navigation to detect text based land marks, object identification, identification of various parts in industrial automation, analysis of technical papers with the help of charts, maps, and electric circuits etc.. This algorithm can handle both scene text images and printed documents. The future work involves extracting text from different types of complex and fast moving text in videos.

REFERENCES

[1] Keechul Jung, Kwang In Kim and Anil K. Jain, "*Text information extraction in images and video: A Survey*", Elsevier, Pattern Recognition, vol.37 (5), pp 977–997, 2004.

[2] H.K.kim, "*Efficient Automatic Text Location Method and Content Based Indexing and Structuring Of Video Database*", Journal of Visual Communication Image Representation,vol.7(4),pp336-344,1996.

[3] Mohieddin Moradi, Saeed Mozaffari, and Ali Asghar Orouji, "*Farsi/Arabic Text Extraction from Video Images by Corner Detection*", 6th, IEEE, Iranian conference on Machine Vision and image processing, Isfahan, Iran, 2010.

[4] Chung-Wei Liang and Po-Yueh Chen, "*DWT Based Text Localization*", International Journal of Applied Science and Engineering, pp.105-116, 2004.

[5] Nikolaos G. Bourbakis, "*A methodology for document processing: separating text from images*", Engineering Applications of Artifcial Intelligence 14, pp. 35-41, 2001.

[6] C. Strouthopoulos, N. Papamarkos, A. E. Atsalakis, "*Text extraction in complex color documents*", The Journal of Pattern Recognition 35, pp.1743–1758, 2002.

[7] Chitrakala Gopalan and D. Manjula, "*Contourlet Based Approach for Text Identification and Extraction from Heterogeneous Textual Images*", International Journal of Electrical and Electronics Engineering 2(8), pp. 491-500, 2008.

[8] Dr.N.Krishnan, C. Nelson Kennedy Babu 2, S.Ravi 3 and Josphine Thavamani, "*Segmentation Of Text From Compound Images", International Conference on Computational Intelligence and Multimedia Applications*", Tamil nadu , India, vol. 3, pp.526-528, 2007.

[9] S.Audithan, RM. Chandrasekaran, "*Document Text Extraction from Document Images Using Haar Discrete Wavelet Transform*", European Journal of Scientific Research ISSN 1450-216X, Vol.36, 2009.

[10] Xiaoqing Liu and Jagath Samarabandu, "*Multiscale edgebase Text extraction from Complex images*", IEEE Multimedia and Expo 2006, International Conference, Tronto, Canada, pp. 1721-1724, 2006.

[11] Roshanak Farhoodi and Shohreh kasaei, "*Text Segmentation From Images With Textured and colored Background*", 13th Iranian conference on Electrical Engineering, 2005 .

[12] Xiao-Wei Zhang, Xiong-Bo Zheng, Zhi-Juan Weng, "*Text Extraction Algorithm Under Background Image Using Wavelet Transforms*", Proceedings of the 2008 International Conference on Wavelet Analysis and Pattern Recognition, Hong Kong, pp.30-31, Aug. 2008.

[13] R. lienhart, F. Stuber, "*Automatic Text Segmentation and Text Recognition for Video Indexing*", The Journal Multimedia Systems , vol(8), pp 69-81,Jan 2000.

[14] M. Padmaja, J. Sushma, "*Text Detection in Color Images*", International Conference on Intelligent Agent & Multi-Agent Systems, Chennai, 22-24 July 2009, pp. 1-6, 2009.

A Survey on 4G Wireless Systems

Er. Harsimranjit Singh[1] and Er. Sukhjit Singh[2]
[1]Department of Electronics & Communication Engineering, ACET, Amritsar, India
Email: er_harsimran@yahoo.co.in, er.sukhjit85@gmail.com

***Abstract* —4G mobile technology is in a determining and standardization stage. 4G wireless technology offers higher data rates and the ability to roam across multiple heterogeneous wireless networks. Since 4G is still in the cloud of the sensible standards creation, but still more work is not done in this field. The ITU and IEEE form several methods for the completion of 4G mobile standards. In this paper we will concentrate on current trends and its underlying technologies to implement the 4G mobile technology. This paper also shows some of the possible scenarios that will benefit the 4th generation technology.**

I. INTRODUCTION

In this fast changing technology, there is a rising requirement for people to communicate and get connected with each other. In continuous development of mobile technology, the major service providers in the wireless market kept on monitoring the growths of 4th generation (4G) mobile technology. 2G and 3G are well-established mobile technology around the world. 3G is stumbling to obtain market share for a different reasons and 4G is achieving some confidence. In 2013, the total mobile users is in North America, Europe and Asia Pacific, is expected to grow up to 2500 millions. 4G mobile technology as an example, will give people a more convenience and ease in lifestyle. With the “anytime, anywhere, anything,” capability, 4G wireless technology will benefit every individual regardless of time and place.

4G mobile technologies are perceived to provide fast and high data rate or bandwidth, and offer packetized data communications. Users’ experiences of latest booming Internet forces industry to investigate means to provide high data rate regardless of mobility. There still have large room for the purpose of service application vision: 3G is being delayed in its commercialization and about a decade of change is left for 4G. However, we believe this paper will promote discussion of 4G services by presenting our vision of 4G services.

In this paper, we also outline the current trend of next generation of wireless communications and investigate 4G candidate technologies. Based on this investigation, four scenarios will be discussed to predict and analyze 4G.

II. CHARACTERISTICS OF 4G SYSTEMS

The chief characteristics of 4G mobile systems would inculcate horizontal communication model that would combine different technologies such as wireless, cellular and wired systems on a common platform which would mean getting different service requirements in a best possible way.

A. CONVERGENCE SERVICES

Convergence in essence refers to the building up of a conducive atmosphere that leads to long haul communication while eliminating any possible source of data interruption which in turn lead to highly reliable and quality service. A great emphasis has been put up by various industries to provide seamless convergence services. This kind of development would inspire many a minds to shift gears and focus attention on fully mobile and ubiquitous convergence of media. The trends from the service perspective include integration of services and convergence of service delivery mechanisms. This platform would lead to more flexible architecture, widely recognized and incorporate new services will be easy to deploy.

B. BROADBAND SERVICES

Broadband is a basis for the purpose of enabling multimedia communications including video service, which requires transmission of a large amount of data. The increasing position of broadband services like Asymmetric Digital Subscriber Line (ADSL) and optical fiber access systems and office or home LANs is expected to lead to a demand for similar services in the mobile communication environment. 4G service application characteristics will give broadband service its advantages;

1) Low cost:
To make broadband services available to the user to exchange various kinds of information, it is necessary to lower charges considerably in order to keep the cost at or below the cost of existing service.

2) Coverage of Wide Area:

One chief requirement of mobile communication is its global availability. Taking this factor into consideration it is important that the upcoming technology clubs with the existing one maintaining high service coverage area.

3) Wide Variety of Services Capability:
Mobile communication is for various types of users. In the future, we expect to make the advanced system performance and functionality to introduce a variety of services not only the ordinary telephone service. Those services must be made easier for anyone to use.

C. FLEXIBILITY AND PERSONALIZED

Flexibility of a 4G system is of immense importance. 4G systems will support comprehensive and personalized services, providing stable system performance and quality of service. To support multimedia services, high-data rate services with good system reliability will be provided. In order to meet the demands of these diverse users, service providers should design personal and customized services for them. Personal mobility is a concern in mobility management. Personal mobility concentrates on the movement of users instead of users' terminals, and involves the provision of personal communications and personalized operating environments. Implementing SDR to 4G offers the following advantage benefits.
1) Increases the effectiveness of the infrastructure resources.
2) Better space efficiency
3) Lessens operational expenditure suitable to reduced need for hardware site upgrades.
4) Lowers capital expenditure because of rise in usage of accessible network elements.

III. SCENARIOS 4G SYSTEMS

Based on these visions and characteristics of the 4th generation (4G) for future wireless telecommunication, new spectrum allocation issue, and technology feasibility, the advent of 4G service will bring a number of changes of competition environment, regulation and policy as well as service change into future wireless communication. Accordingly, it is very important we expect what kinds of possibility we have for the 4G service to prepare well. A key feature of 4G is likely to be the availability of significantly higher data rates than for third-generation (3G) systems. It has been suggested that data rates up to 100 Mbps for high mobility and 1 Gbps for low mobility should be the target value. These data rates suggest higher spectral efficiencies and lower cost per bit will be key requirements for such future systems. Additional important and expected features are likely to be increased flexibility of mobile terminals and networks, multimedia services, and high-speed data connections. Future convergence systems will clearly be another feature. In order to forecast the form of realization of 4G systems, we construct four scenarios in the paper.

A. SCENARIO 1

The first scenario is that current 3G service is evolved into 3GPP LTE (long term evolution) which is one of the candidate technologies for 4G. Under the 3GPP LTE, 4G technology will be used basically in 3G spectrum and platform that means existing carriers maintain present customer base and services are integrated 4G. To support broadband service there will be an additional spectrum band with current 3G spectrum band. 3GPP LTE plan to support All-IP based backbone network to connect with other heterogeneous networks seamlessly. In this scenario, 3G incumbent service providers will maintain current subscribers and 3G services will be integrated to 4G.

B. SCENARIO 2

The second scenario is that fixed wireless, led by IEEE, enhances techniques to support mobility and fulfill 4G characteristics. Especially, mobile WiMAX (IEEE 802.16e) technology, WiBro in Korea, is very close to 4G technology which includes OFDM and MIMO. Moreover, there will be a possibility that IEEE 802.20 technology support high mobility to compensate low mobility of WiMAX. In this scenario incumbent service providers based on 3G get smaller market power and there will be a chance for new service providers which control current subscriber connections. In this scenario, a competition structure in the market will be shifted and changed.

C. SCENARIO 3

The third scenario is that both 3G LTE and WiMAX exist. They are in a complementary relationship with each other. Subscribers would possess both 3GPP LTE and WiMAX terminals, and they will adapt to use either of them according to the usage scene and needs. By the SDR (software defined radio) technology, subscribers will get flexibility between different services. Additionally, there could be s new mobile access service as well in a new spectrum band. In this scenario, a number of different service providers will compete for the leadership break out upon introduction of 4G.

D. SCENARIO 4

The last scenario is that subscribers switch over to cost-free transmission services, due to the fact that 3G LTE and WiMAX do not successfully deployed with high cost and dissatisfied service. Therefore, in this scenario, service providers would get limited market such as high quality sensitive business users. However, handheld manufacturers and contents providers regard the free transmission market that secures subscribers as important, and develop unique products and services that would appeal to the users. If this scenario is realized, there will be a dramatic change in the value chain of wireless communication industry.

IV. SCENARIO ANALYSIS

For the purpose of understanding of each scenario's impact on the wireless communication industry, we analyze each scenario based on feasibilities about technology and expected time plan. In addition that we examine the question as to whether each scenario fulfills 4G characteristics which are expected previous research and market situation. Furthermore, we investigate each scenario and find its own pros and cons.

Scenario1: Existing service providers maintain current Subscribers and services are integrated 4G	Scenario2: New service providers control subscriber connections towards 4G
Scenario3: Co-existence and mutual prosperity of existing service providers	Scenario4: Absence of service providers capable of service integration.

V. TECHNICAL FEASIBILITY

Based on the 4G characteristics as we mentioned earlier in this paper, 4G should fulfill convergence, broadband, flexibility, personalized services and All-IP network, Basically standards of 4G candidate services plan to meet these characteristics. In the first scenario, 3GPP LTE include reduced latency, higher user data rates up to 100 Mbps with high mobility, improved system capacity and coverage, and reduced overall cost for the operator. In the second scenario, mobile WiMAX and IEEE 802.20 also meet all the requirements for mobile Internet access. It supports multiple handoff mechanisms, power-saving mechanisms for mobile devices, advanced QoS and low latency for improved support of real-time applications [9], Furthermore, mobile WiMAX was designed from the ground up to be an All-IP technology.

VI. 4G ANALYSES

Considering 4G characteristics, expected scenarios and market trends, we can find out strengths, weaknesses, opportunities and threats of 4G with better understandings. The lists and findings follow.

A. Strengths in 4G:

- 4G visions take into account installed base and past investments
- Strong position of telecommunications vendors expected in the marketplace.
- Faster data transmission and higher bit rate and bandwidth, allow more business applications and commercialization
- Has advantage for personalized multimedia communication tools

B. Opportunities in 4G:

- Evolutionary approach may yield opportunities for the 4G
- Emphasis on heterogeneous networks capitalizes on past investments
- Strategic alliance and coalition opportunities with traditional nontelecommunication industries
- Sophisticated and mature commercialization of 4G technology would encourage more applications of e-commerce and m-commerce
- Worldwide economy recover stimulates consumption and consumer confidence, therefore bring in opportunities for telecommunication sections
- It is expected and predicted that consumers will continue to replace handsets with newer technology at a fast rate.
- Desirable higher data capacity rates, the growth opportunity for 4G is very bright and hopeful.

C. Threats in 4G:

- Faster rate of growth and developments in other region
- Since 3G mobile is still in the market, it squeezes the market competition in the mobile industry.

VII. CONCLUSION

As we come up with the 4G analysis out of this 4G technology, it is inevitable that 4G would completely replace 3G in a long run. Nevertheless, 4G and 3G tend to keep a co-competitive relationship in a short run. In order for 4G to grow in the future market, it is unavoidable to compete with 3G and acquire 3G's customers. As it was also analyzed and investigated through the scenarios, the comparison was made here that among three candidates for the 4G presented. Every service providers and manufacturers strategize towards high mobility and high data rates whether it is 3GPP, WiMAX or even WiBro oriented. However, the mainstream of service providers concern about regulation, uncertainty of market, and economic burden. There is also new spectrum allocation issue which should be resolved and determined, much as the technology feasibility. Any how, there are still plenty of opportunities for 4G. Under these circumstances, to be flourished in the future telecommunication market, each technology should be finalized its standards soon and developed systems to meet needs of consumer demands in a right time. Furthermore, the technical development, change and innovation should be reflected in a future regulation policy.

REFERENCES

[1] Pioneer Consulting Report, "The WiMAX Report Emergence of Fixed & Mobile Solutions", March, 2006.

[2] Keiji Tachikawa, "A Perspective on the evolution of mobile communications", IEEE Communications Magazine, October 2003.

[3] Lu, Willie W, Open Wireless Architecture and Enhanced Performance, Guest Editorial, IEEE Communication Magazine, June 2003

[4] ETRI Paper, "Broadband Mobile Communications towards a Converged World", ITU/MIC Workshop on Shaping the Future Mobile Information Society, March 4-5, 2004

[5] Andrew Dursch, David C. Yen, Shi-Ming Huang, "Fourth generation wireless communications: an analysis of future potential and implementation", Computer Standards & Interfaces 28, 2005, 13-25.

[6] ITU Report, "Broadband Mobile Communications Towards a Converged World.

Simulative Analysis of FBG-Based Spectral Amplitude Coding Optical CDMA

Manjit Singh[1], Himali Sarangal[2], Butta Singh[3]
[1, 2, 3] Assistant Professor, Department of Electronics & Communication Engineering, GNDU, RC, Jalandhar
Email: manu_kml@yahoo.com, Himali.sarangal@gmail.com, bsl.khanna@gmail.com

***ABSTRACT*- In recent years, spectral amplitude-coding (SAC) OCDMA system attracts more attention because the multi-user interference (MUI) is completely eliminated by spectral coding. In this paper we have design a Spectral Amplitude Coding OCDMA (SAC-OCDMA) system having bit rate 200 MB/s for four users. The system is further compare for NRZ and RZ format. The result shows that the maximum distance travelled when we use NRZ format is 93Km and Q-factor is 5.63. But in case of RZ format the distance covered is 92Km and Q-factor is 5.62 which is approximately as same as NRZ format.**
Keywords: SAC-OCDMA, MUI, NRZ, RZ

I. INTRODUCTION

Due to increasing demand for higher information rate the optical CDMA system plays an important role in optical network. Because it provide the advantage to access same bandwidth simultaneously by multiple users without employing high-speed electronic data processing circuits. It also provides high level security during transmission [1]. OCDMA is a digital technique where, instead of each channel occupying a given wavelength, frequency or time slots, information is transmitted using a coded sequence of pulses. It utilizes the basic principle of spread spectrum transmission where all users share the fiber channel bandwidth simultaneously. OCDMA system can be classified into two systems i.e. synchronous system and asynchronous system. The synchronous system uses spreading codes are of modified prime codes and asynchronous system uses optical orthogonal codes (OOCs) as spreading code. OCDMA offers many advantages, mainly at the networking level. It simplifies network control and management. All the users can operate asynchronously in the uplink. Each channel employs a specific code to transmit and recover the original signal. If correct code arrives, an auto correlation function is high and cross correlation function is zero for incorrect code sequence cross correlation function is high. For incorrect code sequence, cross correlation functions and crosstalk are generated and they create multiple-user interference (MUI). Spectral amplitude coding optical code division multiple access (SAC-OCDMA) systems attract more attentions since the MUI can be completely eliminated by spectral coding existing in conventional systems [2].

In this paper we have design a Spectral Amplitude coding OCDMA (SAC-OCDMA) system having bit rate 200 MB/s for four users. The system is further compare for NRZ and RZ format. The BER can be estimated from equation (1), and requires Q >6 for the BER 10^{-9}. This BER gives the upper limit for the signal because some degradation occurs at the receiver end [3].

$$BER= \tfrac{1}{2}\,\mathrm{erfc}\,(Q/\sqrt{2}) \quad \ldots\ldots\ldots\ldots (1)$$

II. SYSTEM DESCRIPTION

Typical optical spectral amplitude-coding (SAC) OCDMA system transmission system consists of a transmitter (Tx), power combiner, optical fiber as a channel, power splitter and receiver (Rx) is shown in figure1. This model is a subsystem of transmitter telescope, optical fiber and receiver telescope. The design of transmitter and receiver is shown in figure2 and 3. The first stage in transmitter section consists of PRBS (pseudo-random bit sequence) generator. The second stage is coding/ modulation. NRZ coding is generated by the engines of NRZ pulse generator. In this set up we are taking first NRZ format and then RZ format in order to compare the result. The transmitter section consists of data source having a bit rate of 200 Mb/s, white light source the frequency of source is 1550.5nm and power is 10dbm further transmission section consist of MZ modulator. The Mach-Zehnder modulator is an intensity modulator based on an interferometric principle. Transmitter also consists of two uniform fiber bragg grating having frequency 1550.1nm and 1550.9 and bandwidth 0.3nm. Transmitter section is shown in figure 2. The receiver section is consists of power splitter(1×2), four uniform fiber bragg grating having frequency 1550.1nm, 1550.9nm,1548.5nm and 1552.5nm. All UFB's having 0.3 nm bandwidth.

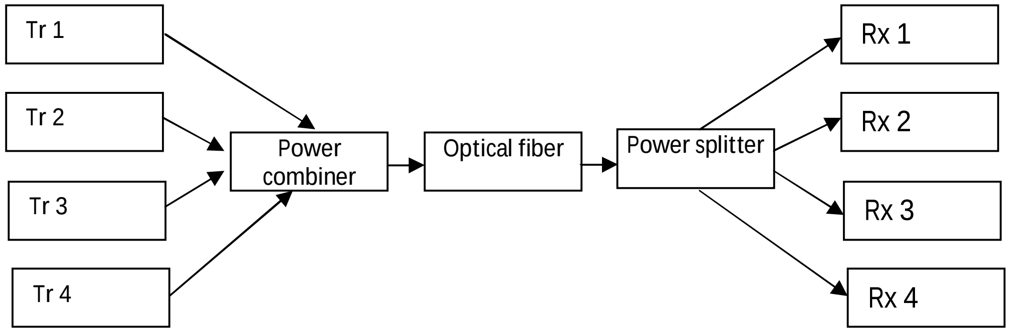


Figure.1. Design layout of FBG based SAC-OCDMA system with 4 users.

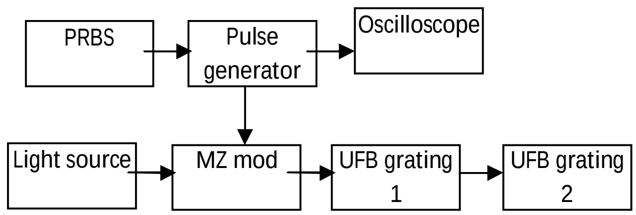


Figure 2. Design of Transmitter

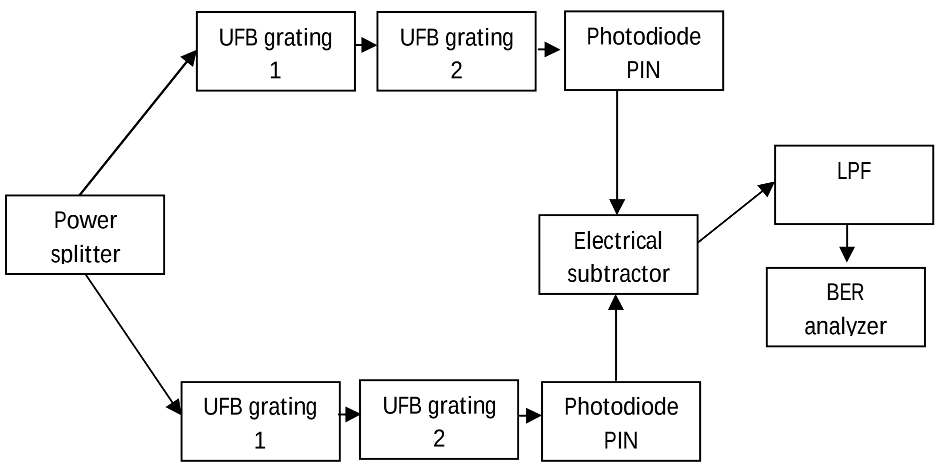


Figure 3. Design of Reciever

Then it consists of photodiode having responsivity and dark current 1A/W, 10nA respectively. The receiver is a PIN type and is followed by an electrical subtractor, low pass filter and BER analyzer. SAC-OCDMA system has been numerically simulated for four users by using optisystem software. The simulation is carried out for NRZ and RZ formats. Parameters used are shown in table 1 as per optisystem software [4].

Table 1. List of the parameters considered in the set up

Sr.no.	Parameter	Value taken
1	Frequency	1550.5 nm
2	Signal bit rate	200Mb/s
3	Power	10dbm
4	Responsivity	1 A/W
5	Dark current	10 nA

III. RESULTS AND DISCUSSION

The right choice of the performance evaluation criteria for the characterization of optical transmission links represent one of the key issuses for an effective design of future long-haul optical systems. The evaluation criteria should provide a precise determination and separation of dominant system limitations, making them crucial for the suppression of propagation disturbances and a performance improvement. The most widely used performance evaluation are the Q-factor and BER [3]. SAC-OCDMA system has been numerically simulated for NRZ and RZ formats for examine the comparison of two in terms of BER and Q-factor is shown in Fig.1. For simulation following parameters are used: frequency of the transmitter is 1550.5nm, transmitted power is 10dbm and signal bit rate is 200Mb/s. The eye diagrams showing the BER and Q-factor for NRZ and RZ are shown in Fig.4-5. For every simulation, the simulation parameters are same.

In case I i.e. when NRZ format is taken in that case the maximum length of the optical fiber should be 93Km in order to maintain the BER. As shown in figure the value of Q-factor at 93Km is 5.63, BER is 7.92e-009, eye height is 119773and threshold value is 197645. So it is clear that if we are using SAC-OCDMA 4 users system having signal bit rate of 200Mb/s the maximum optical fiber length is 93Km for NRZ pulse generator.

In case II i.e. when RZ format is taken in that case the maximum length of the optical fiber should be 92Km in order to maintain the BER. As shown in figure the value of Q-factor at 92Km is 5.62, BER is 9.08e-009, eye height is 51717.8and threshold value is 69936.6. So it is clear that if we are using SAC-OCDMA 4 users system having signal bit rate of 200Mb/s the maximum optical fiber length is 93Km for RZ pulse generator.

The results show that values are approximately same for return to zero and non return to zero formats. The value of Q-factor in case of NRZ is equal to the value of Q-factor in RZ format. There is minor difference in the value of BER in both the cases.

IV. CONCLUSION

It is concluded from the all above observation that the maximum distance travelled when we use NRZ format is 93Km and Q-factor is 5.63. But in case of RZ format the distance covered is 92Km and Q-factor is 5.62 which is approximately as same as NRZ format. 6.

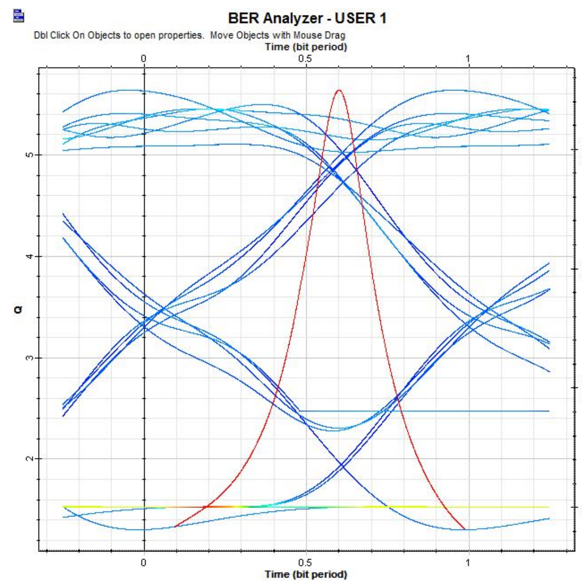


Figure 4. Plot of BER for NRZ format.

Further the results shows that the bit error rate in case of non return to zero format is 7.92e-009, eye height is 119773 and threshold value is 197645 and bit error rate in case of return to zero format is 9.08e-009, eye height is 51717.8 and threshold value is 69936.

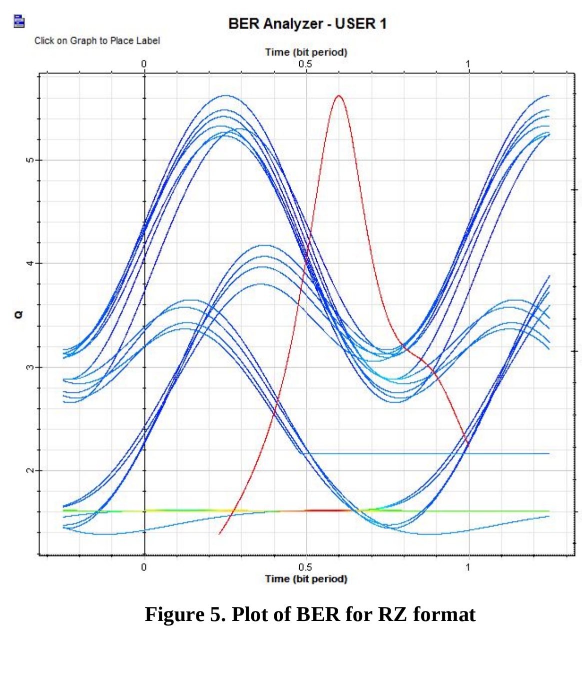


Figure 5. Plot of BER for RZ format

REFERENCES

[1] Cheing-Hong Lin et.al (2005) “Spectral Amplitude Coding optical CDMA system using Mac-zahnder interferometers” Journal of lightwave technology, vol. 23, no 4, pp. 1543-1555.

[2] Hassan Yousif Ahmed. et. al(2011) “Optical Performance analysis of spectral amplitude coding optical CDMA systems with new subtract exclusive OR detection(SED) using vectors combinatorial(VC) code”. Optic 123(2012) 1352-1359.

[3] Manjit Singh et. al (2012) “Simulative analysis of the impact of non-linearity in terms of BER, Q-Factor and Jitter in optical soliton transmission at 40 GB/s”. International journal of computer applications 48(9) 9-12.

[4] http://www.optiwave.com/products/system_overview.html

[5] S.A. Aljunid, Zuraidah Zan, Siti Barirah Ahmad Anas, M.K. Abdullah, A new code for optical code division multiple access systems, Malaysian J. Comput. Sci. 17 (2004) 30–39.

[6] M.K. Abdullah, Feras N. Hasoon, S.A. Aljunid, Sahbudin Shaari, Performance of OCDMA systems with new spectral direct detection (SDD) technique using enhanced double weight (EDW) code, Opt. Commun. 281 (2008) 4658–4662.

[7] Zou Wei, H. Ghafouri-Shiraz, Codes for spectral-amplitude-coding optical CDMA systems, J. Lightwave Technol. 20 (2002) 1284–1291.

[8] S.A. Aljunid, Zuraidah Zan, Siti Barirah Ahmad Anas, M.K. Abdullah, A new code for optical code division multiple access systems, Malaysian J. Comput. Sci. 17 (2004) 30–39.

Mobile Phone Jammer

Shamsher Singh
M.Tech, ECE,
Amritsar College of Engineering and Technology, Amritsar, India
shamsherchandi@gmail.com

***Abstract*--The last few years have witnessed a dramatic boom in the wireless communications industry, hence increasing the number of users of mobile communication devices. This magnified the need for a more efficient and reliable signal scrambler. This paper deals with the Mobile Jamming Technology. The concept of jamming technology is studied in a step-by step approach. The mobile jammer in the frequency range of 890MHz to 960MHz (GSM) is developed. Its circuit analysis simulation is performed using Speace spice Software. Antenna simulation is done by using IE3D software.The jammer circuits designed with minimum cost and high efficiency. The jammer jams the signal within five meter effective radius.**

keywords–Antenna, Jammer, Wideband frequency

I. INTRODUCTION

The Mobile Phone Jammer is designed to block communications between mobile phones and cellular base station, without interfering with other communication systems. Communication jamming devices were first developed and used by military. This interest comes from the fundamental objective of denying the successful transport of information from the sender (tactical commanders) to the receiver (the army personnel), and vice-versa. Nowadays, mobile (or cell) phones are becoming essential tools in our daily life. Here in Jordan, for example, with a rather low population (around 5 million), three main cell phone carries are available; namely; Zain, Orange, and Umniah The first two use the GSM 900system, while the third uses the GSM 1800 system. Needless to say, the wide use of mobile phones could create some problems as the sound of ringing becomes annoying or disrupting. This could happen in some places like conference rooms, law courts, libraries, lecture rooms and mosques. One way to stop these disrupting ringing is to install a device in such places which will inhibit the use of mobiles, i.e., make them obsolete. Such a device is known as cell phone jammer or "GSM jammer", which is basically some kind of electronic countermeasure device. The technology behind cell phone jamming is very simple. The jamming device broadcasts an RF signal in the frequency range reserved for cell phones that interferes with the cell phone signal, which results in a "no network available" display on the cell phone screen. All phones within the effective radius of the jammer are silenced. Various fundamentals discuss in this paper are:

- Effects on Mobile (CDMA/GSM) phones
- Unable to initiate calls or SMS messages
- Unable to receive calls or SMS messages
- Unable to communication between mobile phone and base station
- Disconnect call upon entering covered area

BLOCK DIAGRAM OF JAMMER

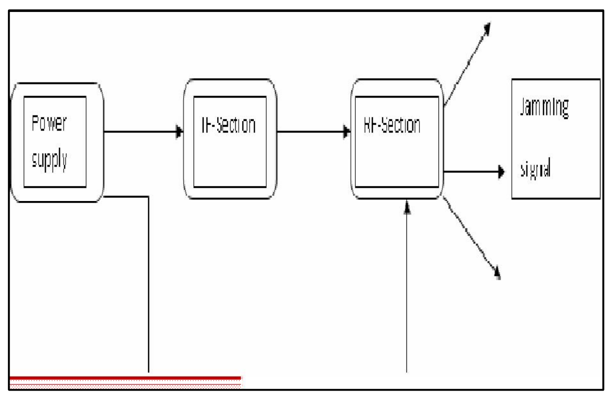


Fig. 1: Various blocks of jammer

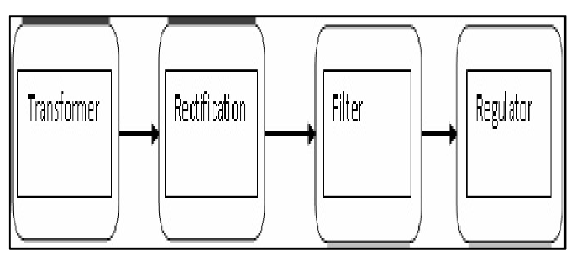


Fig. 2: Power supply

This is used to supply the other sections with the needed voltages. Any power supply consists of the following main parts:

Transformer: -is used to transform the 220VAC to other levels of voltages.

Rectification: - This part is to convert the AC voltage to a DC one. We have two methods for rectification: Half wave-rectification: the output voltage appears only during positive cycles of the inputsignal. Full wave –rectification: a rectified output voltage occurs during both the positive and negative cycles of the input signal.

The Filter: used to eliminate the fluctuations in the output of the full wave rectifier “eliminate the noise” so that a constant DC voltage is produced. This filter is just a large capacitor used to minimize the ripple in the output.

Regulator: This is used to provide a desired DC-voltage.The general parts of the power supply.

THE IF-SECTION

The tuning section of the Jammer sweeps the VCO through the desired range of frequencies. Basically, it is just a triangle or saw tooth-wave generator; offset at a proper amount so as to sweep the VCO from the minimum desired frequency to a maximum. The tuning signal is generated by a triangular wave mixed with noise. The IF section consists of three main parts:

1)Triangle wave generator. (To tune the VCO in the RF section)

2) Noise generator (provides the output noise).

3) Mixer” summer” (to mix the triangle and the noise waves

Noise generation

Without noise, the output of the VCO is just an unmodulated sweeping RF carrier. So, we need to mix the triangular signal with noise (FM modulating the RF carrier with noise). To generate noise signal, we used the Zener Diode operated in reverse mode. Operating in the reverse mode causes what is called avalanche effect, which causes wide band noise. This noise is then amplified and used in our system. We use two amplification stages: in the first stage, we use NPN transistor as common emitter, and in the second stage, we use the LM386IC {Audio amplifier}.

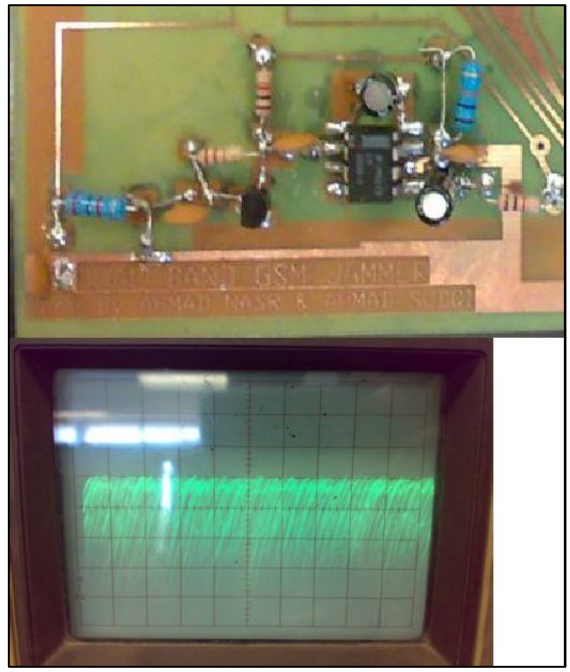


Fig. Noise generated signal

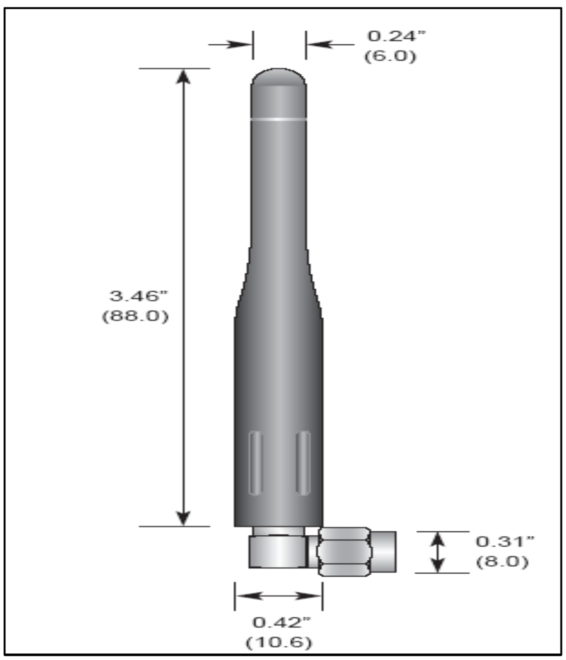
II. The RF-Section

This is the most important part of the jammer, since the output of this section will beinterfacing with the mobile. The RF-section consists of three main parts: voltage controlledoscillator VCO, power amplifier and antenna.The voltage controlled oscillator (VCO)is the heart of the RF-section. It is the device thatgenerates the RF signal which will interfere with the cell phone. The output of the VCO has a frequency which is proportional to the input voltage, thus, we can control the output frequency by changing the input voltage. When the input voltage is DC, the output is aspecific frequency, while if the input is a triangular waveform, the output will span a specific frequency range. In our design, we need to find a VCO for GSM 900 and GSM 1800. There are three selection criteria for selecting a VCO for this application. Most importantly, it should cover the bands that we need, secondly, it should be readily available at low cost, and finally, it should run at low power consumption. Moreover, we need to minimize the size of GSM-jammer. So, we started to search through the internet for

VCO's that work for GSM900 & GSM 1800 bands. The power amplifier: Since 5 dBm output power from the VCO does not achieve the desired output power of the GSM jammer, we had to add an amplifier with a suitable gain to increase the VCO output to 34 dBm. Upon testing, the jammer didn't work properly. It was concluded that amplifier IC does not work at the two bands simultaneously. Such a fact was not indicated in the datasheets. This result was really a big shock, but easily solved by changing the whole RF design. The new design uses two power amplifier IC's instead of one amplifier.

Antenna:

A proper antenna is necessary to transmit the jamming signal. In order to have optimal power transfer, the antenna system must be matched to the transmission system. In this project, we used two 1/4 wavelength monopole antennas, with 50Ω input impedance so that the antennas are matched to the system. We used monopole antenna since the radiation pattern is omni-directional.



III. OPERATION

As with other radio jamming, cell phone jammers block cell phone use by sending out radio waves along the same frequencies that cellular phones use. This causes enough interference with the communication between cell phones and towers to render the phones unusable. On most retail phones, the network would simply appear out of range. Most cell phones use different bands to send and receive communications from towers (called frequency division duplexing, FDD). Jammers can work by either disrupting phone to tower frequencies or tower to phone frequencies.

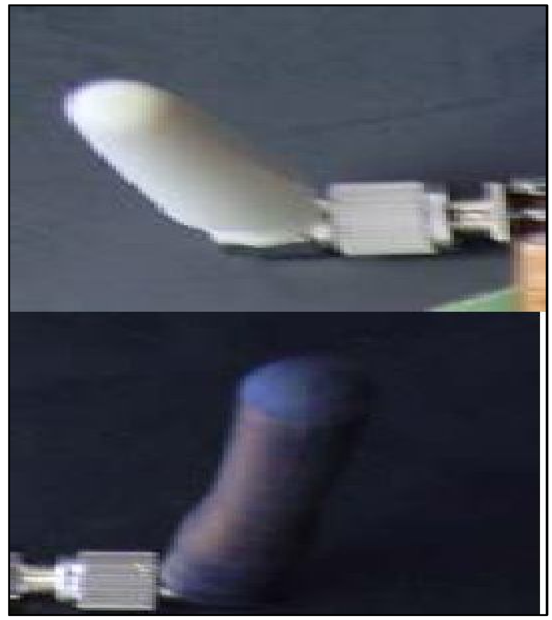


Fig. Antennas of mobile phone jammer

Smaller handheld models block all bands from 800 MHz to 1900 MHz within a 30-foot range (9 meters). Small devices tend to use the former method, while larger more expensive models may interfere directly with the tower. The radius of cell phone jammers can range from a dozen feet for pocket models to kilometers for more dedicated units. The TRJ-89 jammer can block cellular communications for a 5-mile (8 km) radius.Less energy is required to disrupt signal from tower to mobile phone, than the signal from mobile phone to the tower (also called base station), because the base station is located at larger distance from the jammer than the mobile phone and that is why the signal from the tower is not as strong. Older jammers sometimes were limited to working on phones using only analog or older digital mobile phone standards. Newer models such as the double and triple band jammers can block all widely used systems (CDMA, iDEN, GSM, et al.) and are even very effective against newer phones which hop to different frequencies and systems when interfered with. As the dominant network technology and frequencies used for mobile phones vary worldwide, some work only in specific regions such as Europe or North America

FREQUENCY BANDS

Table 1: Operating frequency bands

	UPLINK	DOWNLINK
GSM 900	890-915 MHz	935-960 MHz
DCS 1800	1710-1785 MHz	1805-1880 MHz

The jamming frequency must be the same as the downlink, because it needslower power to do jamming than the uplink range and there is no need to jam the base stationitself. So, our frequency design will be as follows:GSM 900 935-960 MHzGSM 1800 1805-1880

IV. MOBILE JAMMING TECHNIQUES

Type "A" Device: JAMMERS

In this device we overpower cell phone's signal with a stronger signal, This type of device comes equipped with several independent oscillators transmitting 'jamming signals' capable of blocking frequencies used by paging devices as well as those used by cellular/PCS systems' control channels for call establishment. When active in a designated area, such devices will (by means of RF interference) prevent all pagers and mobile phones located in that area from receiving and transmitting calls. This type of device transmits only a jamming signal and has very poor frequency selectivity, which leads to interference with a larger amount of communication spectrum than it was originally intended to target. Technologist Jim Mahan said, "There are two types. One is called brute force jamming, which just blocks everything. The problem is, it's like power-washing the airwaves and it bleeds over into the public broadcast area. The other puts out a small amount of interference, and you could potentially confine it within a single cell block. You could use lots of little pockets of small jamming to keep a facility under control."

Type "B" Device: INTELLIGENT CELLULAR DISABLERS

Unlike jammers, Type "B" devices do not transmit an interfering signal on the control channels. The device, when located in a designated 'quiet' area, functions as a 'detector'. It has a unique identification number for communicating with the cellular base station. When a Type "B" device detects the presence of a mobile phone in the quiet room; the 'filtering' (i.e. the prevention of authorization of call establishment) is done by the software at the base station

Type "c" Device: INTELLIGENT CELLULAR DISABLER

Unlike jammers, Type "C" devices do not transmit an interfering signal on the controlchannels. The device, when located in a designated 'quiet' area, functions as a 'beacon' and anycompatible terminal is instructed to disable its ringer or disable its operation, while within thecoverage area of the beacon. Only terminals which have a compatible receiver would respondand this would typically be built on a separate technology from cellular/PCS, e.g., cordlesswireless, paging, ISM, Bluetooth. On leaving the coverage area of the beacon, the handset must re-enable its normal function. This technology does not cause interference and does not require any changes toexisting PCS/cellular operators. The technology does require intelligent handsets with a separatereceiver for the beacon system from the cellular/PCS receiver. It will not prevent normaloperation for incompatible legacy terminals within a "quiet" coverage area, thus effectivedeployment will be problematic for many years.While general uninformed users would lose functionality, pre-designated"emergency" users could be informed of a "bypass terminal key sequence" to inhibit response tothe beacon.

Type "D" Device: DIRECT RECEIVE & TRANSMIT JAMMERS

This jammer behaves like a small, independent and portable base station, which can directly interact intelligently or unintelligently with the operation of the local mobile phone. The jammer is predominantly in receiving mode and will intelligently choose to interact and block the cell phone directly if it is within close proximity of the jammer. This selective jamming technique uses a discriminating receiver to target the jamming transmitter. The benefit of such targeting selectivity is much less electromagnetic pollution in terms of raw power transmitted and frequency spectrum from the jammer, and therefore much less disruptive to passing traffic. The jam signal would only stay on as long as the mobile continues to make a link with the base station, otherwise there would be no jamming transmission – the technique forces the link to break or unhook and then it retreats to a passive receive mode again

T*ype "E" Device: EMI SHIELD -PASSIVE JAMMING*

This technique is using EMI suppression techniques to make a room into what is called a Faraday cage. Although labor intensive to construct, the Faraday cage essentially blocks, or greatly attenuates, virtually all electromagnetic radiation from entering or leaving the cage – or in this case a target room. With current advances in EMI shielding techniques and commercially available products one could conceivably implement this into the architecture of newly designed buildings for so-called "quiet-conference" rooms. Emergency calls would be blocked unless there was a way to receive and decode the 911 transmissions, pass by coax outside the room and re-transmitted.

V. APPLICATIONS

1] Cell phone jammers can be particularly useful is at a school or university. By blocking cell phone signals, students cannot become distracted by their phones. In addition, they cannot cheat by sending text messages to one another during exams.
2] High security locations, such as prisons and detention centers, can also benefits from cell phone jamming because it can prevent illicit communication between inmates and visitors.
3] Mobile phone jammer can also be beneficial in places such as a movie theater or a library where other patrons expect silence, so they can enjoy their activities.
4] Please note that cell phone jammers are not intended to harm. In fact, even when jamming cell phone signals, emergency services frequencies are not disturbed.
5] In order to give yourself and your patrons the peace and quiet they deserve, while still allowing those in your building or facility to remain safe, you may want to seriously consider purchasing a mobile phone jammer.

VI. CONCLUSION

This paper is mainly intended to prevent the usage of mobile phones in places inside its coverage without interfering with the communication channels outside its range, thus providing a cheap and reliable method for blocking mobile communication in the required restricted a reason. Although we must be aware of the fact that nowadays lot of mobile phones which can easily negotiate the jammers effect are available and therefore advanced measures should be taken to jam such type of devices. These jammers includes the intelligent jammers which directly communicates with the GSM provider to block the services to the clients in the restricted areas, but we need the support from the providers for this purpose.

REFERENCES

[1] Rick Hartley, RF / Microwave PC Board Design and Layout, Avionics Systems.
[2] John Scourias, Overview of the Global System for Mobile Communications,University of Waterloo.
[3] Ahmed Jisrawi, "GSM 900 Mobile Jammer", undergrad project, JUST, 2006.
[4] Limor Fried, Social Defense Mechanisms: Tools for Reclaiming our Personal Space.
[5] Siwiak, K., Radio-wave propagation and Antennas for personal communication.
[6] Pozar, D., Microwave Engineering, John Wiley and Sons, 2005.
[7] "FREQUENCY PLANNING AND FREQUENCY COORDINATION FOR THEGSM 900, GSM 1800, E-GSM and GSM-R LAND MOBILE SYSTEMS (Except direct mode operation (DMO) channels)" by Working Group Frequency Management"(WGFM).

IEEE 802.11 Networks: Load Balancing Techniques

Anna Shukla[1], Mandeep Kumar[2] and Dr. V.K.Banga
[1, 2] Assistant Professor, Department of Computer Applications, ACET, Amritsar, India
[2]Professor & HOD, Department of Electronics & Communication Engineering, ACET, Amritsar, India
Email: anna.shukla@gmail.com, mankumin@yahoo.in, vijaykumar.banga@gmail.com

Abstract* - IEEE 802.11 networks are widely deployed for providing Internet access. In these scenarios, Users (WS) tend to be nomadic. The association and roaming decisions are made by WS to select AP(s) to camp on. In this paper we have discussed load balancing problem and reviewed state-of-the-art solutions towards this problem.

Keywords: IEEE 802.11, Load Balancing, Wi-Fi, Congestion Control

I. Introduction

Wi-Fi Internet access is mushrooming and competing with legacy wireless (cellular) networks to support mobile and data-centric applications. Telecom operators and service providers are gradually changing their marketing strategies and complementing their traditional offer with Wi-Fi [1]. A Wi-Fi network is characterized by a set of base stations (also called Access Points (AP)) placed throughout the environment and connected to the traditional wired LANs or nomadic users (also called Wireless Stations (WS). The primary job of an access point is to broadcast a wireless signal that WS can detect and "tune" into. In order to connect to an access point and join a wireless network, computers and devices must be equipped with wireless network adapters.

This paper is structured as follows: section 2 discusses the various load metric(s). Section 3 provides an overview of a classical approach to associate a WS to AP which leads to the problem of load balancing. Section 4 discusses the possible load balancing techniques and conclusions are given in section 5.

II. Load Metrics

The primary issue while selecting an AP is to measure the workload of APs present in a domain. As we can't measure the APs load directly, we use some tortuous metrics for this, which are as follows:

WS Count: - it is a straightforward load metric which tells the total number of WS associated with an AP. Although WS count is not included as information in probe response or bacon frames of IEEE 802.11, in IEEE 802.11e, the QBSS (QoS Enhanced Based Service Set) load element contains information about current WS population of an AP. Unfortunately, this information is not vital for load balancing as it doesn't make sure whether the associated WS are using the available bandwidth and if yes to how much extent are they using it.

Channel Utilization: it is the percentage of time during which an AP is busy transmitting or receiving data which can tell about the varying traffic conditions. Channel utilization still is not an appropriate AP selection metric, as it does not capture transmission capabilities of respective APs: an 80% utilized IEEE 802.11g AP can offer even more bandwidth than a 40% utilized IEEE 802.11b AP.[1]

Throughput: Network Throughput refers to the volume of data that can flow through a channel in a given amount of time. However, throughput in wireless networks is constrained further by the capabilities of network adapters on WS. Some AP selection heuristics favor APs that maximize expected throughput or potential bandwidth [1].

*Received Signal Strength Indicator (RSSI):*In an IEEE 802.11 system RSSI is a measurement of the power present in a received signal from an AP. Therefore, the higher the RSSI number, the stronger the signal.

III. Association of WS to AP: A Classical Approach

The currently implemented procedure used by most manufacturers for the association of a WS to an AP is as follows:

A WS scans the available channels of each AP in the region and listens to the Beacon or Probe Response Frames. The WS stores the RSSI of Beacon or Probe Response Frames and other relevant information, as ESSID, encryption (on/off), etc. After finishing the scanning procedure, the WS selects that AP with the maximum RSSI. The WS will leave the AP when the RSSI falls under a predefined threshold. This procedure is based on the conviction that the quality of service of the so selected AP is the best. However, this procedure leads to the result that many stations are connected to a few APs, while some other neighboring APs remain idle. [2] This overloading of the AP will lead to performance degradation also called load unbalancing. Therefore, techniques are needed that will take into consideration the status of each AP and it's already associated WS, in order to associate new WS to an AP and this phenomena is called load balancing.

IV. Load Balancing Techniques

We categorize load-distribution schemes as WS-based and network-based, depending on which part of the network is in charge of load distribution.

A. WS- Based Load Distribution

In a WS-based approach, WSs learn of APs' load status somehow and, accordingly, select an AP that maximizes their potential benefits. APs act passively in the whole selection process. [3]

There are two basic techniques for WS to make association with AP:

Least Load First: Many WS-based approaches are not designed to achieve system-wide load balance-WSs select APs simply for their own interests. However, seeking an AP that provides the maximal available bandwidth implicitly implements least-load-first AP selection, a widely-used load-balancing heuristic. The acquisition of APs' load condition may be realized in several ways. A WS may measure channel utilization or the delay between the scheduled and actual transmission time of periodic Beacon frames. [3]

Throughput Based: Throughput is a common metric for quantifying the effectiveness of load balancing schemes. We can use link-layer throughput as a direct measure of load. Most commonly used approach is to select that AP which has maximum throughput or potential bandwidth. However, throughput or bandwidth usage is affected by time-varying channel conditions, so it is very hard to measure the accurate throughput of the AP(s). In this, a WS is first associated with some AP, through which it then accesses throughput from stand-alone server. The server maintains load states for all APs residing in its administrative domain and periodically poll to WS about the load of these APs. With this information of bandwidth consumption, the WS then decides whether a handoff should be conducted to distribute the load. The selection is based on expected bandwidth and round-trip time.[3]

The WS-AP association management can proceed in a static or dynamic fashion. In static cases, a WS performs AP selection prior to its association with the target AP and does not re-associate to other APs as long as the association holds. A drawback of static AP selection is the inflexibility to adapt to network dynamics. With dynamic AP selection, on the other hand, a WS may determine to reassociate with another AP even if the current association still holds. Dynamic AP selection is better suited to highly dynamic networking environments. However, it may also lead to unstable WS-AP associations or so-called ping-pong effects, the phenomenon of repeated association changes from one AP to another.

To avoid ping-pong effect, either static AP selection or back0ff scheme is used to suppress the burst of association migrations. [3]

.

B. Network-based Load Distribution

In a network-based approach, WSs behave passively in modifying AP-WS associations. It is a network-side entity (could be an AP, a switch, or a dedicated server) that controls the distribution of AP's load. There are three basic techniques for APs to control their own load level:

Cell Breathing: Crowded APs can reduce the transmission power of their beacon signal so that new WSs are less likely to discover them. APs may collaborate in adjusting their radio coverage patterns in a way that lightly-loaded APs cover more area than heavily-loaded ones, and there is no coverage hole to ensure continuous coverage.[3]

Scrutinize Entry. An overloaded AP may simply reject new association requests. A non-overloaded AP decides whether it should grant association requests from WSs based on work-load status. The request can be granted only when the predicted load level after the association does not exceed some threshold.

Association Management. A crowded AP may send an unsolicited disassociation frames to selected WSs that are already associated with it, hoping that these WSs would re-associate with other lightly-loaded APs. Theoretically, the best disassociation candidate is the one for which the corresponding re-association balances the load among related APs. However, it is impractical to seek such an optimal solution in a fast-changing networking environment as the optimum holds only for the current state. To finds good candidates, the AP may need to know the load level of neighboring. The ping-pong effects may still be a problem if we allow a disassociated WS to be disassociated again in the future.

For a global view of load distribution, APs may exchange load status information via a wired backbone. Protocols such as IAPP with slight modifications may be used for this purpose. APs identify themselves as overloaded can then use above-mentioned approaches to relieve load. Alternatively, a dedicated server located in the wired infrastructure can be used to collect load-related information. The server learns of the distribution of traffic load and recommends or instructs designated WSs (by communicating with the peers running at these WSs) to change their AP associations. The ability to estimate the bandwidth actually consumed or potentially demanded by each WS may further facilitate association migration decisions, as this information eases the following tasks:

- Determining the set of illegible APs that fulfills the implicit bandwidth demand of a particular WS.
- Predicting load shift between APs for each feasible handoff so that the best handoff target can be determined.

Price-based: In this a WLAN Operator investigates policies to control the traffic generated by its hot-spot users in order to maximize the revenues. The idea is to control the hot spot traffic by implementing a congestion-control pricing policy. At any given time instant, the access cost depends on the current load in the hot spot. When congestion increases, the WLAN Operator increases the access cost until some users give up transmitting [4]. The aim is to maintain the quality of service, and to exploit the revenues. Generally users will be charged for the number of successfully transmitted and received packets. By increasing or decreasing the per-packet cost the WLAN Operator will control the number of active users and hence the congestion in the hot-spot. The aim is to drive the hot spot to operate in a status that maximizes the operator revenues and maintains the system in the most efficient state, i.e., the maximum aggregated throughput. Specifically, we identify as the optimal operating point of the system, the status corresponding to the maximum aggregated throughput. The rational for this choice is based on the following observations:

- When the offered load is below this point, the WLAN Operator can stimulate additional traffic, thus increasing the revenues without congesting the system;
- When the offered load is higher than this point, to maintain the same revenues of the optimal operating point, the WLAN Operator must increase the costs and thus the users have to pay more for a worse quality of service.

This policy, of course, cannot be acceptable by the users that can hence drift to another WLAN Operator implementing a more fair policy and therefore providing a better quality at a lower cost. Simple mechanisms to smooth these fluctuations can be introduced. For example, gradually increasing the costs or the cost increase can be announced to subsets of randomly selected users. Price-based Congestion Control (PCC) policy that can be implemented in an 802.11 based WLAN hot-spot. The distributed mechanism for controlling the congestion in a IEEE 802.11 network to guarantee that the system operates below the optimal operating point. Furthermore, the proposed mechanism asymptotically (with respect to the number of active stations) drives the system close to the optimal operating point. The main drawback of that approach is the requirement that all the network stations use a network interface card obtained by modifying the Wi-Fi cards. Obviously, this constraint is not acceptable in a WLAN hot spot in which the WLAN operator cannot force the use of modified Wi-Fi cards. This policy not only requires additional hardware/software in the Access Point also the user's adaptation standard Wi-Fi cards.

V. Conclusion

In this paper we have reviewed existing solutions to the load balancing problem in IEEE 802.11 networks, which will increase the overall network throughput significantly if traffic load is fairly distributed among available APs. WS-based approaches are generally not customized to system-wide load balance, so these are suitable for once-only AP selection. Network-based approaches have the potential of achieving system-wide load balance if designed appropriately.

REFERENCES

[1] Li-Hsing Yen, Tse-Tsung Yeh, Kuang-Hui Chi," Load Balancing in IEEE 802.11 Networks", Internet Computing,IEEE, Vol.13 Issue-1,Jan-feb 2009.

[2] Ioannis Papanikos, Michael Logothetis, "A Study on Dynamic Load Balance for IEEE 802.11b Wireless LAN",COMCON 8, 2001.

[3] Roberto Battiti, Enrico Gregori, Mikalai Sabel, " Price- Based Congestion-Control in Wi-Fi Hot Spots" Proc. WiOpt'03, INRIA Sophia-Antipolis,2003.

[4] Eduard Garcia, Rafael Vidal, Josep Paradells, "Cooperative Load Balancing in IEEE 802.11 Networks with Cell Breathing" Computers and Communications, ISCC 2008. IEEE , 1133 – 1140, 2008.

[5] [IEEE802] Official Homepage of the IEEE 802.11 Working Group, http://grouper.ieee.org/groups/802/11/.

[6] T. Henderson, D. Kotz, and I. Abyzov, "The changing, usage of a mature campus-wide wireless network," in 10th Mobicom, pp. 187-201, September 2004.

[7] N.Prasad, A. Prasad, "WLAN Systems and wireless IP for next generation communications", Artech House, 2001.

[8] G. Bianchi and I. Tinnirello, "Kalman Filter estimation of the number of competing terminals in an IEEE 802.11 network," in 22nd Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM'03,vol. 2, pp. 844-852, April 2003.

[9] Balachandran, A., Voelker, G.M., Bahl, P., Rangan, P.V.: Characterizing User Behavior and Network Performance in a Public Wireless LAN. ACM SIGMETRICS Int. Conference on Measurement and Modelling of Computer Systems. Marina Del Rey, California (2002).

[10]Chiu, Dah-Ming, Jain, : Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks. Journal of Computer Networks and ISDN, Vol. 17, N. 1. (1989) 1-14

[11] Bob O'Hara and Al Petrick, "IEEE 802.11 Handbook – A Designer's Companion", IEEE Press, 1999.

[12] Shiann-Tsong Sheu and Chih-Chiang Wu, "Dynamic Load Balance Algorithm (DLBA) for IEEE 802.11 Wireless LAN", Tamkang Journal of Science and Engineering, Vol 2, No 1, pp.45-52 (1999).

Review Paper on Iris Recognition System

Anu sharma
M.Tech, ECE, Amritsar college of Engineering and Technology, Amritsar.
anu29_sharma@yahoo.co.in

***Abstract*—Iris detection is an important biometrics technique used for security purposes. It works by extracting the important features of iris and then comparing them with database. It is popular because of unique nature and extreme richness of the human iris. The iris recognition technique consists of segmentation, normalization, encoding and comparison .In this paper various IRIS segmentation and normalization methods have been discussed. Also it is observed that J.Daugman's method provides highest accuracy. In addition to this, advantages and disadvantages and comparison of iris with other biometrics techniques is also discussed.**

***Keywords*— Biometrics, Contour, Template, FAR, FRR.**

I. INTRODUCTION

Iris recognition is a part of biometric identification methods which also include face recognition, fingerprint, and many other biological traits. These all are new methods for a person identification, authentication and security. Currently users have to carry security badges or certain known pin/pass codes in order to get into secure zones or to log in into a computer. Problem with these methods is that users have to remember lot of diff erent passwords and pincodes. These are easy to guess and crack because users prefer passwords that are easy to remember. Cards can be lost and they can be used by anyone else to gain access to a restricted area, place or to a restricted computer .Biometrics on the other hand provide a certain and easy way of authenticating persons, biometrics can be combined with some other method like password, they form up a very strong authentication method.

Biometric identification utilises many psychological and physical characteristics of an individual. Some common features are fingerprints, hand shapes, eyes retinas and many others, including eye's iris. Psychical and behavioural characteristics include typing speed, walking style and signature etc. Out of all physiological properties iris patterns are believed to be one of the most accurate.

The process of iris recognition is real-time and highly accurate. Iris recognition has many practical uses, like it can be used to authenticate a person's identity or to identify a certain person from a large set of databases.

II. PHYSCHOLOGY OF IRIS

The iris is a protected internal organ of the eye, located behind the cornea but in front of the lens . The iris has many features that can be used to distinguish one iris from another. One of the primary visible characteristic is the trabecular meshwork, a tissue which gives the appearance of dividing the iris in a radial fashion that is permanently formed by the eighth month of gestation period. During the development of the iris, there is no genetic influence on it, a process known as chaotic morphogenesis that occurs during the seventh month of gestation, which means that even identical babies means, twins have uncorrelated minutae, i.e. differing irises. In fact, even persons own eyes are uncorrelated.

The most important function of the iris is to control the size of the pupil Illumination, which enters into the pupil and falls on the retina of the eye, is controlled by muscles of the iris. They maintain the size of the pupil and this is what permits the iris to control the amount of light entering the pupil. The change in the size results from involuntary reflexes and is not under conscious control. This feature can be used to guarantee that the image being taken is probably with a high confidence of a living eye and not an artificial image of an eye.

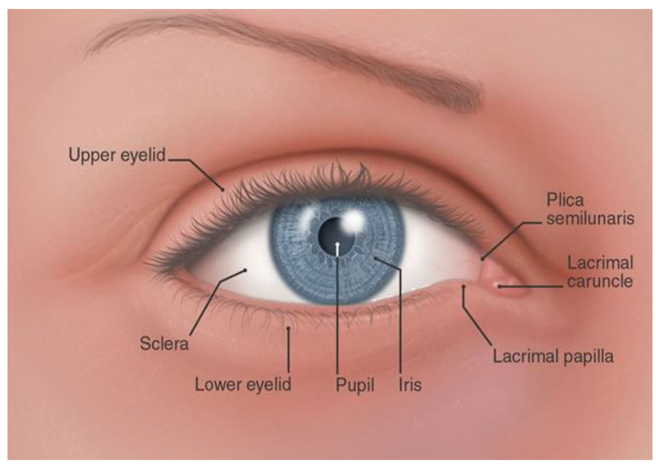
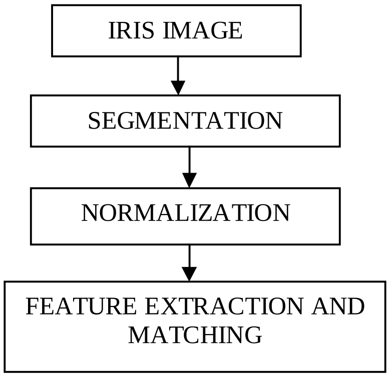


Fig 1 View of Human Eye

III. METHODOLOGY

The system of iris recognition further consists of a number of sub-systems, which correspond to each stage of iris recognition. These stages are:

- Image acquisition-capturing image of eye.
- Segmentation – locating the iris region in an eye image.
- normalization – creating a dimensionally Consistent representation of the iris region.
- Encoding – creating a template containing only the most discriminating features of the iris.
- Matching- The matching module determines how closely the produced code matches the encoded features stored in the database.



IV. DIFFERENT METHODS FOR IRIS RECOGNITION OR SEGMENTATION

A) Phase-based method

D*augman's integro differential operator:* The phase based method recognize iris patterns using the phase information. Phase information is independent Of both imaging contrast and illumination. J.Daugman designed this system in 1994. The pupil and iris boundary was found using integro differential operator given in following Equation

$$\max_{r,xo,yo} \left| G(r) * d/dr \ \square_{r,xo,yo} \ I(x,y)/2\Pi r \, ds \right| \qquad (1).$$

Where I(x,y) is the image in spatial coordinates, r is the radius, (x0,y0) are centre coordinates, the symbol * is for convolution and $G\sigma(r)$ is a Gaussian smoothing Function of scale σ. The centre coordinates and radius are estimated for both pupil and iris by determining the Maximum partial derivative of the contour integral of the image along the circular arc. The iris portion of the image I(x, y) is normalized to the polar form by the mapping function $I(x(r, \theta), y(r, \theta)) \rightarrow I(r, \theta)$. The representation of iris texture is binary coded by quantizing the phase response of a texture filter using quadrature 2D Gabor wavelets into four levels. Each pixel in the normalized iris pattern corresponds to two bits of data in the iris template i. e. A total of 2,048 bits are calculated making template, and an equal number of masking bits are generated in order to mask out corrupted regions within the iris. This creates a compact 256- byte template, which allows for storage and comparison between different iris codes. Iris codes are different for two different samples, the test was performed using Boolean XOR operator applied to 2048 bit phase vectors to encode any two iris patterns, masked (ANDed) by both of their bit vectors. From the resultant bit vector and mask bit vectors, the dissimilarity measure between any two iris patterns is computed using Hamming Distance (HD) given in following Equation

HD= (code A xor code B) □ mask A □ mask B /mask A □ mask B. (2).

Where codeA and CodeB are two phase code bit vectors and mask and mask are mask bit vectors. The HD is a Fractional measure of dissimilarity with 0 representing a Perfect match. A low normalized HD implies more Similarity of iris codes.

B) Texture-analysis based method

Laplacian of Gaussian Filters: In order to encode features of iris, the Wildes et al. system decomposes the iris region by application of Laplacian of Gaussian filters to the iris region image. The filters are given as

$$\blacktriangledown G = -1/2\Pi\sigma^4 \ \{ 1-\rho^4/ 2\sigma^2\} e^{\rho 2/2\sigma 2} \qquad (3).$$

Where σ is the standard deviation of the Gaussian and ρ is the radial distance of a point from the centre of the filter. The filtered image is represented as a Laplacian pyramid which is able to compress the data, so that only significant data remains.

A Laplacian pyramid is constructed with four different resolution levels in order to generate a compact type iris template. The method for iris identification by Emine Krichen use a hybrid method for iris segmentation, Hough transform for outer iris boundary and integrodifferential operator for inner iris boundary, then the iris code was produced using wavelet packets. The whole image is analyzed at different resolutions. 832 wavelets with 4 scales are used to generate 1664 bits code. An improvement of 2% FAR and 11.5% FRR was obtained relative to Daugman method. It was observed that by considering colour information, overall improvement of 2% to 10% was obtained according to threshold value.

C) Approach based on intensity variations

Hough Transform :The Hough transform is a standarized and simple computer based algorithm that can be used to determine the parameters of simple geometric objects, such as lines and circles. The circular Hough transform can be employed to deduce the radius and centre coordinates of the pupil and iris regions. An automatic segmentation algorithm based on the circular Hough transform is employed by Wildes et al. Kong and Zhang , Tisse et al. , and Ma et al.. Firstly, an edge map is generated by calculating the first derivatives of intensity values in an eye image and then thresholding the result values. From the edge map, votes are cast to determine in Hough space for the parameters of circles passing through each edge point. These parameters are the centre coordinates x_c and y_c, and the radius r of the circle, which are able to define any circle according to the equation

$$Xc^2 + Yc^2 - r^2 = 0 \qquad (4)$$

A maximum point in the Hough space will correspond to the radius and centre coordinates of the circle which defines the edge points. Wildes et al. and Kong and

Zhang also make use of the parabolic Hough transform to detect the eyelids, using following equation

$$(-(x-h_j)\ \sin\theta_j + (y-k_s)\cos\theta_j\,)^2 = \ (\ (x-h_s)\ \cos\theta_j\ \ +(y-k_s)\ \sin\ \theta_j\ ) \qquad (5)$$

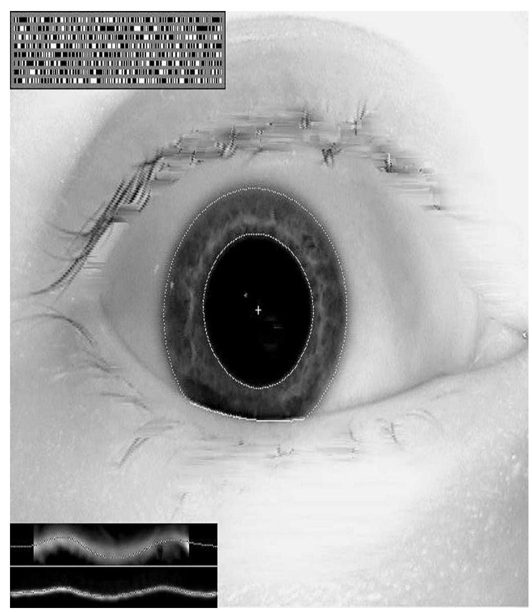
Where, (h_j, k_j) is the peak of the parabola and θ_j is the angle of rotation relative to the x-axis. In performing the preceding edge detection step, Wildes et al. bias the derivatives in the horizontal direction for detecting the eyelids and in the vertical direction for detecting the outer circular boundary of the iris. The motivation for this is that the eyelids are usually horizontally aligned, and also the eyelid edge map will corrupt the circular iris boundary edge map. Taking only the vertical gradients for locating the iris boundary will reduce influence of the eyelids when performing circular Hough transform, and not all of the edge pixels defining the circle are required for successful localization. Not only does this make circle localization more accurate, it also makes it more efficient, since there are less edge points to cast votes in the Hough space but , There are a number of problems with the Hough transform method. First of all, it requires threshold values to be chosen for edge detection, and this may result in critical edge points being removed, resulting in failure to detect circles/arcs. Secondly, the Hough transform is computationally intensive due to its' brute-force' approach and thus may not be suitable for real time applications.

D) .Approach using Independent Component Analysis

The iris recognition system developed by Hamed Ranjzad uses Independent Component Analysis (ICA) to extract iris texture features. Image acquisition is performed at different illumination and noise levels. The iris localization is performed using integrodifferential operator and parabolic curve fitting together. From the inner to outer boundary of iris, fixed number of concentric circles n with m samples on each circle is obtained. This is represented in form of matrix n x m for a specific iris image which is invariant to rotation and size. The independent components are determined from the feature coefficients. The feature coefficients are non-Gaussian and mutually independent. The independent components are estimated and encoded. The centre of iris is determined by competitive learning mechanism which is stored as the iris code for each and every person. The average Euclidean distance classifier is used to recognize iris patterns.

E).Active contours and generalized Coordinates

Iris recognition begins with finding an iris in an image, i. e. finding its inner and outer boundaries at the pupil and sclera, detecting the upper and lower eyelid boundaries if they occlude and then detecting and excluding any superimposed eyelashes or reflections from the cornea or eyeglasses. These processes may collectively be called segmentation. Precision in assigning the true inner and outer iris boundaries, even if they are partly invisible, is important because the mapping of the iris in a dimensionless (i.e., size invariant and pupil dilation invariant) coordinate system is critically dependent on this thing. Inaccuracy in the detection, modeling, and representation of these boundaries can cause different mappings of the iris pattern in its extracted description, and such differences can cause failures in matching. It is natural to start by thinking of the iris as an annulus i.e. one discovers that the inner and outer boundaries are usually not concentric. A simple solution is then to create a non concentric pseudo-polar coordinate system for mapping the iris, relaxing the assumption that the iris and pupil share a common center and requiring only that the pupil is fully contained within the iris. Performance in iris recognition is significantly improved by relaxing both of those assumptions detecting and modeling those boundaries whatever their shapes are and defining a more flexible and generalized coordinate system on this basis. Because the iris outer boundary is often partly occluded by eyelids, and the iris inner boundary may be partly occluded by reflections from illumination, and sometimes both boundaries suffer from reflections by eyeglasses, so it is necessary to fit flexible contours that can tolerate interruptions and continue their trajectory across them on a principled basis. A further constraint is that both the inner and outer boundary models must form closed curves. An excellent way to achieve all of these goals is to describe the iris inner and outer boundaries in terms of "active contours" based on discrete Fourier series expansions of the contour data. By employing Fourier components whose frequencies are integer multiples of $1/(2\pi)$. Selecting the number of frequency components allows control over the degree of smoothness. Then, truncating the discrete Fourier series after a certain number of terms amounts to low-pass filtering the boundary curvature data in the active-contour based model.



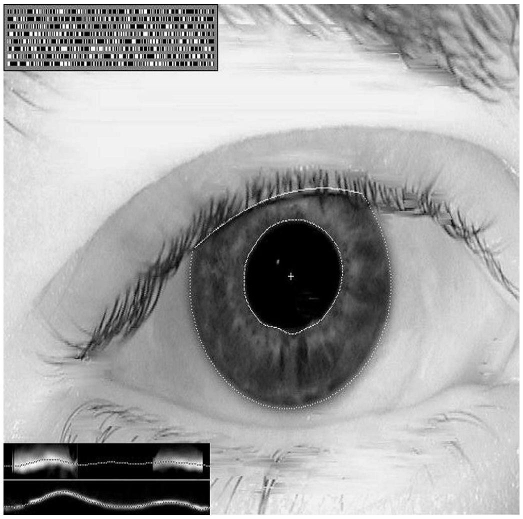


Fig 4

The lower left-hand corner of each figure shows two “snakes,” each consisting of a fuzzy ribbon-like data distribution and a dotted curve, which is a discrete Fourier series approximation of the data. The lower snake in each snake box is the curvature map for the pupil boundary, and the upper snake is the curvature map for the iris outer boundary, with the endpoints joining up at the six o’clock position. The interruptions correspond to detected occlusions by eyelids, which are indicated by separate splines in both images, or by specular reflections. Thus, the relative thickness of each snake roughly represents the sharpness of the corresponding radial edge. If an iris boundary were well-described as a circular edge, then the corresponding snake in each box should be flat and straight but, in general, this is not the case. The estimation procedure is to compute a Fourier expansion of the N regularly spaced angular samples of radial gradient edge data $\{r_\theta\}$ for $\theta = 0$ to $\theta = N - 1$. A set of M discrete Fourier coefficients $\{C_k\}$, for $k = 0$ to $k = M - 1$, is computed from the data sequence $\{r_\theta\}$ as follows:

$$c_{k=}\sum_{\theta=0}^{N-1} r\theta\, e^{-2i\Pi k\theta/N}$$

Here, the zeroth-order coefficient or “DC term” C_0 extracts information about the average curvature of the (pupil or outer iris) boundary, or in other words, about its radius when it is approximated as just a simple circle.

From these M discrete Fourier coefficients, an approximation to the corresponding iris boundary is obtained as the new sequence $\{R_\theta\}$ for $\theta = 0$ to $\theta = N - 1$, which is expressed as follows:

$$R_{\theta=}\ 1/N \int_{k=0}^{M-1} Ck e^{2\Pi ik\theta/N}$$

As it is generally true in case of active-contour methods , there is a tradeoff between how precisely one wants the model to fit all the data (improved by increasing M) versus how much one wishes to impose constraints such as keeping the model simple and of low-dimensional curvature (achieved by reducing M; e.g., $M = 1$ enforces a circular model). Thus, the number M of Fourier coefficients specifies the degrees of freedom in the shape model. It has been found that a good choice of M for capturing the true pupil boundary with appropriate fidelity is $M = 17$, whereas a good choice for the iris outer boundary where the data is often much weaker is $M = 5$. The active-contour models for the inner and outer iris boundaries supports an isometric mapping of the iris tissue between them, independent of the actual shapes of the contours.

V. IMAGE NORMALIZATION

Once the iris region is successfully segmented from an eye image, the next stage is to transform the iris region so that it has fixed dimensions in order to allow comparisons between two iris. The dimensional inconsistencies between eye images are mainly due to the stretching of the iris caused by pupil dilation from varying levels of illumination. Other sources of inconsistency include, varying imaging distance, rotation of the camera, head tilt, and rotation of the eye within the eye socket. The normalisation process will produce iris regions, which have the same constant dimensions, so that two photographs of the same iris under different conditions will have characteristic features at the same spatial location. Most normalization techniques are based on transforming iris into polar coordinates, called as unwrapping process.

A) Daugman’s Rubber Sheet Model

The homogenous rubber sheet model purposed by Daugman re maps each point within the iris region to a pair of polar coordinates (r,θ) where r is on the interval [0,1] and θ is angle $[0,2\pi]$

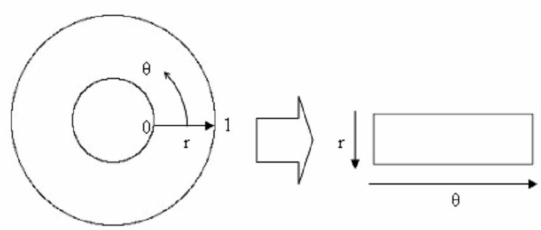


Fig 5. Daugman’s rubber sheet model

The remapping of the iris region from (x,y) Cartesian coordinates to the normalised non-concentric polar representation is modelled as

$$I(x(r,\theta), y(r,\theta)) \rightarrow I(r,\theta)$$

The rubber sheet model takes into account pupil dilation and size inconsistencies in order to produce a normalised representation of image with constant dimensions. In this way the iris region is modelled as a flexible rubber sheet anchored at the iris boundary with the pupil centre as the reference point. Even though the homogenous rubber sheet model accounts for pupil dilation, imaging distance and non-concentric pupil displacement but, it does not compensate for rotational inconsistencies.

B).Wilde's Image Registration

Wildes has proposed an image registration technique for normalizing iris textures. In this method, a newly acquired image, $I_a(u; v)$ would be aligned with an image in the database, $Id(u; v)$ to perform the comparison . The alignment process is a transformation using a choice of mapping function, $(U(x; y); V(x; y))$ that would minimize the function.

$$\square_x \square_y \ (I_d (x,y) - I_a(x - u, y-v))^2 \, dxdy$$

Wildes normalization process is based on a different approach as compared to Daugman's method. In this method, normalization is performed in the matching time as Compared to Daugman's approach; the normalization method would be time consuming in identification applications. However, for verification purposes the method is capable of compensating unwanted factors such as variations in rotation and scale.

VI. FEATURE ENCODING AND MATCHING

In order to provide accurate recognition of individuals, the important information present in an iris pattern must be extracted. Only the significant features of the iris must be encoded so that comparisons between templates can be made.

The dark intensity regions are coded by value 1 and light intensity regions are coded by 0, then comparison is done. One technique for comparing two Iris Codes is to use the Hamming distance, which is the number of corresponding bits that differ between the two Iris Codes. The binary mask computed in the normalization module ensures that the technique compares only bits corresponding to valid iris pixels. Then, if the XOR combination results in 1, there is an error in the code, and 0 corresponds to the correct code i. e ., matching of iris.

VII. ADAVANTAGES OF IRIS DETECTION

a). Very high accuracy.
b.) Verification time is generally less than 5 seconds.
c.) The eye from a dead person would deteriorate too fast to be useful, so no extra precautions have to been taken with retinal scans to be sure the user is a living human being.
d.) Reliable as compared to other biometric techniques.

VIII. DISADVANTAGES OF IRIS DETECTION

a.) Intrusive.

	Eye-iris	Eye - Retina	Fingerprint	Hand's geometry	Writing signature	Voice
Reliability	Very high	Very high	High	High	High	High
Easiness of the use	Average	Low	High	High	High	High
Attack's precaution	Very high	Very high	High	High	Average	Average
Acceptance	Average	Average	Average	High	Very high	High
Stability	High	High	High	average	Averag	Averag
Identification and authentication	Both	Both	Both	Authentication	both	Authentication
Standards			ANSI/NIST ,FBI			SVAPI

b). A lot of memory for the data to be stored.

c.) Very expensive.

IX. CONCLUSION

Iris detection is an important biometrics technique, which is preferred because of its high relaibilty and accuracy as compared to other techniques. Various methods of segmentation are introduced, and among them Hough Transform is the widely used but it has certain drawbacks.On other hand Daugman's integro differential operator provides more accuracy. So a combination of both can be used. Method based on contours is the latest one. Then after segmentation, normalization is done using various methods which converts iris region to fixed dimensions and finally encodimg ad matching is performed.

REFERENCES

[1.] www2.it.lut.fi/Kurssit/0304/010970000/seminars/vartian.pdf

[2.] www.scribd.com/doc/35149629/Iris-Recognition-System
S V Sheela, P A Vijay a, ―Iris Recognition Methods–Survey International Journal of Computer Applications

[3.] J. Daugman. Biometric decision landscapes. Technical Report No. TR482, University of Cambridge Computer Laboratory, 2000.M. Kass, A. Witkin, and D. Terzopoulos, Snakes: Active contour models,"*Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, 1988.

[4.] http://www.ijctee.org/files/VOLUME2ISSUE1/IJCTEE_0212_14.pdf

[5.] http://arxiv.org/ftp/arxiv/papers/1111/1111.5135pdf

[6.] http://www.cl.cam.ac.uk/~jgd1000/NewMethodsInIrisRecog.pdf

[7.] www.collectionscanada.gc.ca/obj/s4/f2/dsk3/owto/TC-OWTU-1045.pdf

[8.] http://www.csse.uwa.edu.au/~pk/studentprojects/libor/LiborMasekThesis.pdf

[9.] http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5647105&url=http://ieeexplore.ieee.org/xpls/abs_all.jsp%3Farnumber%3D5647105

[10.] http://biometrics.pbworks.com/w/page/14811349/Advantages%20and%20disadvantages%20of%20technologies

Comparison of Free Space Optical Communication System at 20 Gb/s by using NRZ, MDRZ, CSRZ and DRZ Formats

Rajniti[1] and Anu Sheetal[2]
[1, 2]Department of Electronics and Communication Engineering, GNDU, RC Gurdaspur, Punjab, India

Abstract-In this paper, we simulate the FSO system operating 20Gb/s for different environmental conditions i.e. clear weather, heavy rain, moderate rain, and heavy haze, using NRZ, CSRZ, DRZ, MDRZ format in order to calculate the faithful transmission distance. It is found that MDRZ data format seems to be the best choice for transmission distance despite slightly more complex transmitter configuration. Further, it is found that the faithful transmission distance at 20Gb/s for MDRZ data format is 170km in clear whether.

Keywords: *FSO, Attenuation, NRZ, CSRZ, DRZ, MDRZ.*

1. Introduction

FSO communications, also called Free Space Photonics (FSP) or wireless, refers to the transmission of modulated visible or infrared (IR) beams through the atmosphere to obtain optical communications, like, fiber FSO uses lasers to transmit data, but instead of enclosing the data streams in a glass fiber, it is transmitted through the air. The receiver's lenses able to collect the photon stream from the transmitter converts the signal back to electrical signal.

In examining FSO performance, it is important to take several system parameters into consideration. In general, these parameters can be divided into two different categories: internal parameters and external parameters. The internal parameters are related to the design of a FSO system and include optical power, wavelength, transmission bandwidth, divergence angle and optical loss on the transmitter side and receiver sensitivity, Bit-Error Rate (BER), receiver lens diameter and receiver Field of View (FOV) on the receiver side. External parameters or non-system-specific parameters are related to the environment in which the system must operate and include visibility and atmospheric attenuation, scintillation, deployment distance, window loss and pointing loss. It is important to understand that many of these parameters are not independent but are linked together in specifying overall system performance. FSO has several attractive characteristics such as (i) Dense spatial reuse (ii) Low power usage per transmitted bit (iii) License-free band of operation and (iv) Relatively high bandwidth.

In free space, normally each common optical wavelength can be used. But because of the atmospheric conditions and due to the laser safety regulations 1550nm is better suited for transmission. The losses due to Mie-scattering in haze or light fog at longer wavelengths (1550 nm) are smaller than at shorter ones (850 nm). Due to atmospheric turbulences the received polarized light changes its direction slightly, compared to the transmitted beam. The maximun value of attenuation for various different climatic conditions i.e. for clear weather, heavy rain, moderate rain and heavy haze have been considered as 1, 28, 14 and 9.8dB/km respectively. As the attenuation in the clear weather is less, it covers the maximum distance because attenuation is inversely proportional to distance covered by the FSO system. Here, issue is to increase the transmission distance by compensating atmospheric attenuation with optimized power and receiver aperture diameter.

Free space optical communication systems do not exhibit the limitations associated with the installation and maintenance of guided wave optical communication systems. It is very beneficial in urban area where the uses of lines are expensive, suitable for factories or industrial environments and the best choice to make a connection across rivers and other natural or artificial obstacles, where cables is not available. In free space, the transmitted light signal is reflected, refracted or absorbed by objects, rain, fog, wind or sunlight. It is the nomadic broadband solution (high data rates without any cabling) for connecting the backbone to the users (last-mile-access) because FSO allow communication through windows without the need for rooftop mounted antennas. A typical FSO system is capable of operating at a range of two to three times that of the naked eye in any particular environmental condition. The FSO-link should have the same reliability as the fiber link. The FSO link is mainly determined by the local weather. Therefore link availability is the key issue in FSO system deployment. Link availability comprises many factors including equipment reliability and network design, but these are well known and fairly quantifiable. The main limiting factor which affects the system performance is the attenuation of the laser power. The attenuation of the laser power depends on two main parameters: Atmospheric attenuation and geometrical loss. The first parameter describes the attenuation of the laser power in the atmosphere. The second parameter, geometrical loss, occurs due to the spreading of the transmitted beam between the transmitter and the receiver. Typically, the beam spreads to a size larger than the receiver aperture.

The biggest unknown is the statistics of atmospheric attenuation. These statistical data can be obtained either by measurements or from theoretical calculations using atmospheric optics theory. Atmospheric attenuation happens when sent signal interacts with air molecules and other particles suspended in the air (aerosols). Atmospheric attenuation of FSO systems is typically dominated by fog but can also be dependent upon low clouds, rain, haze and their various combinations. Rain has the ability to produce the effects of fluctuation in the delivery laser and the quantity of water droplet level influence the visibility of the laser transmission. The rain consists of solid vapor composed of water droplets, which are only a few hundred microns in diameter but can modify light characteristics or completely hinder the passage of light through a combination of absorption, scattering and reflection. Due to all these factors, it attenuates the transmitted signal. Another factor that decreases the FSO link distance of the transmitted signals is haze which occurs because of seasonal forest fires and air pollution.

2. System description

Typical FSO link consist of a transmitter, FSO channel, and receiver. The only difference between general communication system and FSO system is that the communication channel used for FSO is free space i.e the atmosphere. The other two components are optical transmitter and receiver, designed to meet the needs of such a specific communication channel. The FSO model is shown in Fig.1 to examine the performance of FSO system under various environment conditions. i.e clear weather, moderate rain, heavy rain, heavy haze.

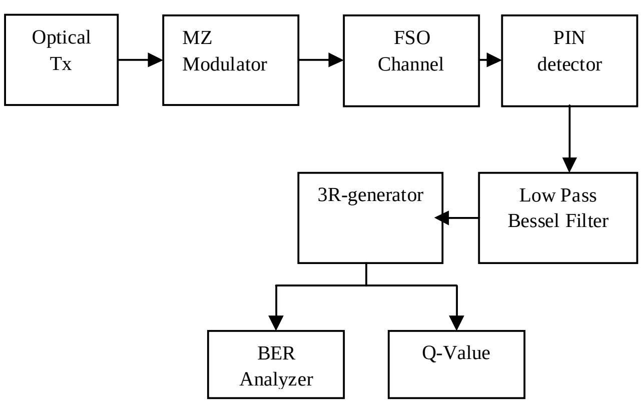


Fig.1. Block Diagram of System Setup

This model is a subsystem of transmitter telescope, free space and receiver telescope.

The Carrier-suppressed return-to-zero (CSRZ) transmitter consists of Pseudo-Random Bit Sequence (PRBS) generator, NRZ electrical driver, and Mach-Zehnder Amplitude Modulator. In this transmitter , the NRZ optical signal after MZ modulator goes through phase modulator by analog sine wave generator at the frequency equal to half the bit rate. That will introduce a ∏ phase shift between any two adjacent bits. The PRBS generator with bit rate of 20Gb/s has been used. The logical sequence generated by PRBS is converted into electrical signal using NRZ electrical pulse generator. The rise time and fall time of NRZ is taken to be 0.05 bit. The modulating signal is generated by the optical source with optical power 20dBm and FWHM line width = 10MHz operating at 1550nm. The transmitter aperture diameter is taken to be 5cm.The extinction ratio of the Mach-Zehnder Amplitude Modulator is 10dB which modulates the electrical signal.

In duobinary return-to-zero (DRZ) transmitter. The duobinary was generated by first creating an NRZ duobinary signal using a duobinary precoder, NRZ generator and a duobinary pulse generator. The generator drives the first MZM, and then cascades this modulator with a second modulator that is driven by a sinusoidal electrical signal with the frequency of 40GHz, phase = -90^{o}. The duobinary precoder used here is composed of an exclusive-or gate with a delayed feedback path. DRZ formats are very attractive, because their optical modulation bandwidth can be compressed to the data bit rate B, that is, the half-bandwidth of NRZ format 2B.

In modified duobinary return-to-zero (MDRZ) transmitter. The MDRZ was generated by first creating an NRZ duobinary signal using a delay-and-subtract circuit that drives the first MZM and then concatenating this modulator with a second modulator that is driven by a sinusoidal electrical signal with frequency of 40GHz and phase =90^{o} . The generation of MDRZ signal is almost identical to the DRZ signal, except the delay-and –add circuit is replaced by delay-and-subtract circuit. In the duobinary signal used earlier where the phase of bits ‘1’s are modified only after a bit ‘0’ appear whereas in modified duobinary signal the phase is alternated between 0 and ∏ for the bits ‘1’. The phase of all the ‘0’ bits are kept constant and a 180 ^{o} phase variation between all the consecutive ‘1’ is introduced.

The receiver consists of a PIN detector followed by low pass Bessel filter, 3R generator and a BER tester. The responsivity of the photodetector is 1A/W and its dark current is 10nA. The low pass Bessel filter has the 3dB cut off frequency = 0.75*bit rate. The depth of the 4th order Bessel filter is taken to be 100dB. We have also considered the ASE noise, shot noise, thermal noise, estimated receiver noise and ASE-ASE noise effects in optical receiver. The signal from the filter is fed to the 3R generator. It reshape and regenerates an electrical signal and feed it to the BER analyser which gives the bit error rate BER, Q factor, Eye height and threshold values of the received signal.

The FSO channel considered has the various parameters described as: the beam divergence angle = 25mrad, the transmitter aperture diameter = 5cm and the receiver aperture diameter = 7.5cm, Transmitter loss = 1dB, additional losses = 1dB. The transmission losses are due to fiber-telescope coupling and transmitter efficiency losses and additional losses are due to scintillation, mispointing etc.

4. Results and discussion

The link distance is a ratio of the received power and receiver threshold and it therefore considers all factors that attenuate the signal, divergence, atmosphere, transceivers, thermal losses and others. Another important factor in determining the link distance is BER. BER shows how many, out of the total received bits are at error. FSO system shown in Fig.1 has been numerically simulated to examine the performance of FSO system under various environment conditions. For various environmental conditions the simulation parameters considered are same. BER versus distance covered by FSO is shown in Fig.2-4. In clear weather condition, the attenuation is 1dB/km. The simulation results shows that the maximum distance covered by MDRZ and DRZ format is 160km and CSRZ and NRZ format is 13km. So the maximum distance travelled by FSO system in this case is 160km for data rate 20Gb/s.

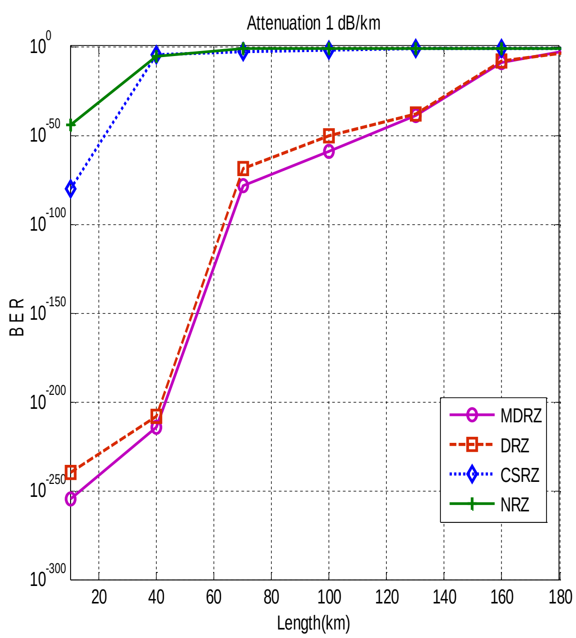


Fig. 2. Bit Error Rate (BER) versus Length for clear weather.

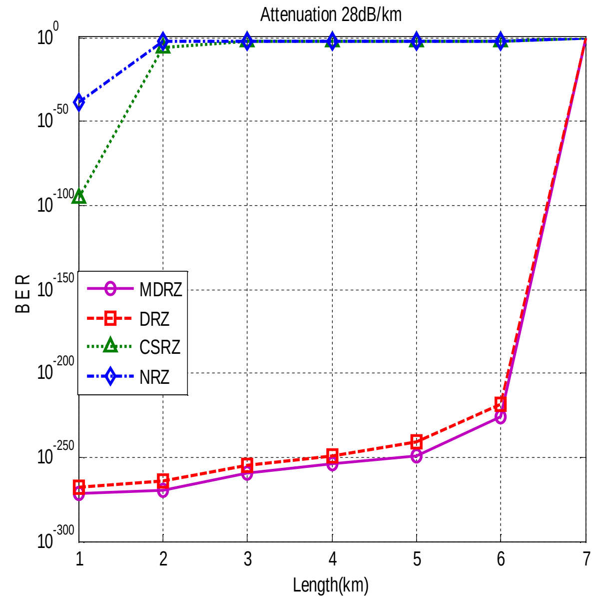


Fig. 3. Bit Error Rate (BER) versus Length for heavy rain.

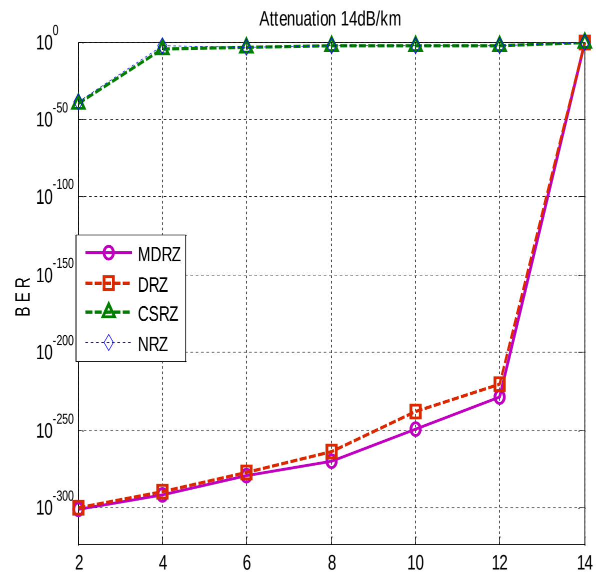


Fig. 4. Bit Error Rate(BER) versus Length for moderate rain.

During heavy rain the attenuation is 28 dB/km. It is obseved that for the heavy rain signal could be faithfully transmitted for the link distance <2km for NRZ and CSRZ data formats and for MDRZ and DRZ the faithful transmission distance is 6km. The signal would be completely lost only after 6km. The results have been shown in Fig.3.

The Fig. 4 shows the results for the moderate rain where the attenuation is 14dB/km. It is obseved that for the moderate rain signal could be successfully transmitted for the link distance upto 13km for MDRZ and DRZ formats and <4km for NRZ and CSRZ data formats.

The Fig. 5 shows that in the case of heavy haze, attenuation would be 9.8dB/km,the acheivable link length is 20km for MDRZ and DRZ formats and <4km for NRZ and CSRZ data formats.

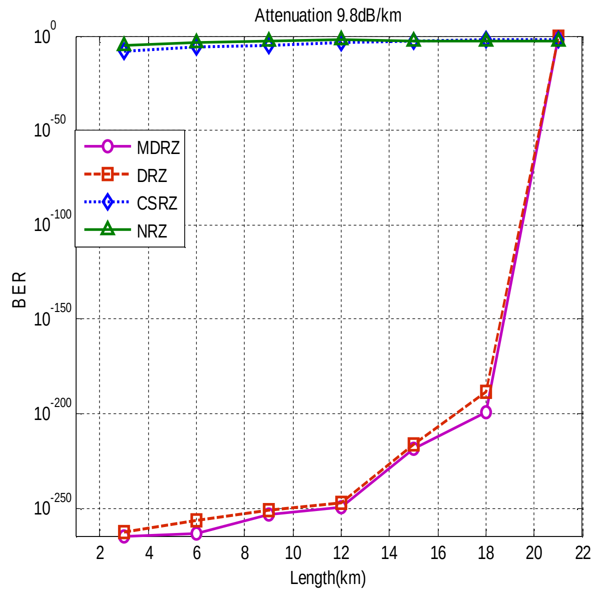
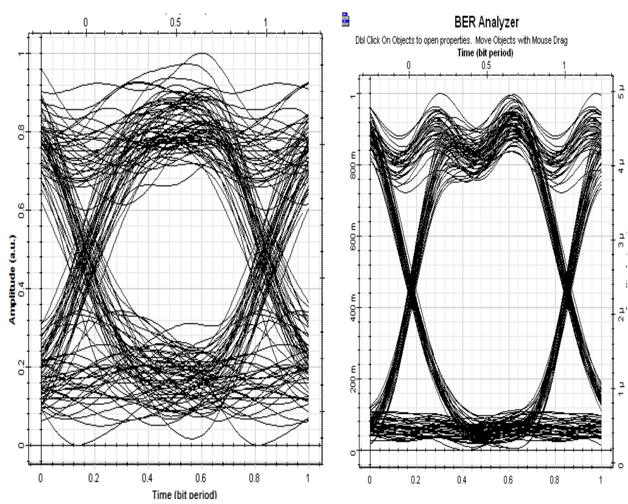
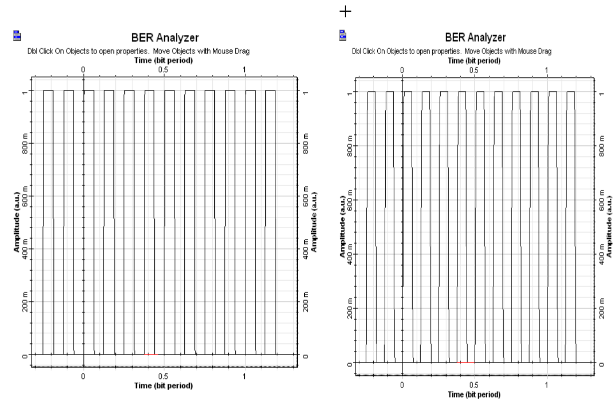


Fig. 5. Bit Error Rate (BER) versus Length for heavy haze.



(a) (b)

Fig.6. Received Eye amplitudes at 20 Gb/s bit for NRZ and CSRZ modulation format with attenuation 1dB/km at 40 km.



(a) (b)

Fig.7. Received optical signal at 20 Gb/s bit with attenuation 1dB/km at 40 km for (a) MDRZ format (b)DRZ format

Table 1 indicates the comparison of results for different data rates under different environmental conditions. From the comparision it is clear that the maximum distance covered under heavy rain is 1.21km. This can be considered as the worst case scenario.

Table 1: List of distance covered under various environment conditions

Case	***Environmental condition***	***Link range for MDRZ***	***Link range for DRZ***	***Link range for CSRZ***	***Link range for NRZ***
Case I	Clear weather	Upto 160 km	Below 160 km	Upto 16km	Upto 13km
Case II	Heavy rain	Upto 6km	Below 6km	Upto 2km	Below 2km
Case III	Moderate rain	Upto 13km	Below 13km	Upto 3.5km	Below 3.5km
Case IV	Heavy haze	Upto 20km	Below 20km	Upto 3km	Below 3km

5. Conclusion

We have simulated FSO system over a transmission distance of 160km using NRZ, CSRZ, DRZ, MDRZ modulation formats. For this we analyze the performance of system at 20Gb/s. Superior performance of MDRZ has been observed as it suppresses all the discreate frequency tones that appear in conventional RZ signal. FSO technology should be seen as supplement to conventional radio links and fiber optics. The main factor which limits the performance of an FSO system is the attenuation, which depends upon the local weather conditions. In this paper, the

FSO system for different environmental conditions(i.e. clear weather, heavy rain, moderate rain, heavy haze), operating at 20Gb/sfor various formats has been studied. The results shows that in clear weather the maximum distance covered by bit rate 20Gb/s is beyond 160km and the signal would be completely lost only after 165km. In heavy haze, the achievable link length is more than 6km. It is concluded that the link length can be increased if we overcome the effect of atmospheric attenuation and further it is overcome if we increase the transmitted power and receiver diameter.

REFERENCES

[1] E.Leitgeb, M.Loschnigg, U.birnbachar, G.Schwarz and A.Merdoning,"High Reilable Optical Wireless links for the Last mile Access", Proceedings of the IEEE, ICTON 2008, pp 178-183, 2008.

[2] Himali Sarangal, Anu Sheetal, and Amarpal Singh"Simulative Analysis of Free Space Optical Transmission System Under different Environment Conditions " CIIT International Journal of programmable device currents and systems, Issue Sept 2009.

[3] A.K Rahman, M.S Anuar, S.A Aljunid and M.N Junita, "Study of rain attenuation consequence in free space optic transmission" Proceedings of IEEE 2008 6th National Conference on Telecommunication Technologies, 26-27 August 2008, Putrajaya, Malaysia..

[4] E.Leitgeb, P.Fasser, M.Gebhart and J.Bregenzer, "Free Space Optics- Extension to fiber networks for the last mile", IEEE Press, Wiley-Intersciences, pp. 495-460, 2002.

[5] Husagic Alma and Wajdi al-Khateeb, "Effect of Weather Conditions on Quality of Free Space Optics Link (with focus on Malaysia)" Proceedings of the international conference on computer and communication engineering 2008, Kuala Lumpur, Malaysia, pp 1206-1210, 2008.

[6] S.Bloom, E. Korevaar, J. Schuster and H. Willebrand, "Understanding the performance of free-space optics", Journal of Optical Networking, vol. 2, no. 6, pp.178-200, June 2003.

[7] Alexander Soibel, Malcolm W. Wright, William H. Farr, Sam A. Keo, Cory J. Hill, Rui Q. Yang, and H. C. Liu, "Midinfrared Interband Cascade Laser for Free Space Optical Communication", IEEE Photonics Technology Letters, vol. 22, no. 2, pp. 121-123, 2010.

[8] Z.Bielecki, W.Kolosowski, J.Mikolajczyk, M.Nowakowski, E.Sedek and J.Wojtas, " Free Space Optical Communication using cascade laser", IEEE, 2008.

Image Steganography Techniques: Survey, Classification and application

Aditi Sharma[1], and Vijaydhir[2]
[1]Lecturer Department of ECE, CT Institutes, Shahpur Campus, Jalandhar, India
[2]AP, Department of ECE, SBBSIET, Padhiana, Jalandhar, India
Email: itsme.adi2710@rediffmail.com, vijaykdhir@yahoo.com

***Abstract* — Steganography is the art and science of hiding some useful data inside some innocent looking canvas. The growth of the World Wide Web has enabled the personal computer to be used as a general communications tool so the requirement to send a message as safely and as securely as possible has been the main concern in communication system. In the last few years, we have seen that new and powerful steganography techniques reported in the literature. This paper intends to give an overview of image steganography, its applications and techniques. The paper gives the description of various techniques used in steganography and attempts to identify the requirements of a good steganographic algorithm.**

***Index Terms* –Least Significant Bits (LSB), Peak Signal-to-Noise Rate (PSNR), Mean Square Error (MSE), Manhattan distance (Mdist), Steganalysis.**

I. Introduction

Steganography means covered or hidden writing i.e., writing that is known to casual observer and is derived from Greek words ‘steganos’ meaning covered or secret and ‘graphy’ meaning writing or drawing. Information is the wealth of any organization therefore security issues are top priority to an organization dealing with confidential data. Steganography is the science that involves communicating secret data in an appropriate multimedia carrier, e.g., image, audio, and video files.

Image Steganography is the technique for hiding information by embedding messages within image. It is widely used in military, diplomatic, personal and intellectual property applications. Steganography is the term applied to any number of processes that will hide a message within an object particularly an image, where the hidden message will not be apparent to an observer. Typically, the message is embedded within another object (image) known as a cover object, by tweaking its properties. The resulting output, known as a stego object or stegogramme is engineered such that it is a near identical perceptual model of the cover object, but it will also contain the hidden message. If anybody intercepts the communication, they will obtain the stegogramme, but as it is so similar to the cover, it is a difficult task for them to tell that the stegogramme is anything but innocent. It is therefore the duty of steganography method to ensure that the adversary regards the stegogramme - and thus, the communication - as innocuous [1, 2].

Steganography differs from cryptography because the latter does not attempt to hide the fact that a message exists. Instead, cryptography merely obscures the integrity of the information so that it does not make sense to anyone but the creator and the recipient. The adversary will be able to see that a message exists, and the inverse process of cryptanalysis involves trying to turn the meaningless information into its original form. Steganography in the modern day sense of the word usually refers to information or a file that has been concealed inside a digital Picture, Video or Audio file. What Steganography essentially does is exploit human perception; human senses are not trained to look for files that have information hidden inside of them. Steganography is employed in various useful applications, e.g., for human rights organizations, as encryption is prohibited in some countries copyright control of materials, enhancing robustness of image search engines and smart identity cards, where details of every person are embedded in their photographs [3].

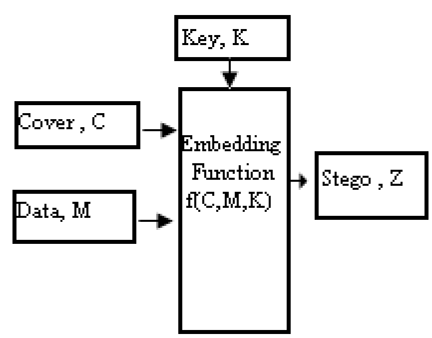


Fig 1: Basic Steganography Model

Other applications are video-audio synchronization, companies’ safe circulation of secret data, TV broadcasting, TCP/IP packets, for instance a unique ID can be embedded into an image to analyze the network traffic of particular users, and also checksum embedding [4]. The basic model for steganography is shown on “Figure 1”. The model shows of steganography consist of Carrier, Message and Password.

Carrier is also known as cover-object, in which message is embedded and serves to hide the presence of the message. The data can be any type of data (plain text, cipher text or other image) that the sender wishes to remain confidential. Password is known as stego-key, which ensures that only recipient who knows the corresponding decoding key will be able to extract the message from a cover-object. The cover-object with the secretly embedded message is then called the stego-object.

II. STEGANOGRAPHY CLASSIFICATION

Almost all digital file formats can be used for steganography, but the formats that are more suitable are those with a high degree of redundancy. Redundancy can be defined as the bits of an object that provide accuracy far greater than necessary for the object’s use and display. The redundant bits of an object are those bits that can be altered without the alteration being detected easily. Image and audio files especially comply with this requirement, while research has also uncovered other file formats that can be used for information hiding.

There are four main categories of file formats that can be used for steganography shown in “Figure 2”. Since, images are quite popular cover or carrier objects used for steganography. In the domain of digital images many different image file formats exist, most of them for specific applications. For these different image file formats, different steganographic algorithms exist. Here, in this paper, we will discuss about the image domain steganography methods. In Image Domain methods secret messages are embedded using the intensity of the pixels values directly. The image domain methods are relatively simple compared to the other methods and are sometimes characterized as the “simple systems”. However, they are generally more sensitive to small changes on the image such as filtering, resizing and squeezing.

To hide information in audio files similar techniques are used as for image files. One different technique unique to audio steganography is masking, which exploits the properties of the human ear to hide information unnoticeably. A faint, but audible, sound becomes inaudible in presence of another louder audible sound. This property creates a channel in which to hide information. Although nearly equal to images in steganography potential, the larger size of meaningful audio files makes them less popular to use than images.

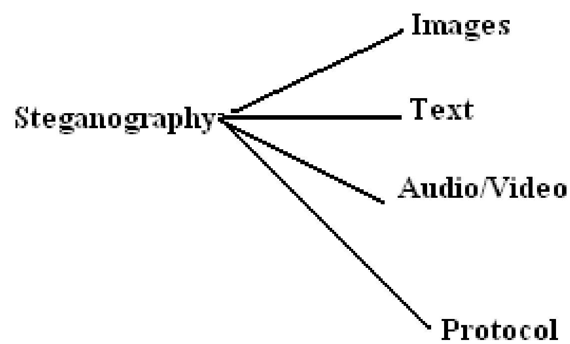


Fig 2: The four main categories of file formats that can be used for steganography

III. VARIOUS IMAGE STEGANOGRAPHY TECHNIQUES

Image steganography techniques can be classified into two broad categories: Spatial-domain based steganography and Transform-domain based steganography.

A. Spatial Domain Method

In spatial domain scheme, the secret messages are embedded directly. Here, the most common and simplest steganography method is the least significant bits (LSB) insertion method. In the LSB technique, the least significant bits of the pixels are replaced by the message bits which are permuted before embedding. Most steganography software hide information by replacing only the least-significant bits (LSB) of an image with bits from the file that is to be hidden. This technique is generally called LSB encoding. One of the most common techniques used in steganography. The following example shows how the letter A can be hidden in the first eight bytes of three pixels in a 24-bit image.

Pixels: (10101111 11101001 10101000)
(10100111 01011000 11101001)
(11011000 10000111 01011001)

Secret message: 01000001

Result: (10101110 11101001 10101000)
(10100110 01011000 11101000)
(11011000 10000111 01011001)

The three bold bits are the only three bits that were actually altered. Since the 8-bit letter A only requires eight bytes to hide it in, the ninth byte of the three pixels can be used to begin hiding the next character of the hidden message. A slight variation of this technique allows for embedding the message in two or more of the least significant bits per byte. This increases the hidden information capacity of the cover-object, but the cover-object is degraded more, and therefore it is more detectable.

Spatial LSB embedding is widely used for its high hiding quality and simplicity to realize. However, the robustness of this method is weak and the message length can be estimated by statistical scheme [5]. In order to solve this problem, some researcher's proposed various methods which are advanced version of LSB techniques. A reversible histogram transformation function-based LSB steganographic method is proposed by Der-Chyuan Lou and Chen-Hao Huto resists statistical steganalysis [6]. The experimental results show that the proposed method resists not only Regular-Singular (RS) attack but also Chi-Square (χ 2) detection methods. Chia-Chun Wu et al. proposed a novel secret image sharing scheme by applying optimal pixel adjustment process to enhance the image quality under different payload capacity and various authentication bits conditions [7]. The experimental result of proposed scheme shows the improvement of image quality of stego images. He also provides several experiments to demonstrate the efficacy of authentication capability of the proposed scheme and therefore maintains the secret image sharing and authentication ability while enhances the image quality.
Xin Liao et. al. improve the embedding capacity and provide an imperceptible visual quality, by give a novel steganographic method based on four-pixel differencing and modified Least Significant Bit (LSB) substitution [8]. The experimental result of proposed method gives not only an acceptable image quality but also provides a large embedding capacity.

As vast channels for communication such as the Internet are becoming popular, the security of digital media becomes a greater concern. The hiding of a message will reduce the probability of detecting this message. This method hides a gray image in one another. The cover is divided into blocks of equal sizes. Each block size equals the size of the embedding image. Edge Based Steganography is in which only the sharper edge regions are used for hiding the message while keeping the other smoother regions as they are. It is more difficult to observe changes at the sharper edges than those in smoother regions. In this method Enhanced Least Significant Bit algorithm is used which can reduce the rate of pixel modification thereby increasing the security both visually and statistically.

Grey Level Modification Steganography Method steganography method is based on image layers. This method divides the host image into blocks and embeds the corresponding secret message bits into each block using the layers which are made by the binary representation of pixel values. It then performs a search on the rows and columns of the layers for finding the most similar row or column. The location of row/column and its differences from the secret message is then marked by modifying minimum number of bits in the least significant bits of the blocks.

B. TRANSFORM DOMAIN METHOD

Here we can embed information in DCT, DFT, FFT domains etc. The main strength offered by transform domain techniques is that they can take advantage of properties of alternate domains to address the limitations of pixel-based methods or to support additional features. A possible disadvantage of spatial techniques is that they are not very robust against attacks. In addition to this, adaptive steganography techniques are a bit more difficult in the spatial domain.

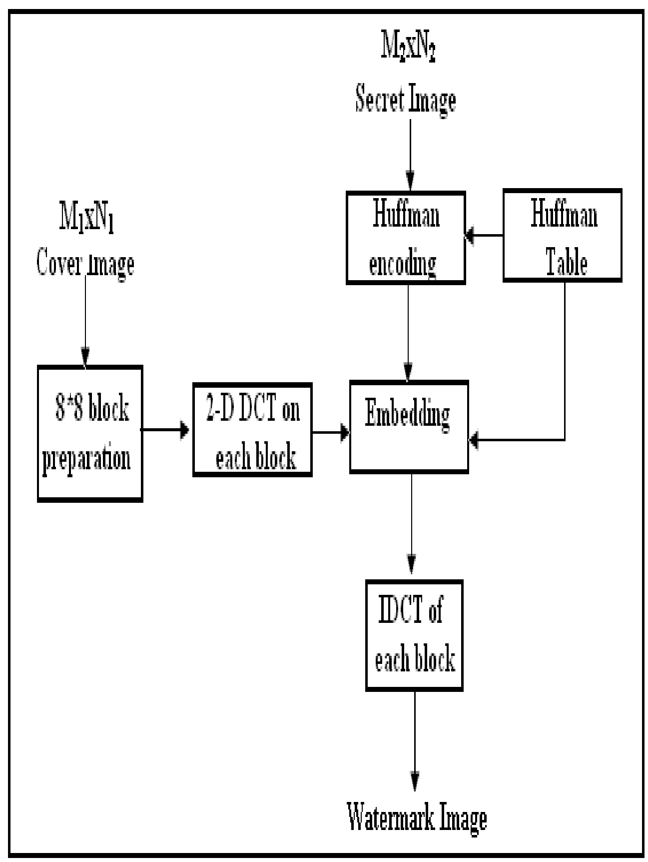


Fig. 3: Insertion of a information/Secret message

Both the robustness and quality of the watermark could be improved if the properties of the cover image could similarly be exploited. For instance, it is generally preferable to hide watermarking information in noisy regions and edges of images, rather than in smoother regions. The benefit is two-fold; Degradation in smoother regions of an image is more noticeable to the HVS, and becomes a prime target for lossy compression schemes.

The idea is to hide information in frequency domain by altering magnitude of all of discrete cosine transform (DCT) coefficients of cover image. The 2-D DCT converts image blocks from spatial domain to frequency domain. The carrier image is divided into non overlapping blocks of size 8×8 and applies DCT on each of blocks of cover image using forward DCT. Now perform Huffman encoding on the 2D secret image of size M2xN2, to convert it into 1D stream.

Huffman code is decomposed in 8-bits blocks. The least significant bit of all of the DCT coefficients inside 8×8 block is changed to a bit taken from each 8 bit block from left to right. Now, perform the inverse block DCT using inverse DCT and obtain a new image which contains secret image. At the receiver side, the stego-image is received, which is in the spatial domain.

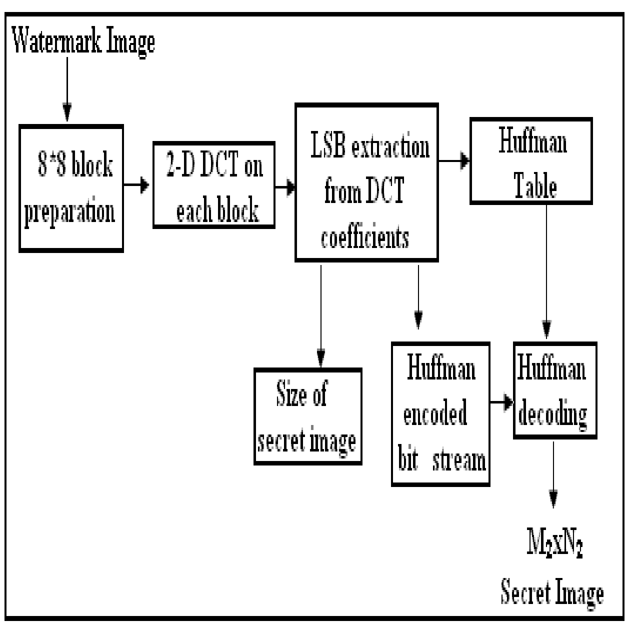


Figure 4: Retrieval of Information/Secret message

IV. CONCLUSIONS

In the given paper, Survey, classification and application of various methods of steganography were discussed. Most of the techniques work on the least significant bits of the pixel values. Most of the existing steganographic methods rely on two factors: the secret key and the robustness of the steganographic algorithm.Without loss of generality; edge embedding maintains an excellent distortion free output whether it is applied in the spatial, DCT or DWT domains. However, the limited payload is its downfall. DWT domain shows promising results and outperforms DCT embedding especially in terms of compression survival. A steganographer should be cautious when embedding in the transformation domains in general; however DWT tends to be more flexible than DCT.

REFERENCES

[1] Abbas Cheddad, Joan Condell, Kevin Curran, Paul McKevitt, "Digital image steganography: Survey and analysis of current methods", Journal of Signal Processing, Elsevier, Vol. 90, 2010, pp: 727–752.

[2] Yambern Jina Chanu, ThemrichonTuithung, Kh. Manglem Singh, "A Short Survey on Image Steganography and Steganalysis Techniques", IEEE 3rd conference on Emerging Trends and Applications in Computer Science (NCETACS), Shillong, 30-31 March 2012, pp: 52 – 55.

[3] Alain Brainos, "A Study of Steganography and The Art of Hiding Information"July,2004, http://www.infosecwriters.com/text_resources/pdf/steganographyDTEC6823.pdf

[4] Paunwala, M.C. Patnaik, S., "Sheltered Identification with Hiding Biometrics", International Conference on Signal and Image Processing (ICSIP), IEEE, Surat, 15-17 Dec., 2010, pp: 191-196.

[5] R. Shreelekshmi, M. Wilscy, C.E. Veni Madhavan, "Cover Image Preprocessing for More Reliable LSB Replacement Steganography", Proc. of International Conference on Signal Acquisition and Processing, IEEE, Trivandrum, 9-10 Feb., 2010; pp: 153-156.

[6] Der-Chyuan Lou and Chen-Hao Hu, "LSB steganographic method based on reversible histogram transformation function for resisting statistical steganalysis", Journal of Information Sciences, Elsevier, Vol. 188, 2012, pp: 346–358.

[7] Chia-Chun Wu, Shang-Juh Kao and Min-Shiang Hwang, "A high quality image sharing with steganography and adaptive authentication scheme", Journal of Systems and Software, Elsevier, Vol. 84, 2011, pp: 2196– 2207.

[8] Xin Liao, Qiao-yan Wen and Jie Zhang, "A Steganographic method for Digital Images with Four-pixel Differencing and Modified LSB Substitution", Journal of Visual Communication and Image Representation, Elsevier, Vol. 22, 2011, pp: 1–8.

Applications of Wireless Optical Communication and Issues

Sanamdeep Singh
M.Tech (ECE), Amritsar College of Engineering and Technology, Amritsar.
singh.sanamdeep@ymail.com

***Abstract-* Wireless optical communication is the future of wireless technology where visible light can be used as communication medium. It uses visible light between 400 THz (780 nm) and 800 THz (375 nm). It provides very high data rate along with high security. This technique consumes less power and also serves the purpose of lightning along with data transfer. The bit rate achieved using this wireless communication cannot be achieved in Radio and Microwave links. Light beams are propagated through air which can carry data at rates up to 500 Mbps over five meter distance. In this paper, knowledge is provided about applications, advantages and disadvantages of this technology in different situations. This paper also introduces its dominance over existing wireless technologies.**

Keywords- LiFi (Light Fidelity), FSO (Free Space Optics)

I. Introduction

High speed, high security, large bandwidth and biologically friendly communication are the main advantages of wireless optical communication. The main device used as transmitter is high brightness LEDs. These have good efficiency, long life, less heat dissipation and are also resistant to humidity. LEDs having good switching capabilities are used so that high rate data can be transmitted effectively. The transmitters and receivers are easy to install, also there is no interference from radio waves. Different problems like limited bandwidth, interference, data security and harmful effects to health can be overcome.

In 2010, frequency modulated white LED used by Siemens and Heinrich Hertz institute in Berlin, to transmit data at 500 Mbps over 5 meters. Also, Harald Haas demonstrated his invention using an ordinary table lamp that successfully transmitted data at speeds exceeding 10Mbps using light waves from LED light bulbs to a computer located below the lamp. The technology uses fluorescent lamps (ordinary lamps, not special communications devices) to transmit signals at 10 kbps, or LEDs for up to 500 Mbps. Li-Fi is the term first used by Harald Haas in his TED Global talk on Visible Light Communication. In October 2011 a number of companies and industries formed the Li-Fi Consortium, to promote high-speed optical wireless systems and to enhance the limited bandwidth provided by radio-based wireless spectrum available. The consortium believes it is possible to achieve more than 10 Gbps speed using this optical wireless technology also known as Li-Fi. The communication is done by deploying transmitter and receiver in direct line of sight manner. Affected if line of sight is not used, it will affect the speed of date transmission. It is also more secure than other wireless networks as only photo receptors are used, which can receive data within transmitted cone of light signal.

II. APPLICATIONS

Optical wireless communication broadly has four major applications. These are:-

- High speed wireless indoor communication (also known as Li-Fi)
- Vehicular visible light communication network.
- Free space optical communication.
- Underwater communication over short distance.

2.1 HIGH SPEED WIRELESS INDOOR COMMUNICATION

Just like Wifi, a new approach has been used to provide wireless communication in homes, offices and buildings to access internet wirelessly. Here LEDs with 100 MHz bandwidth are used as source or medium. Speed up to 500 Mbps can be achieved along with a much wider bandwidth than WLAN in radio links.

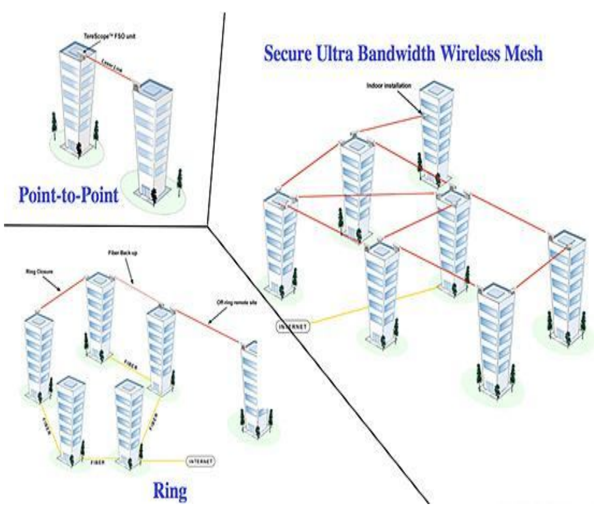
The main issue regarding this application is Line of Sight tracking. The receiver and transmitter must be in direct line of sight. Another problem occurs regarding maximum coverage of a room or office. Line of sight can be solved by using high brightness LEDs having a broad conical beam. This technique can also cover maximum area to provide access to almost every user.

Benefits – high speed and bandwidth, harmless to eyes and health and other electronic devices, no interference from radio waves.

Drawbacks – narrower beam of LEDs can cause line of sight problems, Line of Sight is essential, interference from other light sources is possible.

2.2 FREE SPACE OPTICAL COMMUNICATION

It is short range system used to connect buildings of campuses and industries via laser beams. LAN-to-LAN connections, connections in a city, a metropolitan area network, at Fast Ethernet or Gigabit Ethernet speeds can be made.

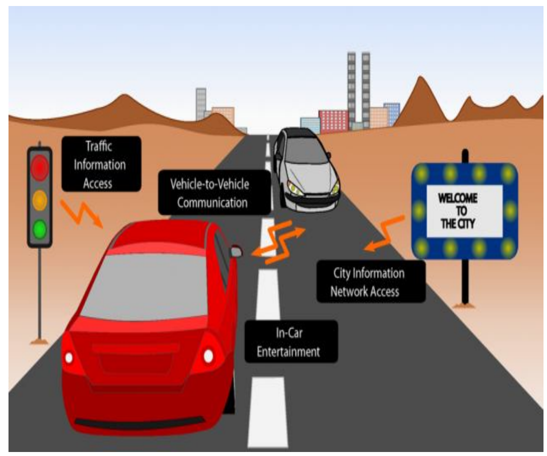


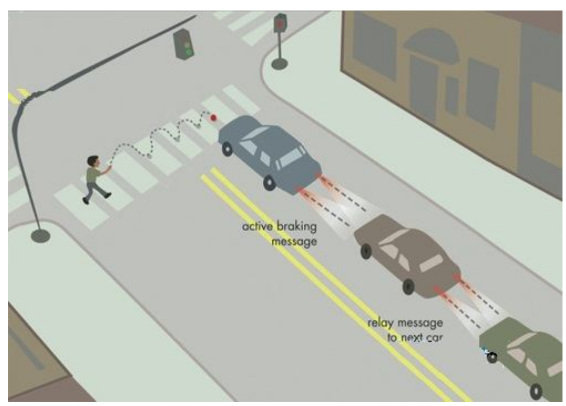
Benefits – easy deployment, low bit error rates, immunity to interference, high security, range 1-2 kms and economical.

Drawbacks - beam dispersion, absorption, rain, fog, snow, background light, shadowing, pointing stability in wind, pollution, if the sun goes exactly behind the transmitter, it can swamp the signal. Due to these factors attenuation of signal can occur and also bit error rate increases.

2.3 VEHICULAR VISIBLE LIGHT COMMUNICATION

This kind of communication can be used in vehicle to vehicle communication, infrastructure to vehicle broadcasting, vehicle to infrastructure communication.





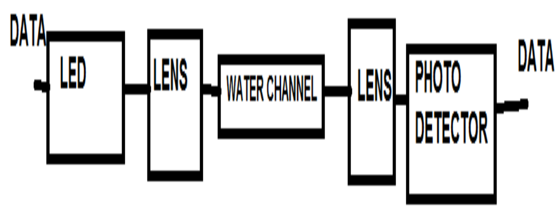
It can be used in traffic signal violation warning, curve speed warning, left turn assistant, lane change warning, forward collision warning, emergency electronic back lights. Its main objective is to provide vehicle safety and also provides internet access applications. Headlights and backlights of vehicle can act as transmitters and receivers can be attached to other parts of vehicle. It can be used in dense traffic roads and can overcome the effects of packet collisions.

Benefits – provides vehicle to vehicle communication, entertainment inside car, connectivity to global internet, access to traffic information.

Drawbacks – noise (day time) from sunlight and (night time) from other light sources, sideways communication not possible.

2.4 UNDERWATER COMMUNICATION

This kind of communication is helpful to divers. A diver can communicate with another one with the help of microphone installed in LED light and its voice can be sent to another diver over light path. Other diver will receive the audio signal using light wave. This will also require line of sight communication. With the help of this technology communication can be done between divers, under water vehicles and ships. The block diagram of system is shown below-



Benefits – high reliability within short range (5-10 meters), economical as LEDs are used, can be used in water-air interface communications, can be used deep inside oceans, 100 meter link can be established and high data rate.

Drawbacks – noise due to ambient light, scattering due to suspended particles, absorption of optical signals in water.

Noise due to ambient light can be overcome by using monochromatic light and other drawbacks can be overcome by using different kind of modulation schemes.

Applications above are demonstrating the efficient use of wireless optical technology, but there are some negative points which have to be considered. The main point of consideration is abundance of bandwidth and very high speed. So, this technology is more advantageous than other wireless technologies in modern era. Certain conditions have to be taken into account to implement this technology, like line of sight, ambient noise, area coverage, efficiency and switching capabilities of LEDs and LASERS. If above conditions are fulfilled; an efficient use of technology can be made. Also the research is being carried on by certain research groups like VLCC (Visible Light Communications Consortium) to overcome the drawbacks.

III. DOMINANCE OVER EXISTING WIRELESS TECHNOLOGIES

Wi-Fi is a technology that uses radio waves to provide wireless communication path with a bit rate of 2Mbps up to a distance range from 32 meters - 95 meters. In contrast to this wireless optical technique provides a speed of 500 Mbps up to a distance of 5 meters and 100Mbps up to 100 meters using LEDs and LASERs. Also these devices consume much less power consumption. Here only source of interference is another light source.

Optical signals cannot cancel each other, so no fading occurs. Data is encoded into light signal by varying the flicker rate of LED. It can be used in areas where Wi-Fi is prohibited like in aircrafts, medical devices and hospitals. It can be used in underwater according to Jamie Condliffe as shown in last application above.

IV. CONCLUSION

In this paper applications of wireless optical communication technology are discussed and their drawbacks and benefits have been studied. It is a rapidly growing field because of its unmatched advantages. It extends the use of wireless technology beyond the radio spectrum range. Research is being carried out to increase the speed beyond 10 Gbps. Due to its highly secure transmission it will have a bright future in defense and military applications also. Despite of some problems like Line of Sight, ambient noise etc., this technology will be boon for wireless communication because of its large bandwidth and higher bit rate. The problems or drawbacks are easy to fix and overcome. Also less hardware is required for its implementation which is also very economical.

Due to its harmless nature it opens a pathway for wireless communication in prohibited areas. All in all, this technology has good scope in communication field, we must carry on research to enhance and implement it.

REFERENCES

[1] Visible Light Communication (VLC) Sustainable Energy-Efficient Wireless Applications Using Light IEEE Communication Magazine, December 2010

[2] Short-Range Underwater Wireless Communication Using Visible Light LEDs

Yousuke Ito, *Nonmember*, Shinichiro Haruyama, Member and Masao Nakagawa, Fellow, IEEE

[3] Visible Light Communication Using OFDM Mostafa Z. Afgani, Harald Haas, Hany Elgala, and Dietmar Knipp School of Engineering and Science International University Bremen 28759 Bremen, Germany

[4] Free-Space Optical communication using visible light HU Guo-yong†, CHEN Chang-ying, CHEN Zhen-qiang

[5] Enabling Vehicular Visible Light Communication (VVLC) Networks Cen Liu Rice University cen.liu@rice.edu

[6] Indoor optical wireless communication by ultraviolet and visible light Kaiyun Cui, Gang Chen, Qunfeng He, Zhengyuan Xu* Department of Electrical Engineering, University of California 900 University Avenue, Riverside, CA 92521 USA

[7] Design and Analysis of High Speed Free Space Optical Communication Link with Different Parameter Jitendra Singh Department of ECE NIT Hamirpur- India, Vinod Kapoor Department of ECE NIT Hamirpur- India

[8] http://en.wikipedia.org/wiki/Li-Fi

[9] http://www.lificonsortium.org/

[10]http://en.wikipedia.org/wiki/Visiblelight communication

Quality Measures of Color Image Compression under Various Algorithms using DB Wavelet Transform

Varun Sood[1], Anuradha Sood[2]
[1]Assistant Professor, Electronics and Communication Department, Sai College of Engg and Technology, Manawala,
[2]PG Scholar in Master of Computer Application, Amritsar College of Engg and Technology, Manawala.
Email: aryanvarunsood@gmail.com, sood458@gmail.com

***Abstract*— Images require essential storage and transmission resources, thus image compression is advantageous to reduce these requirements. This paper covers some background of wavelet analysis, data compression and how wavelets have been and can be used for image compression. The paper examines a set of wavelet functions (wavelets) for implementation in a still image compression system and discusses important features of wavelet transform in compression of still images, including the extent to which the quality of image is degraded by the process of wavelet compression and decompression. The effects of different wavelet functions, image contents and compression ratios are calculated. On the basis of this we recommended the best compression algorithm for an image for the storage and transmission.**

Keywords— **PSNR, MSE, CR, BPP.**

I.INTRODUCTION

Image compression is an important field of research that has been studied for nearly three decades now. Image Compression addresses the problem of reducing the amount of data required to represent the digital image. Compression is achieved by the removal of one or more of three basic data redundancies: (a) Coding redundancy, which is present when less than optimal (i.e. the smallest length) code words are used; (b) Inter-pixel redundancy, which results from correlations between the pixels of an image & (c) psycho visual redundancy which is due to data that is ignored by the human visual system (i.e. visually nonessential information).Demand for communication of multimedia data through the telecommunications network and accessing the multimedia data through Internet is growing explosively. Another important application is browsing, where the focus is on getting high compression.

There are two types of image compression: lossless and lossy. With lossless compression, the original image is recovered exactly after decompression. Unfortunately, with images of natural scenes it is rarely possible to obtain error free compression at a rate beyond 2:1. Much higher compression ratios can be obtained if some error, which is usually difficult to perceive, is allowed between the decompressed image and the original image [1]. This is lossy compression. In many cases, it is not necessary or even desirable that there be error-free reproduction of the original image. For example, if some noise is present, then the error due to that noise will usually be significantly reduced via some de-noising method. In such a case, the small amount of error introduced by lossy compression may be acceptable. Another application where lossy compression is acceptable is in fast transmission of still images over the Internet. Unlike lossless compression, lossy compression reduces image quality. You can't get the original image back after using lossy compression methods. You will lose some information [2]. Lossless image compression is usually used in artificial images that contain sharp-edged lines such as technical drawings, textual graphics, comics, maps or logos. This is because lossy compression methods produce compression artifacts to images and sharp-edged lines become fuzzy especially when using strong compression. Instead, lossy compression is a good choice for natural images such as photos of landscapes where minor loss on sharpness is acceptable to achieve smaller file size. With the naked eye it is very hard to see any differences between uncompressed natural image and one with compressed by lossy methods if the compression is not too strong [3]. The most widely used methods of lossless compression in images are run-length encoding (RLE), entropy coding and dictionary coders. Digital images are composed of pixels that represent colour information. When a pixel differs only slightly from its neighbors, its value can be replaced theirs. This will lose some information but it is usually barely noticeable with human eye if the algorithm is good enough. After this e.g. RLE or Huffman coding can be used to compress data.

II.IMAGE COMPRESSION USING WAVELET TRANSFORM

Wavelets are functions defined over a finite interval and having an average value of zero. The basic idea of the wavelet transform is to represent any arbitrary function (t) as a superposition of a set of such wavelets or basis functions. These basis functions or baby wavelets are obtained from a single prototype wavelet called the mother wavelet, by dilations or contractions (scaling) and translations (shifts).

There are two types of image compression: lossless and lossy. With lossless compression, the original image is recovered exactly after decompression. Unfortunately, with images of natural scenes it is rarely possible to obtain error-free compression at a rate beyond 2:1. Much higher compression ratios can be obtained if some error, which is usually difficult to perceive, is allowed between the decompressed image and the original image. This is lossy compression. In many cases, it is not necessary or even desirable that there be error-free reproduction of the original image. We shall concentrate on wavelet-based lossy compression of grey-level still images. When there are 256 levels of possible intensity for each pixel, then we shall call these images 8 bpp (bits per pixel) images. Images with 4096 grey levels are referred to as 12 bpp.

The methods of lossy compression that we shall concentrate on are the following: the EZW algorithm, the SPIHT algorithm, the WDR algorithm, and the ASWDR algorithm. These are relatively recent algorithms which achieve some of the lowest errors per compression rate and highest perceptual quality yet reported. After describing these algorithms in detail, we shall list some of the other algorithms that are available. Before we examine the algorithms listed above, we shall outline the basic steps that are common to all wavelet-based image compression algorithms. The five stages of compression and decompression are shown in Figs. 1 and 2.

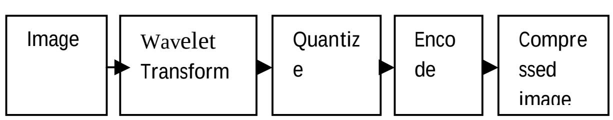


Figure 1: Compression of an image

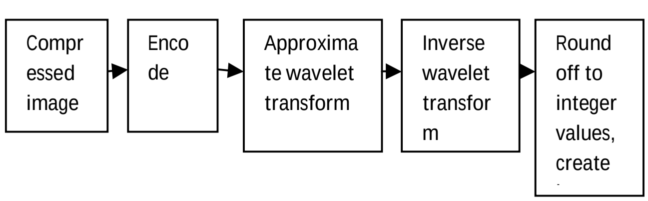


Figure 2: Decompression of an image

All of the steps shown in the compression diagram are invertable, hence lossless, except for the Quantize step. Quantizing refers to a reduction of the precision of the floating point values of the wavelet transform, which are typically either 32-bit or 64-bit floating point numbers. To use less bits in the compressed transform which is necessary if compression of 8 bpp or 12 bpp images is to be achieved these transform values must be expressed with less bits for each value. This leads to rounding error. These approximate, quantized, wavelet transforms will produce approximations to the images when an inverse transform is performed. Thus creating the error inherent in lossy compression. The relationship between the Quantize and the Encode steps, shown in Fig. 1, is the crucial aspect of wavelet transform compression. Each of the algorithms described below takes a different approach to this relationship. The purpose served by the Wavelet Transform is that it produces a large number of values having zeroed, or near zero, magnitudes. Two commonly used measures for quantifying the error between images are Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR). The MSE between two images f and g is defined by

$$MSE = \frac{1}{N}\sum_{j\,,\,k}(f[j\,,k] - g[j,k])^2$$

where the sum over j; k denotes the sum over all pixels in the images, and N is the number of pixels in each image. The PSNR between two (8 bpp) images is, in decibels,

$$PSNR = 10\log_{10}\left(\frac{255^2}{MSE}\right)$$

1. EZW algorithm

The EZW algorithm was one of the first algorithms to show the full power of wavelet based image compression. It was introduced in the groundbreaking paper of Shapiro [7]. Many algorithms build upon the fundamental concepts that were first introduced with EZW. EZW stands for Embedded Zerotree Wavelet. We shall explain the terms Embedded, and Zerotree, and how they relate to Wavelet-based compression. An embedded coding is a process of encoding the transform magnitudes that allows for progressive transmission of the compressed image. Zerotrees are a concept that allows for a concise encoding of the positions of significant values that result during the embedded coding process. The embedding process used by EZW is called bit-plane encoding. EZW stands for Embedded Zerotree Wavelet. We shall explain the terms Embedded, and Zerotree, and how they relate to Wavelet-based compression. An embedded coding is a process of encoding the transform magnitudes that allows for progressive transmission of the compressed image. Zerotrees are a concept that allows for a concise encoding of the positions of significant values that result during the embedded coding process. We shall first discuss embedded coding, and then examine the notion of zerotrees.

2) Set Partitioning in Hierarchical Trees (SPIHT encoding)

The SPIHT [8-9] image coding algorithm was developed in 1996 by Said and Pearlman and is another more efficient implementation of the embedded zerotree wavelet algorithm by Shapiro.
Some of the best results-highest PSNR values for given compression ratios for a wide variety of images have been obtained with SPIHT. Consequently, it is probably the most widely used wavelet-based algorithm for image compression, providing a basic standard of comparison for all subsequent algorithms. SPIHT stands for Set Partitioning in Hierarchical Trees. The term Hierarchical Trees refers to the quadtrees that we defined in our discussion of EZW. Set Partitioning refers to the way these quadtrees divide up, partition, the wavelet transform values at a given threshold. By a careful analysis of this partitioning of transform values, Said and Pearlman were able to greatly improve the EZW algorithm, significantly increasing its compressive power. Our discussion of SPIHT will consist of three parts. First, we shall describe a modified version of the algorithm introduced in [5]. We shall refer to it as the Spatial orientation Tree Wavelet (STW) algorithm. STW is essentially the SPIHT algorithm, the only difference is that SPIHT is slightly more careful in its organization of coding output. Second, we shall describe the SPIHT algorithm. It will be easier to explain SPIHT using the concepts underlying STW. Third, we shall see how well SPIHT compresses images.

3. WDR Algorithm

One of the defects of SPIHT is that it only implicitly locates the position of significant coefficients. This makes it difficult to perform operations, such as region selection on compressed data, which depend on the exact position of significant transform values. By region selection, also known as region of interest (ROI), we mean selecting a portion of a compressed image which requires increased resolution. This can occur, for example, with a portion of a low resolution medical image that has been sent at a low bpp rate in order to arrive quickly. Such compressed data operations are possible with the Wavelet Difference Reduction (WDR) algorithm of Tian and Wells. The term difference reduction refers to the way in which WDR encodes the locations of significant wavelet transform values, which we shall describe below. Although WDR will not typically produce higher PSNR values than SPIHT, we shall see that WDR can produce perceptually superior images, especially at high compression ratios. The only difference between WDR and the Bit-plane encoding described above is in the significance pass. In WDR, the output from the significance pass consists of the signs of significant values along with sequences of bits which concisely describe the precise locations of significant values.

4. ASWDR algorithm

One of the most recent image compression algorithms is the Adaptively Scanned Wavelet Difference Reduction (ASWDR) algorithm of Walker. The adjective adaptively scanned refers to the fact that this algorithm modifies the scanning order used by WDR in order to achieve better performance. ASWDR adapts the scanning order so as to predict locations of new significant values. If a prediction is correct, then the output specifying that location will just be the sign of the new significant value the reduced binary expansion of the number of steps will be empty. Therefore a good prediction scheme will significantly reduce the coding output of WDR.

III.RESULTS

Compressing colour images efficiently are one of the main problems in multimedia applications. So we have tested the efficiency of colour image compression using EZW, SPIHT, WDR and ASWDR algorithm. Reconstructed image is verified using human vision and PSNR.

1. QUALITY ASSESSMENT FOR DB1

	MSE	PSNR	CR	BPP
EZW	10.1191	38.0794	18.0445	4.3307
SPIHT	12.5269	37.1524	12.2345	2.9363
WDR	10.1191	38.0794	20.4188	4.9005
ASWDR	10.1191	38.0794	19.5653	4.6957

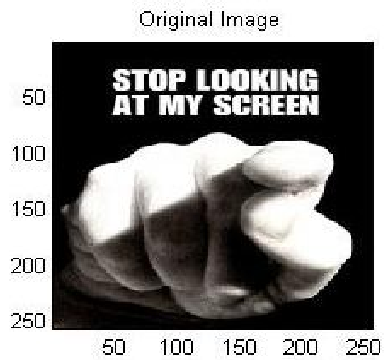


Fig1.1 Original 256 x 256 image

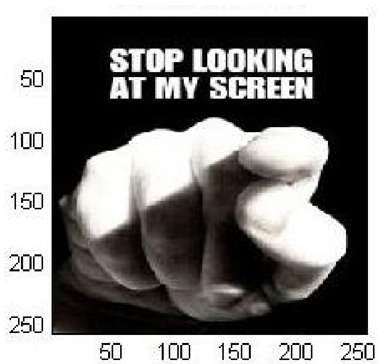


Fig 1.2Compressed image using ASWDR

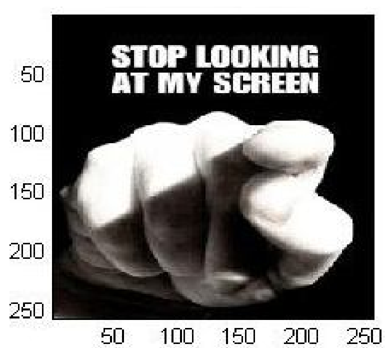


Fig1.3 Compressed image using EZW

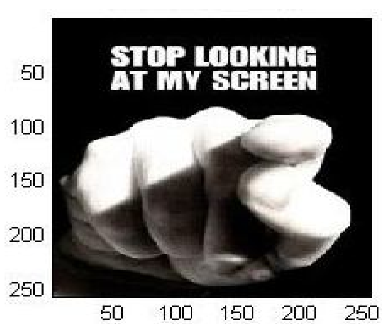


Fig1.4 Compressed image using SPIHT

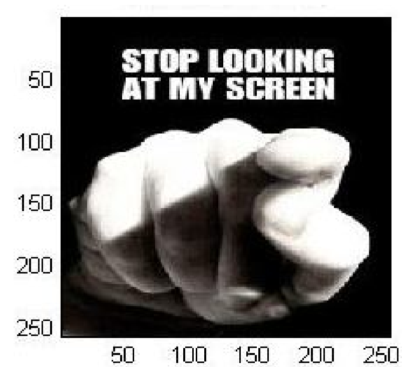


Fig1.5 Compressed image using WDR

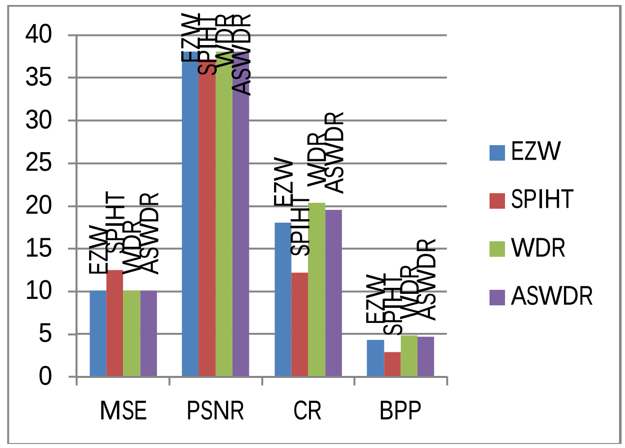


Fig1.6 Graphical Representation of DB1 Wavelet

2. QUALITY ASSESSMENT FOR DB2

	MSE	PSNR	CR	BPP
EZW	**9.8623**	**38.1910**	**17.5440**	**4.2106**
SPIHT	**12.5252**	**37.1530**	**11.4502**	**2.7480**
WDR	**9.8623**	**38.1910**	**19.4768**	**4.6744**
ASWDR	**9.8623**	**38.1910**	**18.7439**	**4.4985**

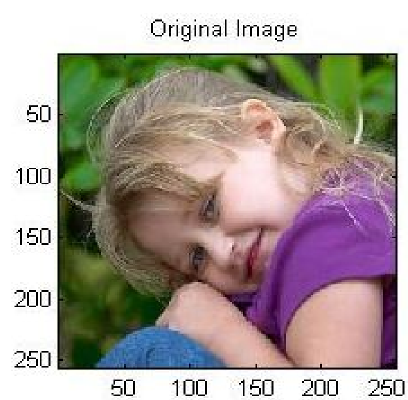


Fig2.1 Original 256 x 256 image

Fig2.2Compressed image using ASWDR

Fig2.3 Compressed image using EZW

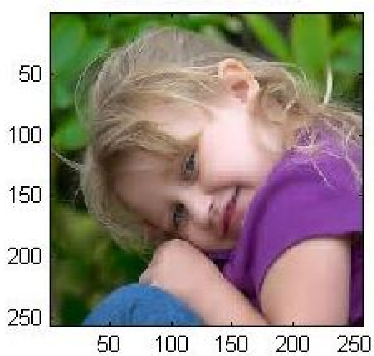


Fig2.4 Compressed image using SPIHT

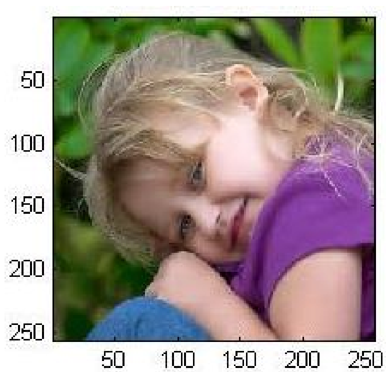


Fig2.5 Compressed image using WDR

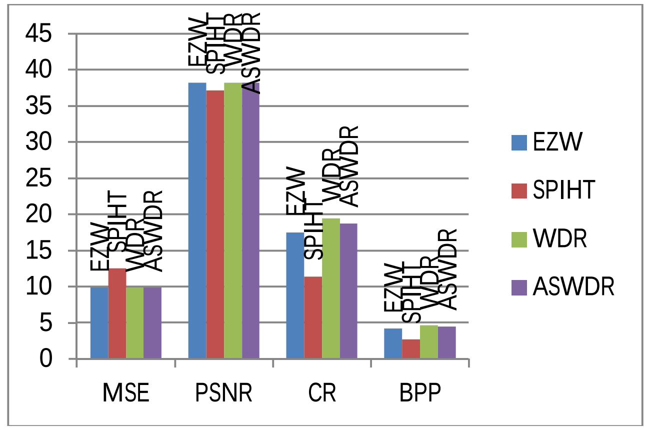


Fig2.6 Graphical Representation of DB2 Wavelet

3. QUALITY ASSESSMENT FOR DB7

	MSE	PSNR	CR	BPP
EZW	**10.1840**	**38.0516**	**18.1396**	**4.3535**
SPIHT	**13.0109**	**36.9877**	**11.4090**	**2.7382**
WDR	**10.1840**	**38.0516**	**20.1324**	**4.8318**
ASWDR	**10.1840**	**38.0516**	**19.7184**	**4.7324**

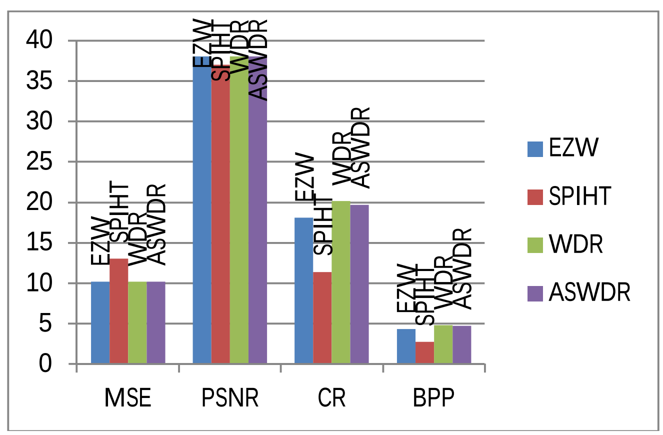


Fig3.6 Graphical Representation of DB7 Wavelet

IV.CONCLUSIONS

We have reviewed and summarized the characteristics of image compression, need of compression and its principles and EZW and SPIHT image compression algorithms based on Wavelet. We use 256×256 color image for comparison. Any of the two approaches is satisfactory when the 0.5 bits per pixel (bpp) is requested.

In the DB1 Wavelet technique, EZW is the best approach as it gives better quality of images(PSNR) and least mean square error(MSE) whereas the WDR gives the better compression ratio(CR).

In the DB2 Wavelet technique, EZW and WDR both give the better quality of images and least mean square error (MSE) whereas only WDR gives the better compression ratio (CR).

In the DB7 Wavelet technique, EZW is the best approach as it gives better quality of images(PSNR) and least mean square error(MSE) whereas the WDR gives the better compression ratio(CR).

However if For practical applications, we conclude that (1) Wavelet based compression algorithms are strongly recommended, (2) DCT based approach might use an adaptive quantization table, (3) VQ approach is not appropriate for a low bit rate compression although it is simple, (4) Fractal approach should utilize its resolution-free decoding property for a low bit rate compression.

REFERENCES

[1] L. R. Iyer and A. E. Bell. "Improving image compression performance with balanced multiwavelets." Proc. of 35th Asilomar Conference on Signals, Systems and Computers, November 2001, (Invited Paper).

[2] Majid Rabbani, Paul W.Jones*;* "Digital Image Compression Techniques" Edition-4, 1991 page 51.

[3] Steven L. Eddins Roberto H. Bamberger and Veyis Nuri. "Genralized symmetric extension for size-limited multirate filter banks.", IEEE Trans. on Image Proc., 3(1):82-87, January 1994.

[4] S. Rout and A. E. Bell. "Color image compression: multiwavelets vs. scalar wavelets." Proceedings of International Conference on Acoustics, Speech, and Signal Processing, 4:3501{3504, 2002}.

[5] Subramanya A, "Image Compression Technique,"Potentials IEEE, Vol. 20, Issue 1, pp 19-23, Feb-March 2001,

[6] David Jeff Jackson & Sidney Joel Hannah, "Comparative Analysis of image Compression Techniques," System Theory 1993, Proceedings SSST '93, 25th Southeastern Symposium, pp 513-517, 7 –9 March 1993.

[7] Sonal, Dinesh Kumar," A study of various compression techniques", National Conference on Challenges & Opportunities in Information Technology, 2007.

[8] Dr. A. E. Bell, Chair, Dr. A. L. Abbott, Dr. B. D. Woerner, Survey of Lossless Image Coding Techniques, SPIE Proceedings on Digital Image Processing Applications, pp. 92-100, 1989.

[9] *Andreas E. Savakis* ,"Evaluation of lossless compression for gray scale document images.", Journal Of Electronic Imaging, 2002.

Coordinated Multi-Point (CoMP) in the Cellular System

Satwinder Singh
M.Tech ECE, Amritsar College of Engineering and Technology
Email :- satwindersingh92gmail.com

***Abstract*-This paper provides a overview of CoMP which is currently considered for the evolution of LTE, referred to as LTE-Advanced. Theoretical research on coordinated multi-point (CoMP) in the cellular uplink claims large improvements in spectral efficiency and fairness. However, the real-world implementation of CoMP is linked with major challenges such as multi-cell synchronization and multi-cell channel estimation, which have to be addressed to make sure that CoMP finds its way into next generation cellular systems (e.g. LTE-Advanced). CoMP concepts do in fact yield significant spectral efficiency gains in an outdoor deployment of two cooperating base stations and two terminals. CoMP Increases 3G LTE Data Rates.**

Keywords: CoMP, LTE, MIMO, ICI, CSI, RRH, UE.

I. INTRODUCTION

The increasing demand for higher transmission rates in cellular mobile communication systems requires that spectrum is used as efficiently as possible, which requires that radio resources are reused in each cell. The occurring inter-cell interference, however, is not sufficiently addressed in LTE Release 8 [1], which leads to a strong performance degradation of cell-edge users.

CoMP transmission/reception

CoMP is a technology which sends and receives signals from multiple sectors or cells to a given UE. By coordinating transmission among multiple cells, interference from other cells can be reduced and the power of the desired signal can be increased.

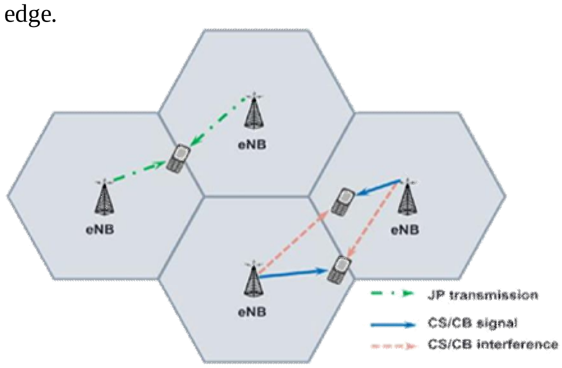
Carrier aggregation and CoMP are the two most important techniques that boost the data rate of the LTE-Advanced to a new threshold. If we call CA a road of the LTE-Advanced, CoMP surely will be a car which the LTE-Advanced drives. In traditional telecommunication systems, each UE will be basically served by only one base station (BS) at a moment. Signals come from other BS's will become interference to the UE. When the UE moves to the cell edge, it will communicate with more than one BS's to prepare for handover. However, it is still being served by its original BS. This is also the time when the UE receives strong interference, and data rate will be very low. The situation will become worse if the UE is moving with high speed.

Coordinated multipoint can be considered as a distributed MIMO system, in that geographically distributed nodes form multiple antennas and they cooperate to transmit to and/or receive from UE's. CoMP has been studied as a solution for increasing the system throughput, especially at cell edge areas where inter-cell interference is severe with traditional approach. Due to the potential advantage, CoMP techniques received a lot of attention at the initiatory stage of the LTE-Advanced standardization. However, in practice, there are critical issues in CoMP, such as excessive feedback overhead, backhaul delay and burden and interference channel estimation. Accordingly, the discussion on CoMP was suspended in release 10, but it is being discussed again in release 11.

Coordination

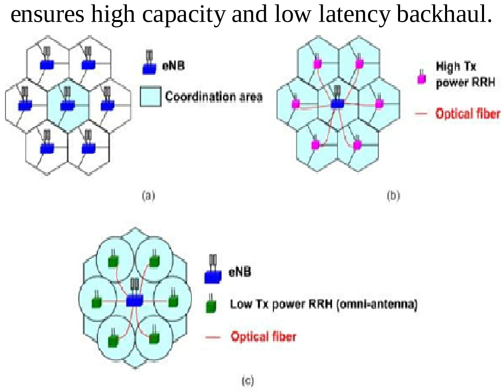
In a conventional cellular system, the BS is located in the cell center and it only serves the users in its coverage area. The signals transmitted from other BSs cause interference, especially at the cell-edge, where different coverage areas overlap, and giving rise to InterCell Interference (ICI) which reduces the spectral efficiency of the cell. When the channel state information (CSI) of various links are made available to an entity (central unit) then the interference from other cells can be avoided by designing a beamformer. This pre-canceling of interference by beamforming and power allocation is called Precoding. When a transmission to a user is collaborated by multiple BSs or network points, acting together to remove interference, this is referred to as CoMP transmission. For this to occur, the CSI from all the BSs needs to be available at the central unit for precoding. This constitutes the centralized joint processing algorithm, where a set of BSs form a cluster of cooperative cells. But, coordinating BSs for coherent joint processing puts tremendous requirements for high speed backhauling (10 Gbps over fiber or up to 4 Gbps over microwave links) for the CSI to be available at the central unit. Hence, various joint processing schemes are developed to reduce the burden on backhauling. The partial joint processing algorithm is one such scheme, where only a subset of BSs are allowed to transmit based on a threshold. Thus, reducing the load on backhauling.Contrary to the centralized joint processing, the precoding can be done locally at each BS, which gives rise to distributed joint processing.

Coordinated multipoint can be applied to both the DL and UL. DL CoMP techniques can be classified according to the amount of information shared among cells. Joint processing is available when neighboring cells share transmit data as well as the channel state information. The joint processing can be realized in the form of joint transmission or dynamic cell selection. In joint transmission, cooperating eNB's **jointly transmit** data to one or more corresponding UE's. Dynamic cell selection is a kind of fast cell selection; UE's are handed over to the best cell considering interference situation. However, joint processing generally requires high-capacity X2 interface between eNB's for sharing transmit data, and thus can cause excessive backhaul overhead and latency. On the other hand, *Coordinatedscheduling/coordinatedbeamforming (CS/CB)* can be realized only if the channel stateinformation and scheduling information are shared among eNB's; data sharing is not required. In the CS/CB, a UE receives data from only one eNB, its own serving node, while the precoding and scheduling are coordinated among related eNB's in such a way to reduce interference and improve the throughput. Figure 1 illustrates how the joint processing and CS/CB serve UE's at the cell edge.



Joint processing and CS/CB
Figure 1

For the case of UL, joint detection and interference prediction are considered. Joint detection can be considered as a UL counterpart of the DL joint transmission. For joint detection, eNB's need to share received signal samples as well as channel state information and scheduling information. The basic principle of interference prediction is to perform link adaptation based on predicted SINR values. Interference prediction is possible by exchanging resource allocation information among cells. Another emerging UL CoMP scheme of importance is interference-aware distributed precoding, which can be implemented in fully distributed manner without sharing even channel state information among eNB's. According to R1-110564 in 3GPP, CoMP techniques can be applied in three different scenarios, as illustrated in Figure 2. Currently, various CoMP schemes are being evaluated by several institutes under the scenarios. The scenarios of particular interest are the two scenarios with remote radio head (RRH), which ensures high capacity and low latency backhaul.



Different Scenarios of CoMP
Figure 2

By coordinating and combining signals from multiple antennas, CoMP, will make it possible for mobile users to enjoy consistent performance and quality when they access and share videos, photos and other high-bandwidth services whether they are close to the center of an LTE cell or at its outer edges.

The main idea of CoMP is as follows: when a UE is in the cell-edge region, it may be able to receive signals from multiple cell sites and the UE’s transmission may be received at multiple cell sites regardless of the system load. Given that, if the signaling transmitted from the multiple cell sites is coordinated, the DL performance can be increased significantly. This coordination can be simple as in the techniques that focus on interference avoidance or more complex as in the case where the same data is transmitted from multiple cell sites. Forthe UL, since the signal can be received by multiple cell sites, if the scheduling is coordinated from the different cell sites, the system can take advantage of this multiple reception to significantly improve the link performance.

It is well-known that an information exchange among base stations for the purpose of coordinated multi-point (CoMP) detection or transmission allows exploiting inter-cell signal propagation rather than treating it as a curse, yielding large spectral efficiency and fairness gains. However, the benefits of CoMP come at a high cost in terms of complexity and additional infrastructure required. Some important technical challenges are the synchronization in time and frequency of all cooperating entities, the estimation of the CoMP channel, as well as backhaul-efficient multi-cell signal processing.

Transmissions between mobile devices and base stations during the field tests made use of the 2.6 GHz frequency band, which is expected to be the predominant band for introduction of commercial LTE services in Europe. Signals transmitted from mobile devices were received by two active remote radio heads deployed on two buildings located 500m from one another, then forwarded across an optical fiber link to a central unit comprising the modem and controller elements of an Alcatel-Lucent LTE base station (eNode B). The signals were then combined with one another to increase the strength of the signal.

The configuration of this solution differs from that of basic MIMO primarily in the deployment and positioning of antennas. In MIMO, antennas involved in the solution are deployed on a single site. CoMP interconnects antennas deployed at a number of sites that are in proximity to one another. Tight coordination of the transmission and reception of signals at these multiple access points reduces interference and increases efficiency.

Coordinated Multi-point Transmission/Reception:

The implementation of intracell/inter-cell orthogonalization on the uplink and downlink in LTE Rel. 8 contributed to meeting the requirements of capacity and cell-edge user throughput. On the downlink, simultaneously connected UE are orthogonalized in the frequency domain. On the other hand, they are orthogonalized on the uplink, in the frequency domain as well as the code domain, using cyclic shift and block spreading. It is possible to apply fractional frequency reuse (A control method which assigns different frequency ranges for cell-edge UE) to control interference between cells semi-statically, but this is done based on randomization in LTE Rel. 8. Because of this, we are planning to study CoMP technology, which performs signal processing for coordinated transmission and reception by multiple cells to one or more UE, as a technology for Rel. 11 and later in order to extend the intracell/ inter-cell orthogonalization in LTE Rel. 8 to operate between cells.

Independent eNode B and Remote Base Station Configurations:

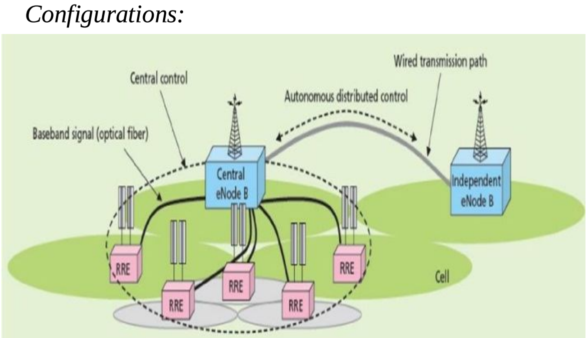
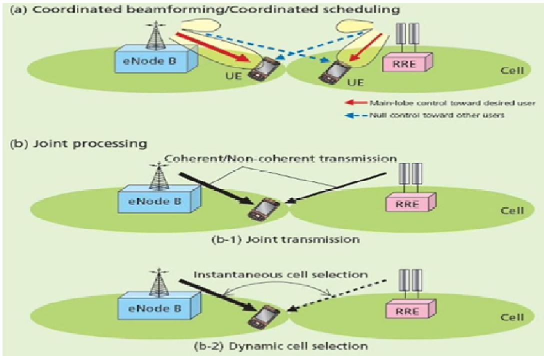


Figure 3

There are two ways to implement CoMP technology: autonomous distributed control based on an independent eNode B configuration, or centralized control based on Remote Radio Equipment (RRE) (Figure 3). With an independent eNode B configuration, signaling over wired transmission paths is used between eNode B to coordinate among cells. Signaling over wired transmission paths can be done with a regular cell configuration, but signaling delay and overhead become issues, and ways to increase signaling speed or perform high-speed signaling via UE need study. With RRE configurations, multiple RREs are connected via an optical fiber carrying a baseband signal between cells and the central eNode B, which performs the baseband signal processing and control, so the radio resources between the cells can be controlled at the central eNode B. In other words, signaling delay and overhead between eNode B, which are issues in independent eNode B configurations, are small in this case, and control of high speed radio resources between cells is relatively easy. However, high capacity optical fiber is required, and as the number of RRE increases, the processing load on the central eNode B increases, so there are limits on how this can be applied. For these reasons, it is important to use both distributed control based on independent eNode B configurations and centralized control based on RRE configurations as appropriate, and both are being studied in preparation for LTE-Advanced.

Downlink Coordinated Multi-point Transmission:

Downlink coordinated multi-point transmission can be divided into two categories: Coordinated Scheduling/ Coordinated Beamforming (CS/CB), and joint processing (Figure 4). With CS/CB, a given subframe is transmitted from one cell to a given UE, as shown in Fig. 4 (a), and coordinated beamforming and scheduling is done between cells to reduce the interference caused to other cells. On the other hand, for joint processing, as shown in Fig. 4 (b-1) and (b-2), joint transmission by multiple cells to a given UE, in which they transmit at the same time using the same time and frequency radio resources, and dynamic cell selection, in which cells can be selected at any time in consideration of interference, are being studied. For joint transmission, two methods are being studied: non-coherent transmission, which uses soft-combining reception of the OFDM signal; and coherent transmission, which does precoding between cells and uses in-phase combining at the receiver.



Downlink Coordinated Multi-point Transmission
Figure 4

Uplink Multi-cell Reception:

With uplink multi-cell reception, the signal from a UE is received by multiple cells and combined. In contrast to the downlink, the UE does not need to be aware of whether multi-cell reception is occurring, so it should have little impact on the radio interface specifications.

Coordinated Multipoint Transmission's key benefits:

Although LTE Advanced CoMP, Coordinated Multipoint is a complex set of techniques, it brings many advantages to the user as well as the network operator.

Makes better utilization of network: By providing connections to several base stations atonce, using CoMP, data can be passed through least loaded base stations for better resource utilization.

Provides enhanced reception performance: Using several cell sites for each connectionmeans that overall reception will be improved and the number of dropped calls should be reduced.

Multiple site reception increases received power: The joint reception from multiple basestations or sites using LTE Coordinated Multipoint techniques enables the overall received power at the handset to be increased.

Interference reduction: By using specialized combining techniques it is possible to utilizethe interference constructively rather than destructively, thereby reducing interference levels. *Helps improve bandwidth* scalability by boosting transmission rates not only in theconnection from the network to the user's mobile device (downlink), but from the mobile device to the network (uplink), a unique function that will become indispensable as Web 2.0 applications become increasingly prevalent and a growing number of users send videos and photos from their mobile devices.

Improves quality of service by demonstrating consistently high transmission rates on theuplink from the phone to the network, even at the edges of a 'cell' where transmission quality is typically poor and difficult to maintain; data rates greater than 5Mbps were observed for the vast majority of locations.

CONCLUSION

Coordinated Multipoint Transmission/Reception, an evolution of the inter-cell interference coordination, is yet another technology to enhance the performance of the LTE-Advanced systems. CoMP is an appealing tool considered in the 3GPP LTE Advanced for mitigating interference and boosting spectral efficiency. CoMP can maximizes the use of existing network infrastructure to achieve these higher transmission speeds without necessarily requiring deployment of additional antennas. Usage of CoMP in conjunction with relaying, showing that both technologies can complement each other well in providing efficient and ubiquitous mobile broadband access. Presently, this framework has only support for single antenna MSs. This can be extended to multiple antennas, as future mobile devices are bound to have more than one antenna. Enhanced coverage and higher cell edge data rates requirements of 4G systems could be met by the relaying technology. LTE release 10 will support this technology component with more enhancements being incorporated in latter releases.

REFERENCES

[1].http://www.cambridge.org/gb/knowledge/isbn/item6488965/?site_locale=en_GB

[2].Coordinated Multi-Point in Mobile Communications, From Theory to Practice By: Patrick Marsch,

[3].http://www.ijcaonline.org/volume13/number5/pxc3872449.pdf

[4].http://3g4g.blogspot.in/2010/02/coordinated-multi-point-comp.html

[5].https://www.academypublisher.com/~academz3/ojs/index.php/jcm/article/view/040314615 4/49

[6].http://www1.i2r.a-star.edu.sg/~wlho/Pubs/Mc_VTC.pdf

[7]. http://www.3gpp.org/ftp/workshop/2009-12-17_ITU-R_IMT-Adv_eval/docs/pdf/REV-090004%20Radio%20layer%202%20and%20RRC%20aspects.pdf

[8].http://www.isrn.com/journals/cn/2012/682090/

[9].http://publications.lib.chalmers.se/records/fulltext/125971.pdf

Modulation Formats For Optical Networks –A Review

Rupanjali Banerjee[1], Er. Manjit Singh[2], Er.Kuldeep Sharma[3]
RIET Phagwara, GNDU (RC), Jalandhar, RIET Phagwara
rupanjan10@gmail.com,manu_kml@yahoo.co.in,sharma02.87

Abstract — These **With the increasing growth and demand for capacity in national, regional and even metropolitan optical networks, high bit rate fiber transmission has recently become an essential part of communications. The high bit rate transmission improves spectral utilization which results in increased overall system capacity and reduces overall cost. Nonlinearity, dispersion, attenuation are some of the major factors which degrades the system's performance. The modulation format plays a very important role in the overall system performance in all communication technologies. Thus in this paper the different modulation formats used in high speed data rate light wave systems has been explained.**

***Index Terms*—RZ, NRZ, Modulation Formats**

I. Introduction

In this paper the methods for the optical signal generation are introduced. The focus is set on modulation formats employing the amplitude modulation of the optical carrier, because of their importance in today's optical transmission systems. The generation and transmission characteristics of conventional and novel modulation formats are presented. The optical signal used for the optical communication network can be generated with different modulation techniques. All modulation formats can be divided into three groups: NRZ-based, RZ-based and novel modulation formats. In the paper, the NRZ modulation formats are introduced. Non-return-to zero (NRZ) and duo binary modulation belong to this group. [1]

II. NRZ BASED MODULATION FORMATS

A basic classification of the various ASK-based modulation formats can be made according to the shape of the optical pulses. All modulation formats can be divided into three groups: NRZ-based, RZ-based and novel modulation formats. In this section, the NRZ modulation formats are introduced. Non-return-to zero (NRZ) and duo binary modulation belong to this group.

2.1.1 Non-return-to-zero (NRZ) modulation

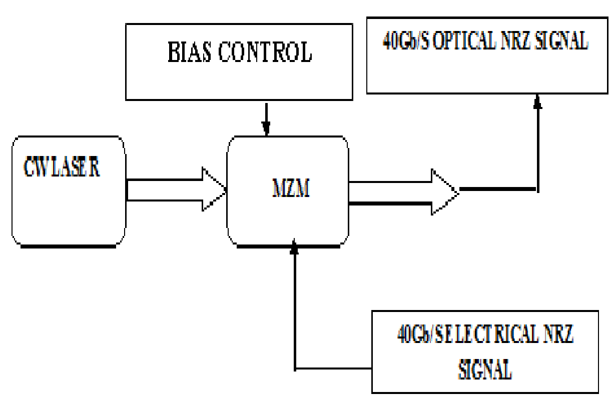
The non-return-to-zero (NRZ) has been the dominant modulation format in intensity modulated-direct detection (IM/DD) fiber-optical communication systems for the last years. The reason behind using NRZ in the early days of fiber-optical communication systems were-

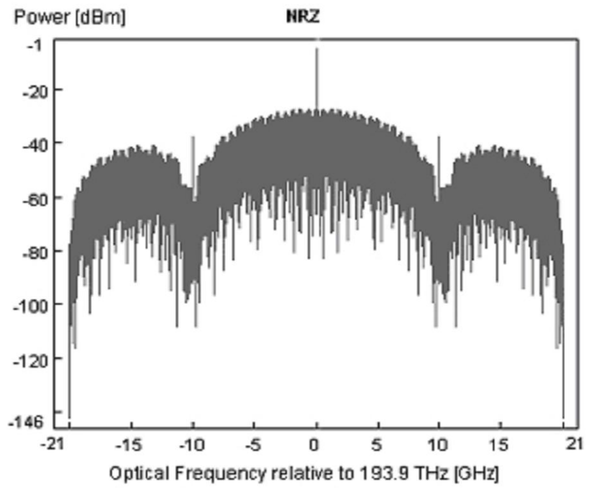
a) it is not sensitive to laser phase noise like Phase shift keying

b) it requires a relatively low electrical bandwidth for transmitters and receivers compared with RZ

c) it has the simplest configuration of transmitter and receiver.

NRZ transmitter and the spectrum of the modulated signal have been shown in figure 2.1. In the generator the intensity of the carrier light wave is modulated by the applied electric field which voltage varies with a determined function. The Mach-Zehnder modulator (MZM) is driven at the quadrature point of the modulator power transfer function with an electrical NRZ signal. In NRZ modulation the spectrum of carrier more strong than the signal component so the carrier contains the more power than the signal. It is also noted from the spectrum that the dispersion tolerance is improved on the cost of ISI effects. [2, 3]





(b)

Fig.2.1a) NRZ Transmitter and b) Optical spectrum of NRZ signal with 10 GB/s of data rate [2, 3]

2.1.2 *Duo binary modulation*

It is a combination of a conventional ASK-based modulation and phase shift keying (PSK). Duo binary transmission technology was introduced for the first time by A. Lender [4] in the 1960s as a mean of transmitting binary data over an electrical cable with high-frequency cutoff characteristics.

Figure 2.2a illustrates an optical spectrum and figure 2.2b a signal waveform of a 40 GB/s duo binary signal. Due to binary filters used in this method the spectral width is twice as compared to the condensational NRZ signals so the reduced spectral with means tolerance to dispersion is more. So the spectral efficiency of the optical network is increased .In the due binary modulation the SSB effects is also introduced due to supersession of carrier. [4, 5]

.

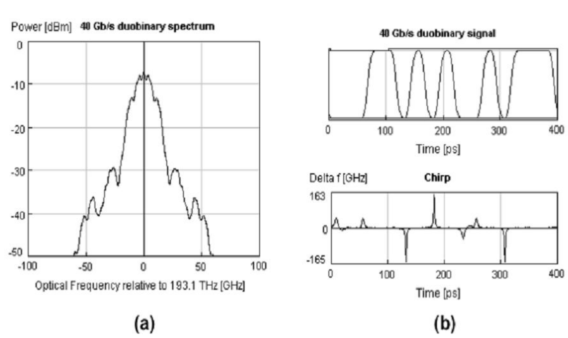


Figure 2.2: 40 GB/s duo binary signal: a) optical spectrum b) signal shape and chirp [4-5]

RZ-BASED MODULATION FORMATS:

This section elaborates the signal generation of RZ-based modulation formats. The RZ-based modulation formats considered here are return-to-zero (RZ), carrier-suppressed RZ (CSRZ), single side-band RZ (SSBRZ) and chirped RZ (CRZ) modulation. The collective characteristics of these formats are a duty ratio smaller than 1 and a broad signal spectrum.

2.2.1 Return-to-zero (RZ) modulation

The generation is given in Figure 2.3 is the frame about the principle of the generation of RZ which is all composed by the two concatenation of MZM. The technology of RZ code prevails recently, which is used in the high speed of 40 GB/s optical transmission system. In the pulse sequence of RZ code, the transition area which connects “1” amplitude of electric field has the independent time envelope. Because modulation format of RZ has different transition all the time in the code bits, thus it can bring more “neatness” optical signal in order to unscramble the receiver. The advantage of RZ is the low average of optical power; higher ability on anti-non-linearity effect and anti-polarization mode dispersion (PMD) [3, 5]. RZ code is also more conducive to clock recovery. Because the consecutive “1” of NRZ is a whole, the eye pattern of RZ code stretches bigger, the better ability of anti-error code performance, and the greater improvement on 3dB of the optical signal noise ratio (OSNR). The RZ pulses occupy just a part of the bit slot, resulting in a duty cycle value smaller than 1.

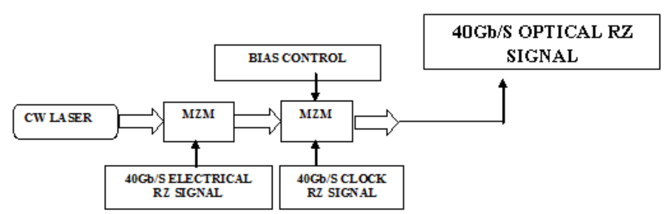


Figure 2.3 RZ Transmitter sections [4, 5]

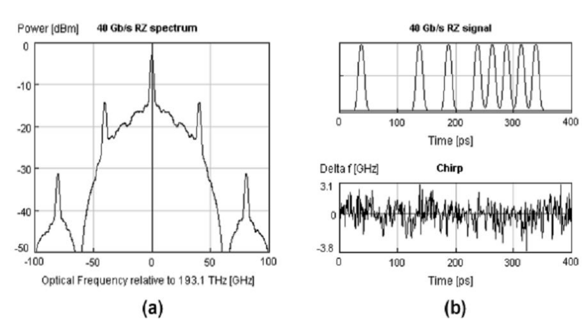


Figure 2.4: 40 GB/s RZ signal: a) optical spectrum b) signal shape and chirp [4-5]

2.2.2 Carrier-suppressed RZ (CSRZ) modulation

Carrier-suppressed RZ (CSRZ) modulation is also a newly proposed modulation formats for high bit rate transmission systems, which has been intensively investigated in numerical and experimental works [2-5]

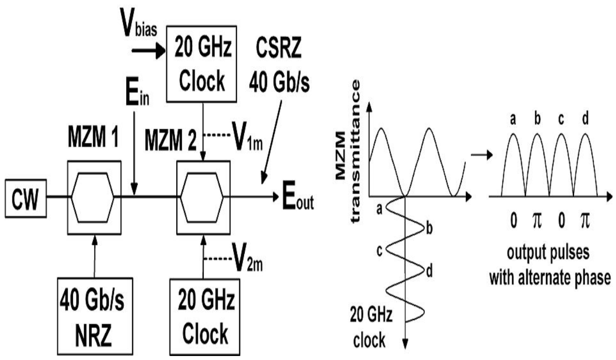


Figure 2.5: Generation of 40 GB/s CSRZ signal [5]

The main target of this modulation format is a reduction of the nonlinear impacts in a transmission line and an improvement of the spectral efficiency in high bit rate WDM systems. Additionally, it can be expected that the dispersion tolerance of the transmission can be improved as well by CSRZ modulation, due to its reduced spectral width compared to conventional RZ modulation. The generation of CSRZ pulses is presented in Fig. 2.5. The first MZM modulator (MZM 1) (Fig. 2.6) generates a 40 GB/s NRZ optical signal by external modulation of the CW-pump light. MZM 1 is driven with a 40 GHz electrical NRZ signal at the"quadrature"-point of the modulator power function. The CSRZ signal forming is realized in the second MZM (MZM 2), which is biased at the"zero"-point.

The spectrum and the shape of figure 2.6showes that the spectral with is reduced with a factor of two as compared to RZ .the tolerance due to non linearity's is also increased due to phase difference between adjacent bits which is a great advantage of CSRZ.[3,5,6]. The nonlinear tolerance of CSRZ modulation can be enhanced by the implementation of pre-chirp at the transmitter side Therein, the amount of pre-chirp has to be carefully optimized in order to avoid an increase of the linear crosstalk and waveform distortions. [7]. Due to an RZ pulse shape, the CSRZ modulation offers better receiver sensitivity than conventional NRZ modulation [4-5]. By the implementation of sophisticated filtering methods (e.g. asymmetrical filtering) the robustness of CSRZ modulation to narrow-band filtering can be improved, which can be beneficial for DWDM systems By the use of an optimized narrow-band filtering or a polarization multiplexing] in 40Gb/s CSRZ based DWDM transmission systems, a spectral efficiency beyond 0.4 bit/s/Hz can be realized.[3-7]

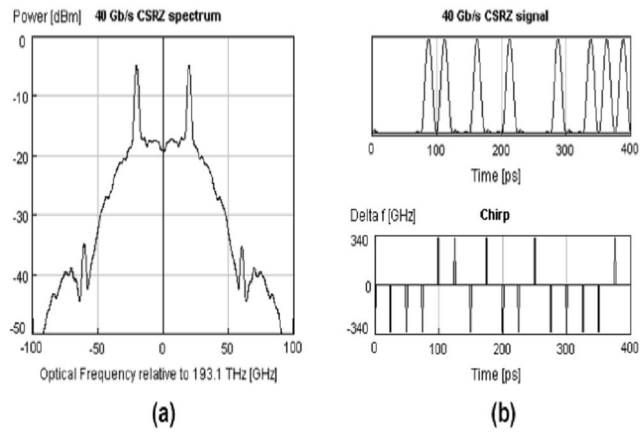


Figure 2.6: 40 GB/s CSRZ signal: a) optical spectrum b) signal shape and chirp. [7]

2.2.3Chirped RZ (CRZ) modulation

Chirped RZ (CRZ) modulation is a special case of RZ modulation realized by the implementation of the pre-chirp on the conventional RZ pulses at the transmitter side. To date, CRZ modulation is basically used in long-haul under-sea transmission systems at channel data rates up to 10 GB/s. The pre-chirping of RZ pulses can be realized in various manners. The conventional CRZ generation is realized by a phase modulation in an additional modulation stage. spectrum of the CRZ signal caused by the phase modulation with a 40 GHz clock signal and relatively high phase modulation index (typically, b=1.5 rad) Recent experimental studies showed that an system implementation of the CRZ modulation format can reduce the waveform distortions[8,9].

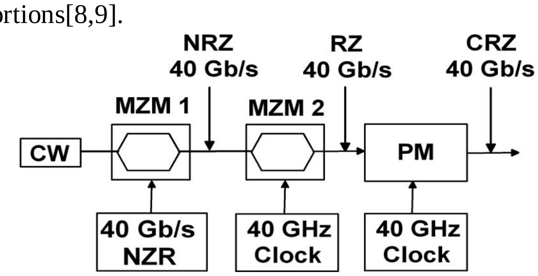


Figure 2.7: Generation of 40 GB/s CRZ signal [8]

The signal generation of CRZ pulses is presented in Fig. 2.7. In a CRZ transmitter, a CW laser is modulated with a conventional data stream in order to generate a NRZ signal. The CRZ signal is formed by re-modulating the NRZ signal's amplitude and phase with a sinusoidal electrical drive into an amplitude MZM and PM. The phase modulation index of PM impacts the spectral forming of the

CRZ signal and the spectral changes are governed by the PM drive signal. For example, higher clock frequencies would cause larger spectral broadening. In CRZ pulses a sine-clock is used with the frequency equal to the signal bit rate (e.g. 40GHz).

The optical spectrum and signal waveform of a 40 GB/s CRZ signal are shown in Fig. 2.8. It can be seen that the signal shape (Fig. 2.8b) remains unchanged compared to the RZ case, but changes occur in the optical

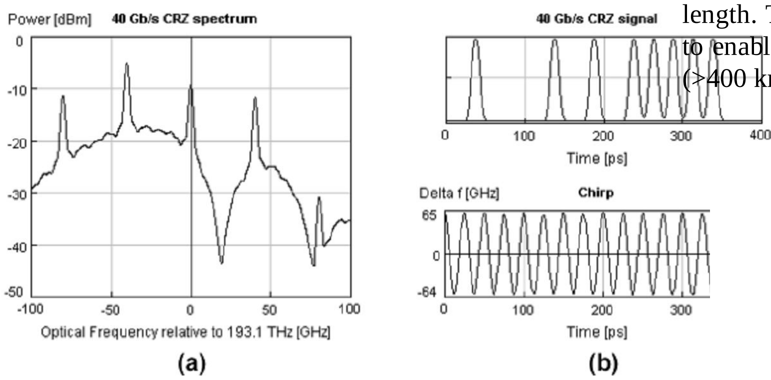


Figure 2.8: 40 GB/s CRZ signal: a) optical spectrum b) signal shape and chirp [9]

Associated with conventional transmission formats. CRZ pulses enable an improvement of transmission characteristics due to the enhanced nonlinear tolerance. The drawback of CRZ is relatively wide optical spectrum reducing the system spectral efficiency and dispersion tolerance. Due to its good transmission characteristics CRZ modulation can be of interest for 40 GB/s WDM systems with a spectral efficiency smaller than 0.4 bit/s/Hz. [9]

2.3 Novel modulation formats:

As we want to improve the transmission distance at high bit rates in optical networks new formats has been devolved known as Novel modulation formats.

2.3.1 Alternate Chirped NRZ (alCNRZ) modulation

In alCNRZ modulation formats the pre-chirping is done at the transmitter which helps to improve the transmission distance prechirping can be done by using either active components or the passive components .the tolerance to dispersion is reduced by prechirping because it increases the spectral width so pre chirping should be chosen carefully.
The pre-chirping of optical pulses at the transmitter

Modulation formats with a reduced nonlinear tolerance (e.g. NRZ-based formats) could primarily profit from the implementation of a pre-chirp. The main reason against pre-chirping of NRZ pulses until recently was the induced spectral broadening resulting in a stronger impact of group velocity dispersion (GVD). This effect can be partly avoided by implementing a carefully chosen amount of phase modulation on NRZ pulses. This is realized in alternate chirped NRZ (alCNRZ) pulses. The alCNRZ modulation could be a good candidate for a performance improvement in deployed optical transmission systems because of its simple generation. The improved nonlinear tolerance would provide an enhancement of the maximum transmission length. Therefore, the main target of alCNRZ modulation is to enable a high speed optical transmission over long-haul (>400 km) distances. [5, 7.10, 11]

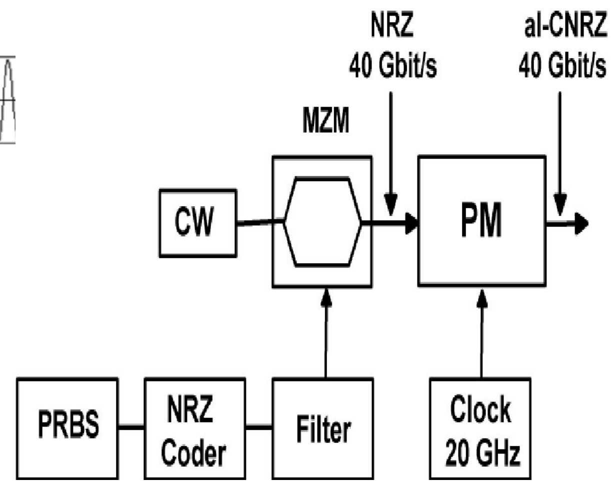


Figure 2.9: Generation of 40 GB/s alCNRZ signal [5-7]

The alCNRZ generator is presented in Fig. 2.9. Starting from a conventional generation of NRZ signals, alCNRZ pulses are generated by a phase modulation of NRZ pulses in an additional phase modulator (PM], which is driven by a clock signal at half the bit rate. The neighboring pulses are alternate chirped. [12, 13]

The signal shape and the optical spectra for different phase modulation indices in an alCNRZ signal are presented in Fig. 2.10. Because of phase modulation, new spectral components arise (Fig. 2.10), hence broadening the signal spectrum depending on the implemented amount of phase modulation. At the same time, the modulation index governs the phase shifts between adjacent bits (Fig.2.10 bottom line). By the implementation of an optimum amount (m=2.25 rad) of the alternate phase shift, alCNRZ modulation can enable an even better transmission performance than a conventional RZ modulation in 40 GB/s based WDM systems.

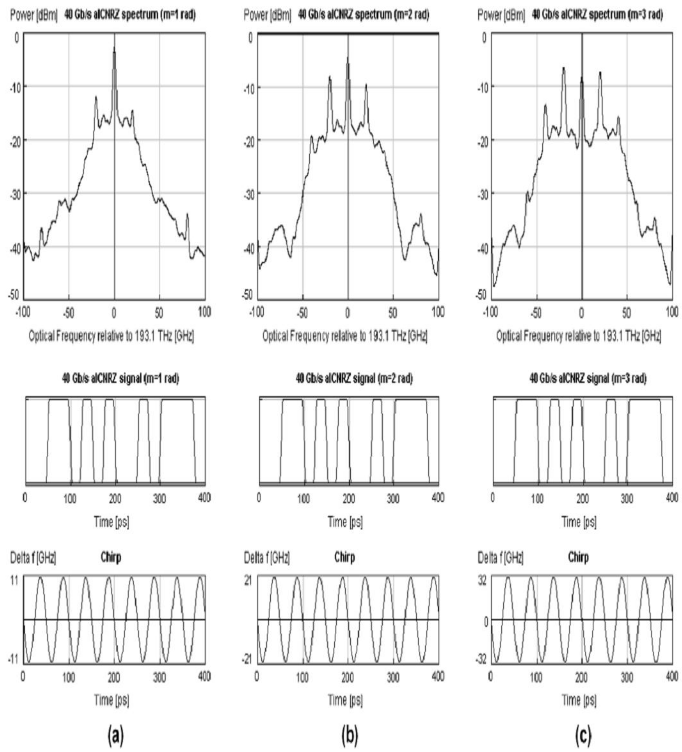


Figure 2.10: 40 GB/s alCNRZ optical spectrum and signal shape for different values of modulation index:
a) m=1 rad b) m=2 rad c) m=3 rad[5,7,10,11]

2.3.2 Alternate chirped RZ (al-RZ) modulation

The alternate-chirped RZ (al-RZ) modulation format was proposed for the first time at OFC 2001Its spectral form shows similarities to a CSRZ spectrum (Fig. 2.6a) with a small amount of additional pre-chirp. The generation of the al-RZ pulses is illustrated in Fig. 2.11. Using a phase modulator (PM), the output of a CW-pump is modulated by a 20 GHz clock signal with an oscillation voltage of V_. The narrow-band filter extracts the carrier and the first side-bands. A flat-top interleaver with 3 dB bandwidth of 0.35nm is used as the narrow-band filter. The intensity modulator (IM) is used for encoding 40 GB/s data. The modulation used here does not require any bias-voltage control between modulation stages and is driven by a single clock at half the bit rate. [9, 10, 14] The optical spectrum and a signal waveform of 40 GB/s al-RZ are shown in Fig. 2.12. The al-RZ spectrum consists of three dominant components: carrier and two side-bands at 20 GHz left and right (Fig. 2.12b). The spectral width is similar to CSRZ and the signal shape is same as the conventional RZ pulses. This combination of the phase modulator and the narrow band filter provides the generation of 40 GB/s al-RZ pulses, but also the realization of a sinusoidal chirp, causing the alternate chirp inversion between neighboring bit slots (Fig. 2.12b), which can be useful for the suppression of the interferences between neighboring pulses[15].Due to a strong concentration of the optical spectrum around the carrier and a narrow spectrum, al-RZ could be a potential candidate for 40 GB/s based DWDM systems.

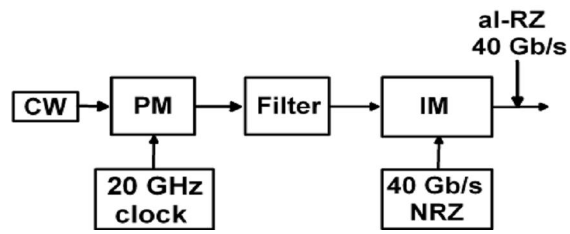


Figure 2.11: Generation of a 40 GB/s al-RZ signal[14-15]

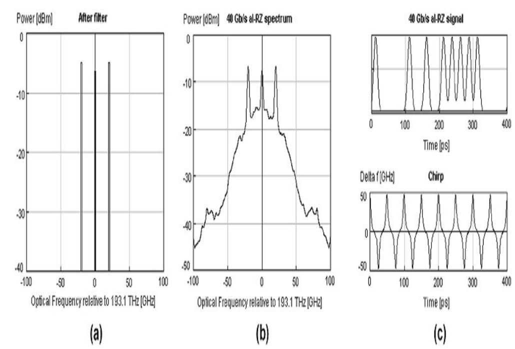


Figure 2.12: 40 GB/s al-RZ Signal: a) Optical spectrum b) Final spectrum c) signal shape and chirp [15]

III. CONCLUSION

From the study given above it is fond that each modulation format has discrete characteristics so to obtain the desire characteristics from an optical network the choice of a suitable data format has a great significance.

REFERENCES

[1] Govind P. Agrawal, “Fiber-optic communication systems”, Third edition, 2002

[2] H. Takahashi, I. Nishi, and Y. Hibino, “10GHz Spacing Optical Frequency Division Multiplexer Based on Arrayed-waveguide Grating,” IEE Electronics Letters, Vol. 28, No. 4, pp. 380–382, 1992.

[3] S. Suzuki and A. Sugita, “Recent Progress in Silica-based Planar Light wave Circuits (PLCs),” NTT Technical Review, Vol. 3, No. 7, pp. 12–16, 2005.

[4] Y. Hibino, “Photonic Integration Technologies for Large capacity Telecommunication Networks,” Proc. of LEOS Annual Meeting 2009, Paper PLE3, Belek-Antalya, Turkey, 2009

[5] Y. Miyamoto and S. Suzuki, “Advanced Optical Modulation and Multiplexing Technologies for High-capacity OTN Based on 100 Gb/s Channel and Beyond,” IEEE Commun. Mag., Vol. 48, No. 3, pp. S65–S72, 2010.

[6] C. Kurtzke. Kapazit¨atsgrenzen digitaler optische ¨ Ubertragungssysteme. PhD thesis, Technische University¨at Berlin, 1995.

[7] B. Wedding and B. Franz. Unregenerated optical transmission at 10 Gbit/s via 204 km of standard single mode fiber using a directly modulated laser diode. Electronics Letters, 29(4):402–403, February 1993.

[8] B.Wedding, B. Franz, and B. Junginger. 10-Gb/s optical transmission up to 253 km via standard singlemode fiber using the method of dispersion-supported transmission. Journal of Lightwave Technology, 12(10):1720–1727, October 1994.

[9] B. Wedding. Analysis of fibre transfer function and determination of receiver frequency response for dispersion supported transmission. Electronics Letters, 30(1):58–59, January 1994.

[10] R. Gross. Differential encoder for QPSK systems. IEE Electronics Letters, 27(14):1256–1257, July 1991.

[11] M. Rhode, C. Caspar, N. Heimes, M. Konitzer, E. J. Bachus, and N. Hanik. Robustness of DPSK direct detection transmission format in standard fiber WDM systems. Electronics Letters, 26(17):1483–1484, August 2000.

[12] C. Wree, J. Leibrich, and W. Rosenkranz. RZ-DQPSK format with high spectral efficiency and high robustness towards fiber nonlinearities. European Conference on Optical Communication (ECOC), 4(9.6.6), September 2002.

[13] R. Griffin, R. Johnstone, R. Walker, S. Wadsworth, and A. Carter. Integrated DQPSK transmitter for dispersion-tolerant and dispersion-managed DWDM transmission. Optical Fiber Communication Conference (OFC), 2(FP6):770–771, March 2003

[14] A. Chowdhury , “Recent advances in optical phase conjugation and its application to 40GB/s transmission”, IEEE Photonic Technology Letters, (2005)

[15] A. Singh, “Modulation Formats for High-Speed, Long-Haul Fiber Optic Communication Systems,”Inphi Corporation, 2393 Towns gate Rd #101, Westlake Village, CA 91361.

A Review paper on Image Compression Using Wavelet Transformation

Mandeep Kaur
MTech. ECE, Amritsar College of Engg. & Technology, Amritsar
Email:mandip.khehra@gmail.com

***Abstract*- With the advancement in communication and multimedia technology, it expands the need of image storage and transmission. For image compression, it is desirable that maximum redundant data must be compressed and image quality is preserved. Discrete wavelet transform is one of the most popular compression technique as DWT is multiresolution in nature helps to decompose the images into set of sub images called shapes with different resolution correspond to different frequency bands. Wavelet transformation based image compression provides considerable improvement in picture quality at high compression ratio as compare to other compression techniques.**

***Keywords:* Discrete wavelet transforms, Wavelet transformation, Image compressionand Compression ratio.**

1. INTRODUCTION

Image compression is required for better storage and transmission of data over communication media. As high resolution images require more space and need to be compressed. It reduces the number of bits required to represent an image by removing special and spectral redundancies. Wavelet transform is useful for both stationary and non stationary signals. There are basically three steps used in image compression Firstly, the image coefficients are transformed to wavelet coefficients using wavelet transformation algorithms,then they are quantized using quantizerand lastly a threshold is selected,so that all redundant data is compressed. The output bits are reduced required to represent the compressed image.Wavelet transformation is important for applications where scalability and tolerance degradation is important.

2. WAVELET TRANSFORMATION

The discrete wavelet transformation decomposes the signal into basis function and this basis function is known as wavelets.By dilation and shifting a wavelet y(t) is obtained from mother wavelet.

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right)$$

Where a is the scaling parameter and b is shifting parameter. The wavelet transformation is computed for different frequencies so that it gives different resolution. The multiresolution analysis gives excellent time resolution and poor frequency resolution at higher frequencies but it gives excellent frequency resolution and poor time resolution at lower frequencies. So wavelet transformation is good for signals having higher frequency components for short duration and low frequency components for long duration. The 2-D wavelet transform is obtained by performing two 1-Dtransforms. Signal is passed through high and low pass filters and then split the image into four bands LL, LH, HL, and HH.The inverse transform is obtained by up sampling all sub bands by factor of 2 and uses a reconstruction filter.

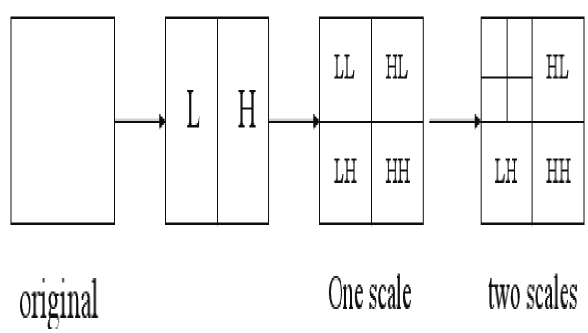


Fig. 1: Two level wavelet decomposition

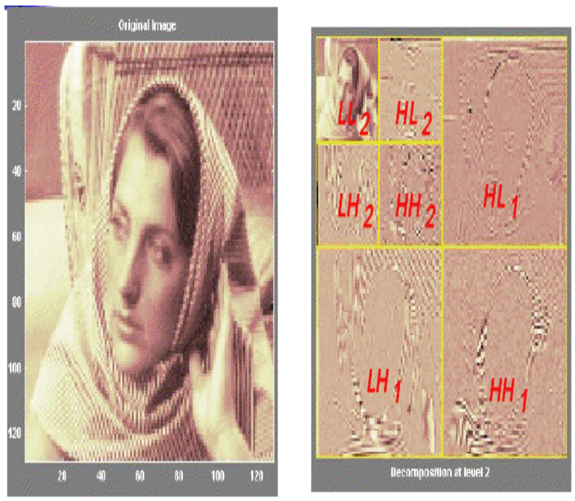


Fig 2: Two level wavelet decomposition applied on an image

DWT is basically used for stationary signals as it has limitation regarding non stationary signals like poor directionality, absence of phase information and aliasing. To process an image symmetric bioorthogonal wavelets are used.EZW is the popular algorithm used for wavelet transformation.

1.1 EZW ALGORITHM

Embedded Zero Tree Wavelet use intra band dependencies between coefficients in one subband and coefficients describing the same special region in another subband.The low coefficient value in low frequency sub band is correlated with low values for all coefficients relating to same special locationin high frequency subbands.This phenomenon is known as zero tree and occur in smooth regions of image.Since each coefficient in a particular subband is related to four coefficients in the next higher subband,recognizing and coding the occurrence of a zero tree can considerably reduce number of coefficientswhich need to be transmitted. The single coefficient is defined as tree and all coefficients in high frequency sub bands corresponding to parts of spatial region described by root. as root is sown in sub band HH3 together with child nodes in the high frequency sub bands.EZW is an embedded algorithm which encodes quantized wavelet coefficient in a bit plane order. Each bit plane is coded stating from lower frequency sub band and on the higher frequency subband.

EZW algorithm is described as-

a) INITIALIZATION:Place all the wavelet coefficients on the dominant list. set all the initial threshold to T0=2

b) DOMINANT PASS**:** Scan the coefficient on the dominant list using the current threshold T1 and each coefficient has one of the four symbols.
POSITIVE SIGNIFICANT:It means that all coefficientsare significant relative to current to current threshold T1 and positive.
NEGATIVE SIGNIFICANT**:** It means that all coefficientsare significant relative to current to current threshold T1 and negative.
ISOLATED ZERO: It means that all coefficientsare significant relative to current to current threshold T1 and one or more of its descendants are significant.
ZERO TREE ROOT:It means that current coefficients and all its descend ants are insignificant relative to the current threshold T1.

c) SUBORDINATE PASS**:**It outputs a 1or 0 for all coefficients on the subordinate list depending on whether the coefficients is in the upper or lower half of the quantization interval.

d) LOOP:It reduces the current threshold by two, until the target fidelity or bit rate is achieved.

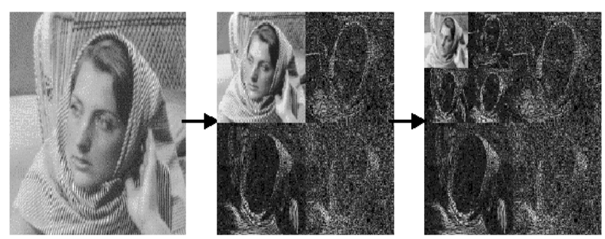


Fig 3: Dyadic decomposition into subbands for test image “Barbara”

2. THRESHOLDING

Threshold is selected in such a way that it preserves only certain number of coefficients among all.It neglects all wavelet coefficients that fall below a certain threshold.The smaller value of DWT has less information of image so they are neglected.Some details of picture are lost during thresholding but they cannot be identified by human eye.

2.1 HARD THRESHOLDING

In hard thresholding,a tolerance is selected;any wavelet whose tolerance value falls below the tolerance is set to zero with the goal to introduce many zeros without losing a great amount of detail.It is difficult to select a threshold as larger as threshold is selected,more errors are introduced.

$$c_i' = \begin{cases} c_i & \text{if} \quad c_i > T \\ 0 & \text{otherwise} \end{cases}$$

2.2 SOFT THRESHOLDING

In soft thresholding, a tolerance h is selected. If the absolute value is less than tolerance, then thatentry is set to zero. The relation for soft thresholding is

$$c_i' = sgn(c_i) \cdot max(|c_i| - T, 0)$$

3. ENTROPY CODING

The transformation and thresholding process the signal but no compression is done up to this point. So Huffman entropy coding is used so that the integer sequence q is changed to shorter sequence. The strings of zeros are coded by numbers1 through 100,105 and 106, whilenon zero are coded by 101 through 104and 107 through 254.

4. QUANTIZATION

The last step is known as quantization,as it converts the sequence of floating numbers w into a sequence of integer numbers q.The simplest form is to round to nearest integer. The vector quantization is best for quality and compressing images. In this encoding is performed by appropriating the sequence to be encoded by vector belonging to codebook. LGB is the most common algorithm for creating this codebook.

.

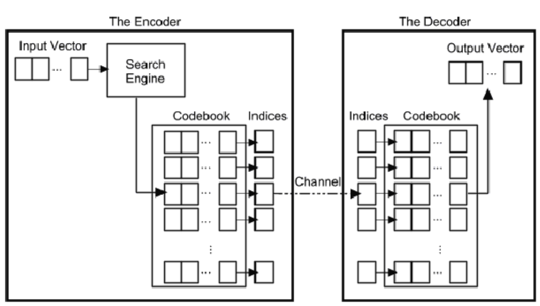


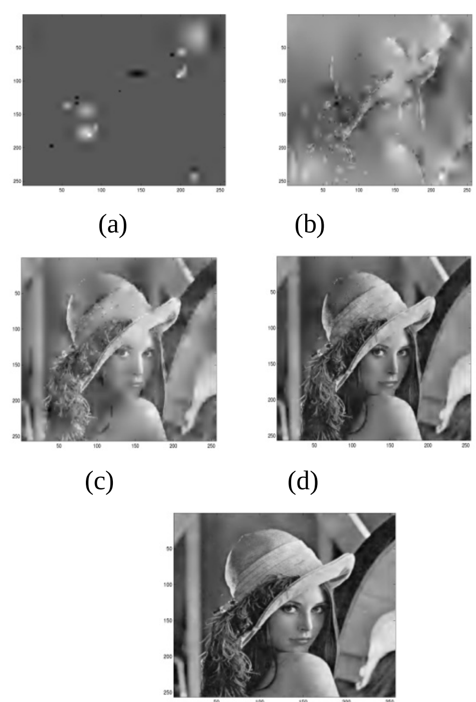
Fig 4: schematic of vector quantizer

4.1 LGB ALGORITHM

- Determine the number of code words N or the size of codebook.
- Select N code words at random and let that be the initial codebook. Theinitialcode words can be randomly chosen from set of input vector.
- Using the Euclidean distance measure clusterize the vectors around each codeword. This is done by taking each input vector and finding the Euclidean distance between it and each codeword. The input vector belongs to the cluster of codeword that the minimum distance.
- Compute the new set of code words. This is done by obtaining the average of each vector and divide by number of vector in the cluster.
- Repeat step 2 and 3 until the either of code words do not change or the change in codeword is small.

5 .IMPLEMENTATION OF DWT COMPRESSION

Fig 5: Original image of fabulous Lena



(a) (b) (c) (d) (e)

Fig.6:Compressed images using DWT

The practical implementation of DWT compression is tested on Lenna's image using different threshold and compression ratios. The image shows the effect of different compression ratios. The compression ratio is applied according to the demand rate of channel and requirement of picture quality. The compression ratio 5:1 gives the best picture quality.

(a) Threshold = .5, Bases included = 19, Compression ratio = 3400:1
(b) Threshold = 1, Bases included = 470, Compression ratio = 140 : 1
(c) Threshold = 2, Bases included = 2383, Compression ratio = 27 : 1
(d) Threshold = 4, Bases included = 6160, Compression ratio = 10 : 1
(e) Threshold = 8, Bases included = 12378, Compression ratio = 5 : 1

6.COMPARISON OF DWT WITH DCT

- DWT is highly flexible as wavelet function can be freely chosen.
- In DWT,there is no need to divide the input coding into non-overlapping 2-D blocks,it has higher compression ratio to avoid blocking artifacts.
- Using DWT,the transformation of the whole image introduces inherent scaling.

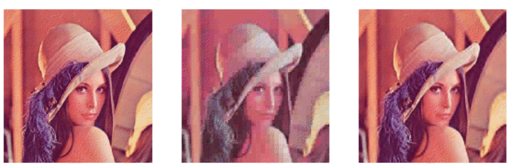


Fig 7:(a)Original image 256*256 pixels
b) DCT compressed image with compression ratio 43:1
c) DWT compressed image with compression ratio 43:1

7. CONCLUSION AND FUTURE SCOPE

DFT has been recognized as a faster approach. The use of DWT provides a good tradeoff between spatial and frequency resolution. This unique property of wavelet does not occur in other transform. It has better energy concentration property. Choice of transform effect on image quality and system design issue.. It is able to compress large images and require single decompression architecture. It provides transmission in noisy environment and robustness to bit errors. Its compression ratio is high as it avoid blocking artifacts.As a future work, improvement in EZW performance by including a preprocessing wavelet coefficient stage and reduce complexity of EZW to implement a real time motion wavelet video coder.

REFERENCES

[1] SHAPIRO JM Embedded Image Coding Using Zero Treeof Wavelet Coefficients 1993(12)
[2] C. Ben Amar and Q Jemai,"Wavelet Networks Approach For Image Compression", ICGST, 2003
[3] M. Vetterli, J.Kovavcevic, wavelets and subband coding, Ist(Edn.),Prentice-Hall,eaglewood Cliffs,NJ 1995
[4] Mallat S.,"A Theory For Multiresolution Signal decomposition:The Wavelet representation"'IEEE trans. Pattern anal. Machine intell.11 (1989).
[5] CCSDS, Image Data Compression recommended standard CCSDS 122.0-B-1Blue Book, Nov 2005.
[6] Sidney Burrus, 1998. Introduction to wavelets and Wavelet Transforms, prentice Hall.
[7] X.Zixiang,K.Ramchandran,M.T.Orchand and Y.Q.Zhang.A Comparative Study of DCT and Wavelet based Image Coding,IEEE tans. Circuits syst.Video Technol.,PP:692-695 1999.
[8] Yogendra Kumar Jain & sanjeev Jain" performance analysis and comparison of wavelet Families using for Image Compression", international Journal of Soft Computing 2(1): 161-171, 2007.
[9] Yogendra Kumar Jain & sanjeev Jain" performance Evaluation of Wavelets for Image compression",international Journal of Soft Computing 2(1): 1104-112,2006.
[10]H,Wang,Wu L.A novel Design of lifting Scheme from general wavelet.IEEE Trans signal processing.2001,49:1714-1717

WiMAX Physical layer: A Survey

Er. Saurabh kher[1], Sh. Gurpadam singh[2]
[1]Assistant Prof, E.C.E ,GNDU RC, Gurdaspur
[2]Associate Prof. E.C.E ,BCET Gurdaspur

Abstract- WiMAX provides Line of sight and non line of sight wireless connectivity. WiMAX uses orthogonal frequency division multiple access technique in wireless communication. WiMAX physical layer based on the IEEE 802.16e standard under different combinations of digital modulation schemes and different communication channels. This technique uses Adaptive modulation coding (AMC) on physical layer of WiMAX. Adaptive modulation technique uses the concept of cyclic prefix that adds additional bits at the transmitter end. The receiver removes these additional bits. Cyclic Prefix is used to combat intersymbol inter-ference (ISI) and intercarrier interference (ICI) introduced by the multipath fading channel. This paper investigates the performance of WiMAX network by varying physical layer parameter such as modulation and coding scheme and cyclic prefix. It also investigates the performance of WiMAX network by increasing traffic (number of downloading nodes) in network with different cyclic prefix

1. INTRODUCTION

WiMAX (Worldwide Interoperability for Microwave Access) is based on the IEEE802.16 standard for Metropolitan Area Networks (MAN). Its goal is to deliver wireless broadband access to customers using base stations with coverage distances in the order of miles. The IEEE WiMax/802.16 is a promising technology for broadband wireless metropolitan area networks (WMANs) as it can provide high throughput over long distances and can support different qualities of services. IEEE 802.16m-based WiMAX 2.0 system is required to provide up to 1 Gb/s peak transmission rate. The most efficient solution to achieve this challenging objective is to utilize wider channel bandwidth. Multicarrier is the technology to utilize wider bandwidth for parallel data transmission across multiple RF carriers. IEEE 802.16 wireless technology can be an excellent choice for Cellular/Mobile applications due to its robust bandwidth and long range. Residential Broadband: Practical limitations like long distance and lack off return channel prohibit many potential broadband customers reaching DSL and cable technologies. IEEE 802.16 can fill the gaps in cable and DSL coverage. In many rural areas, there is no existence of wired infrastructure. IEEE 802.16 can be a better solution to provide communication services to those areas. IEEE 802.16e supports mobility, so the mobile user in the business areas can access high speed services through their IEEE 802.16/WiMAX enabled handheld devices like PDA, Pocket PC and smart phone. The IEEE 802.16 standard supports multiple physical specifications due to its modular nature. The first version of the standard only supported single carrier modulation. Since that time, OFDM and scalable OFDMA have been included to operate in NLOS environment and to provide mobility. The standard has also been extended for use in below 11 GHz frequency bands along with initially supported 10-66 GHz bands. The exact frequency of operation for any given system is dependent on the propagation conditions that are encountered during its use. The frequencies higher than 10 GHz are practical only for fixed line-of-sight (LoS) type services. Non-line of sight (NLoS) communications perform better when the frequencies of operation are kept under 10 GHz. The frequencies below 6 GHz have better propagation properties and are better suited for mobile communications because they most likely guarantee service to all the niches of the coverage area [4].

Table 1: Different Flavour of IEEE 802.16

	IEEE 802.16-2001	IEEE 802.16a	IEEE802.16-2004	IEEE 802.16e-2005
Completed	December 2001	January 2003	June 2004	December 2005
Spectrum	10-66 GHz	2-11 GHz	2-11 GHz	2-6 GHz
Popagation/channel conditions	LOS	NLOS	NLOS	NLOS
Bit Rate	Up to 134 Mbps (28 MHz channelization)	Up to 75 Mbps (20 MHz channelization)	Up to 75 Mbps (20 MHz channelization)	Up to 15Mbps (5 MHz channelization)
Modulation	QPSK, 16-QAM (optional in UL), 64-QAM (optional)	BPSK, QPSK, 16-QAM, 64-QAM, 256-QAM (optional)	256 subcarriers OFDM, BPSK, QPSK, 16-QAM, 64-QAM, 256-QAM	Scalable OFDMA, QPSK, 16-QAM, 64-QAM, 256-QAM (optional)
Mobility	Fixed	Fixed	Fixed/Nomadic	Portable/mobile

In mobile WiMAX FFT size can varies between 128 and 2048 and to keep the subcarrier spacing at 10.94 KHz, the FFT size should be adjusted which is helpful to minimize Doppler spreads. Since there are different channel bandwidth like, 1.25, 5, 10 and 20 MHz etc, so FFT sizes are 128, 512, 1024 and 2048 respectively.

802.16, 802.16a and 802.16c were all together consolidated and a new standard was created which is known as 802.16. In the beginning, it was published as a revision of the standard under the name 802.16REVd, but the changes were so genuine that the standard was reissued under the name 802.16 at September 2004. In this version, the whole family of the standard is ratified and approved. This amendment was included in the current applicable version of standard IEEE 802.16 in December 2005[12]. This includes the PHY and MAC layer enhancement to enable combined fixed and mobile operation in licensed band.

II. SIMULATION MODEL

The structure of the baseband part of the implemented transmitter and receiver is shown in Fig. 1. This structure corresponds to the physical layer of the IEEE 802.162004 Wireless MANOFDM air interface. In this setup, we have just implemented the mandatory features of the specification, while leaving the implementation of optional features for future work.

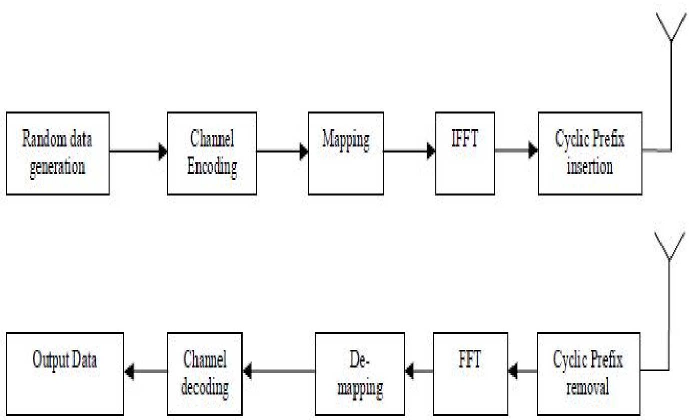


Fig. 1

Channel coding part is composed of three steps Randomization, Forward Error Correction (FEC) and interleaving. FEC is done in two phases through the outer Reed-Solomon (RS) and inner Convolution Code (CC).

The complementary operations are applied in the reverse order at channel decoding in the receiver end. The complementary operations are applied in the reverse order at channel decoding in the receiver end.

The complete channel encoding setup is shown in Fig. 2 while corresponding decoding setup is shown in Fig. 3

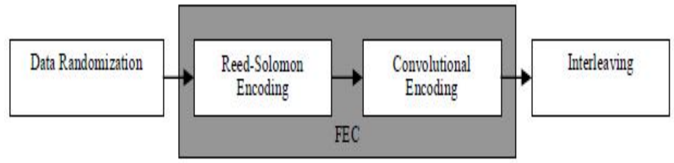


Fig. 2

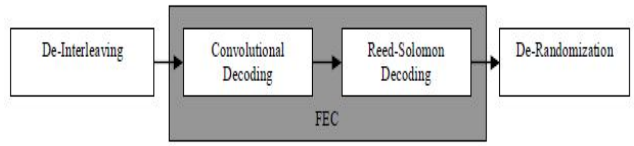


Fig.3

The forward error control (FEC) consists of a Reed-Solomon (RS) outer code and a rate compatible Convolutional Code (CC) inner code. A block Reed Solomon (255,239,8) code based on the Galois field GF (28) with a symbol size of 8 bits is chosen that processes a block of 239 symbols and can correct up to 8 symbol errors calculating 16 redundant correction symbols. Reed Solomon Encoder that encapsulates the data with coding blocks and these coding blocks are helpful in dealing with the burst errors. The block formatted (Reed Solomon encoded) datast ream is passed through a convolutional interleaver. Here a code rate can be defined for convolutional codes as well. Ift here are k bits per second input to the convolutional encoder and the output is n bits per second, the code rate is k/n. The convolutionally encoded bits are interleaved further prior to convert into each of the either four complex modulation symbols in BPSK, QPSK, 4- QAM, 16-QAM modulation and fed to an OFDM modulator for transmission.

Stanford University Interim (SUI) Channel

There are six channels selected to address three different terrains. These three terrains are defined as A, B and C are terrain A is hilly terrain with heavy tree density, B is hilly terrain with light tree density and C is mostly flat terrain with light tree density . These models can be used for simulation, design, and development and testing of technologies suitable for fixed broadband wireless applications. The following parameters were selected on the basis of some statistical models. We are presenting here the parametric view of all the six channels in Table 1

Table 1: Terrain Types for SUI Channel

Terrain Type	**SUI Channels**
C (Flat terrain)	SUI-1, SUI-2
B (Hilly terrain)	SUI-3, SUI-4
A (Hilly terrain)	SUI-5, SUI-6

III. CYCLIC PREFIX ADDITION

The subcarrier orthogonality of an OFDM system can be jeopardized when passes through a multipath channel. CP is used to combat ISI and ICI introduced by the multipath channel. CP is a copy of the last part of OFDM symbol which is appended to the front of transmitted OFDM symbol. The length of the CP (*Tg*) must be chosen as longer than the maximum delay spread of the target multipath environment. Fig 4 depicts the benefits arise from CP addition, certain position within the cyclic prefix is chosen as the sampling starting point at the receiver, which satisfies the criteria $\tau max < Tx < Tg$[8] . Where τmax is the maximum multipath spread. Once the above condition is satisfied, there is no ISI since the previous symbol will only have effect over samples within [0, τmax].

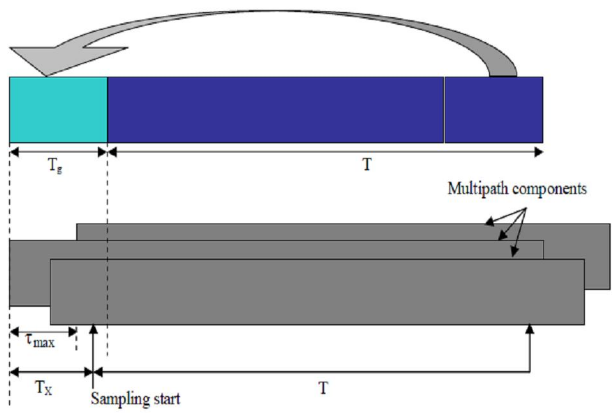


Fig.4

And it is also clear from the figure that sampling period starting from *Tx* will encompass the contribution from all the multipath components so that all the samples experience the same channel and there is no ICI.

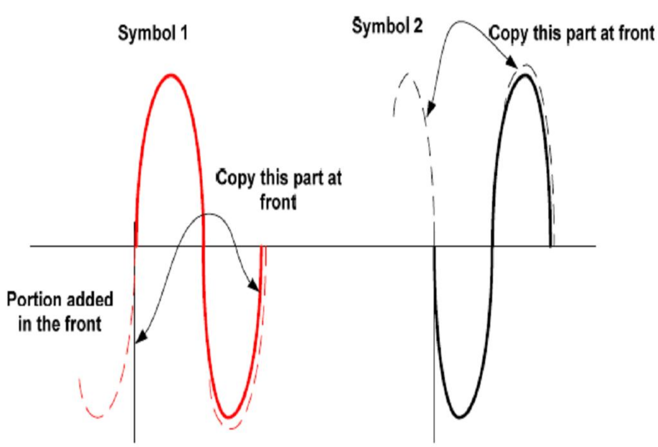


Fig.5

Adaptive Modulation and Coding

The specified modulation scheme in the downlink (DL) and uplink (UL) are binary phase shift keying (BPSK), quaternary PSK (QPSK), 16 quadrature amplitude modulation (QAM) and 64QAM to modulate bits to the complex constellation points. The FEC options are paired with the modulation schemes to form burst profiles. The PHY specifies seven combinations of modulation and coding rate, which can be allocated selectively to each subscriber, in both UL and DL. There are tradeoffs between data rate and robustness, depending on the propagation conditions. Table shows the combination of those modulation and coding rate.

Modulation	Uncoded Block Size (bytes)	Coded Block Size (bytes)	Overall coding rate	RS code	CC code rate
BPSK	12	24	1/2	(12,12,0)	1/2
QPSK	24	48	1/2	(32,24,4)	2/3
QPSK	36	48	3/4	(40,36,2)	5/6
16-QAM	48	96	1/2	(64,48,8)	2/3
16-QAM	72	96	3/4	(80,72,4)	5/6
64-QAM	96	144	2/3	(108,96,6)	3/4
64-QAM	108	144	3/4	(120,108,6)	5/6

REFERENCES

[1] M.A. Hasan, “Performance Evaluation of WiMAX/IEEE 802.16 OFDM Physical Layer,” Master of Science in Technology, Espoo, June 2007.

[2] Ebian, M. Shokair, and K. H. Awadalla. “Performance Evaluation of WiMAX System Using Convolution Product Code (Cpc)” Progress in electromagnetic Research C, Vol. 5, 125–133, 2008.

[3] Yang Xiao,"WiMAX/Mobile MobileFi: advanced Research and Technology",Auerbach Publications,2008.

[4] Md. Ashraful Islam, Md.Zahid Hagan “Performance Evaluation of WiMax physical layer under Adaptive Modulation Techniques And Communication Channels”,(IJCSIS) International Journal of Computer Science and Information Security,Vol. 5, No.1, 2009

[5] Gazi Faisal Ahmed Jubair ,Muhammad Imran Hasan ,Md. Obaid Ullah. “Performance Evaluation of IEEE 802.16e (Mobile WiMAX) in OFDM Physical Layer ”. MEE09:56, pp. 93. ING/School of Engineering, 2009.

[6] Omar Arafat ,K.Dimyati i “Performance Parameter of Mobile WiMAX”, Second International Conference on Computer Engineering and Applications 2010

[7] Mohamed A. Mohamed, Heba M. Abd-El- Atty , “Performance Simulation of IEEE 802.16e WiMAX Physical Layer”,.2010 IEEE

[8] Kene J.D,Kulat k.D “Performance evaluation of IEEE 802.16e WiMax physical layer ”, Engineering (NUiCONE), 2011 Nirma University International Conference Dec 2011

[9] Jamal Mountassir , Horia Balta , Marius Oltean, Maria Kovaci, Alexandru Isar " Simulating the WiMAX Physical Layer in Rayleigh

Fading Channel ,"Journal of Wireless Networking and Communications. 2011.

[10]Tapan kumar Godder, Md Sipon Miah , Md. Azziur Rahman in their paper "Performance Analysis comparison Btween the fixed & Mobile WiMAX Physical Layer"VOLUME 02, ISSUE 01, MANUSCRIPT CODE: 110723 2011

[11]Anand Nayyar in their paper, "Performance Analysis of WIMAX Physical Layer for Various Modulation Schemes," National Conference on Recent Advancements in Science and Engineering (RAISE 2012)

[12] http://en.wikipedia.org/wiki/WiMAX

To Detect Maturity Level of Fruit Using Color Image Segmentation

Meenu Dadwal[1] and Dr.Vijay Kumar Banga[2]
[1]M.Tech scholar, Electronics & Communication Engineering , Amritsar College of Engineering and Technology,Amritsar,
[2]H.O.D., Electronics & Communication Engineering ,Amritsar College of Engineering and Technology,Amritsar, India,
er.meenudadwal@gmail.com,vijaykumar.banga@gmail.

***Abstract*— In this paper, A low-cost visual-based color classification system is presented. Here we represent a general approach to find the ripeness of fruits using RGB color space. We check ripeness by estimating the color of fruit. To detect a particular color we set threshold values of three primary colors - Red, Green and Blue in RGB color space. In an image of fruit where these values will match our threshold values that color will be displayed in resulted image.**

Keywords-color image segmentation,RGB color space

I. INTRODUCTION

A colored image is combination of different colors. From the scientific viewpoint we can see colors if visible light reflects over its surface. A visible light is a set of electromagnetic radiations which can be recognize by human eye. In terms of wavelength its range lies between 390 to 750 nm and in term of frequency from 400 to 790 THz. In case of fruit color detection we fully dependent on techniques to find true colors of fruit images.

In older days, human depends upon its vision qualities to differentiate between ripe and unripe fruits. But this method had high rate of errors because of illness, distraction and other factors during working hours [6]. This also may effects the working speed of system. So to decrease this failure rate human started to invent new methods. These days, there are various methods to detect the ripeness of fruits and vegetables. In some methods we apply chemicals on fruits and sometimes we use machines. As we know, chemical may effects human health so usually machines are used for this purpose. Machine use their visual-based color classification system that provide reliability, high speed and repeatable operation. Hence the production increases and reduces its dependency on manpower.

In machine vision system computer uses different method to analyze the given image of fruit and vegetable. Previously, computer systems were not robust enough to operate on large and real colors of images, so mostly gray scale images had been the main focus for researchers. But today, computer system has been developed enough to work on large and true color images [2].

To increase the efficiency of computer system different researchers perform various experiments on different fruits and vegetables to check their maturity levels. In 2004, F. Mendoza, et al converted the RGB image of bananas into CIELAB format and accuracy reaches to 98% [4]. In 2003 and 2008, author uses average values of RGB to evaluate maturity levels of peaches, apples and oranges [5].In 2010, Zhi-yuan Wen, et all used machine vision to detect the maturity of citrus fruits. B. Ojeda-Magana, et al used different partitional clustering algorithms to detect ripeness level of bananas and tomatoes [1]. Fatma Susilawati mohamad and Azizah Abdul Manaf used histogram matching method to find ripeness of oil palm fruits. Scanlon proposed an approach to quantify colour of potato chips [3] and Choi used Colour image analysis to detect tomatoes maturity rate [8]. In 2011, Chiunhsiun Lin, et al proposed a method to check the maturity rate of tomatoes [2].

This paper is organized as follows; section II gives description about color image segmentation. Section III defines the RGB color space. Section IV depicts the outline of technique used to find the ripeness level of fruits and section V gives experimental results. Finally, some conclusions are represented in section VI.

II. COLOR IMAGE SEGMENTATION

Color image segmentation is a process of partitioning an image into meaningful regions with respect to colors. The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Firstly monochromatic images are used to perform image segmentation operation. But in these images intensity is the only information source. It has been said that human eye can recognize thousands of color shades and intensities but in case of gray scale images it can recognize only two dozens of gray shades. So segmentation of color images has been more preferred as compare to gray scale images. It is easy to segment image on basis of colors as compare to texture, shape and size. The main reason behind this color images provide more information, more capacity and high speed to process the information [1].With the development of technology, the research work in field of color image segmentation also increases. So according to techniques applied, this research work can be categorized as: 1. Image domain based, 2. Physics based, 3.Feature space based [9].Image domain based techniques are mainly used to provide image segmentation for closely connected regions [9]. In first type of approach we try to develop regions until an homogeneity constrain is held. This development of region can be done with neighboring pixels or by merging and splitting regions. In second type of approach we detect image discontinuities, so regions are limited by the pixels containing the color boundary. The resulting segmentation may or may not be homogeneous region in feature space. Physics based techniques use models based on the physical properties of light and of objects in an image. Such techniques attempt to find actual material boundaries and ignore boundaries due to illumination changes in an image [10]. In feature space based techniques algorithms are concerned with the presence of collective massive pixels within a features spectral space [9]. After determined relevant color classes, the segmentation have been done according to regions of pixels corresponding to same color class. But the final segmentation result is not so visible because of some limitations. These methods don't use its neighborhood pixel information.

III. RGB COLOR SPACE

RGB color space is the most common color representation that has been adopted in large amount input/output devices for color information or RGB color images.

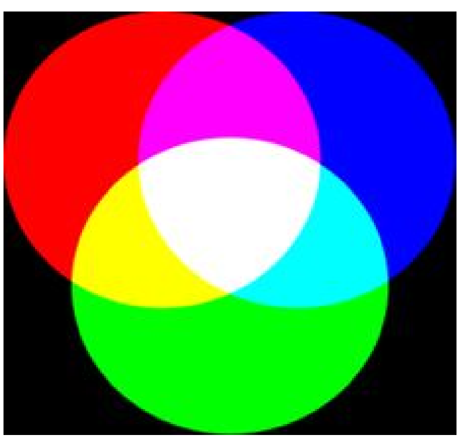


Figure 1

An RGB color image is stored as a three-dimensional (m-by-n-by-3) array of integers in the range [0, 255] (uint8) or [0, 65535] (uint16). According to [11], it is an M*N*3 array of pixels, where each color pixel is a triplet corresponding to the red, green and blue components of an RGB image at a specific spatial location. An RGB image may be viewed as a "stack" of three gray- scale images that, when fed into the red, green and blue inputs of a color monitor, produce a color image on the screen.

The vertices of the cube shows the primary colors- red, green and blue and secondary colors – cyan, magenta and yellow. Secondary colors are formed by the combination of primary colors. Whenever we change the intensities of the primary colors we will get a new color as shown in figure 1.

To define the number of bits in any pixel of RGB image we use word "bit depth". Bit depth is the number of bits used to represent the pixel values of the component image. For example, if each component image is an 8- bit image, the corresponding RGB image is said to be 24 bit depth. To determine the number of color in RGB we use formula:

Number of colors in an RGB image = (2b)3

Where b is the number of bits in each component image. We can represent RGB color space by using RGB color cube as shown in figure 2.

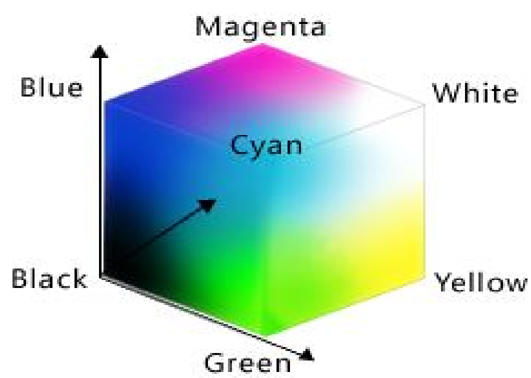


Figure 2 RGB color cube

IV. TECHNIQUE USED TO CHECK RIPENESS LEVEL OF FRUIT

As we know, each color is made up of combination of three primary colors- red, green and blue. To represent a color in colored image each pixel has a fixed value of red, green and blue components. In RGB color space pixel p (i) is defined by ordered triplet of red, green and blue coordinates (r(i), g(i), b(i)),which represents the intensities of red , green and blue light respectively. The intensity value varies from 0 to 255.

In proposed scheme we are trying to detect ripeness level of apple by color image segmentation. Firstly we studied pixel values of different shades of red color for apple. From the observation we find that there is fixed range of r, g and b values which corresponds to same color shades of apple. Therefore, we set the three basic rules according to which we detect the ripeness level of individual fruit.

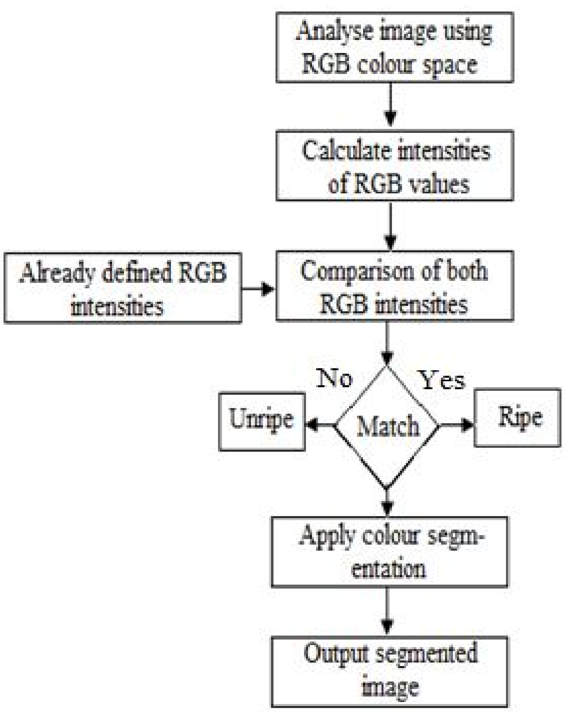
1. r (i) > a: means that the primary color component (red) should be larger than a .

2. g (i) < b: means that the primary color component (red) should be larger than b.

3. b (i) < c: means that the primary color component (red) should be larger than c.

The value of “a”=20,”b”=70 and "c" = 72 are experimentally found to be satisfactory in the color segmentation. The first rule means that the value of r (i)- the intensity of red light should be larger than a. The second rule means that the value of g (i) **-** the intensity of green light should be less than b. The third rule means that the value of b (i) **-** the intensity of blue light should be less than c. In other words, if the pixels of the input image satisfy the above 3 rules, then the pixels are regarded as ripe fruit. You can give (a, b, c) different values for distinct applications and the output will be altered.

In flow chart 1.1 we describe the whole procedure in which way our sheme will perform. In first step we give input RGB image of fruit.



Flow chart 1.1

Secondly we calculate the intensity level of RGB values of input image. In third step, we compare our input image’s RGB intensities with already defined values.In already defined values, we will set some threshold values of red, green and blue colors. Here We get two types of outputs – Yes and No . If Yes, that part of fruit matches our predefined threshold values is ripe part. If No, that part is not matching with threshold values called underipe part.Hence on applying color segmentation the color intensities which will satisfy our predefined input values show the ripe part of fruit else will show unripe part. Atlast we will get our segmented image.

V. EXPERIMENTAL RESULTS

Figure 3(a) illustrated the original input RGB color image of ripe fruit, Figure 3 (b) exemplified the result of ripeness level of apple using color image segmentation — only the red color will be remained. Most of the part of ripe fruit matches with the already set threshold value.

Figure 3(a) Original input RGB image of ripe fruit

Figure 3(b) ripeness level of ripe fruit

Figure 4(a) overripe fruit

Figure 4(b) Ripeness level of overripe fruit

Figure 5(a) unripe fruit

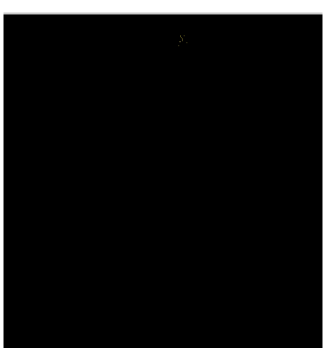


Figure 5(b) Ripeness level of unripe fruit

Figure 4 (a) is the original input RGB color image of overripe fruit, Figure 4 (b) exhibited the result ripeness level of overripe fruit using color image segmentation. Only some part of image has ripe part. The rest part is not good to eat. Similarly in Figure 5(a), we take an unripe fruit of green color. In Figure 5(b), the resulted image is black i.e. no any color matches with our threshold values. Clearly from results we can easily detect the ripeness level of fruit without touching or feeling it.

CONCLUSION

In this paper, a novel color segmentation algorithm that can examine the ripeness level of apple based upon color segmentation is proposed. Our system is based on the derived inherent properties of RGB color space. The proposed color segmentation algorithm is very effective in examining the maturity evaluation of fruits and vegetables. This approach can operate directly on RGB color space without the need of color space transformation. Moreover, the system can be applied to different applications without difficulty; by merely changing the values of the parameters (a, b, c).

REFERENCES

[1] B. Ojeda-Magana, R. Ruelas, J. Quintanilla-Domingvez and D. Andina, “Colour Imge Segmentation by Partitional Clustering Algorithns”, IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society, 7-10 Nov., Glendale, AZ, pp.2828 -2833, 2010.

[2] Chiunheium Lin, ching- Hung Su,Hsuan Shu Huang and Kuo-Chin Fan, “Colour Image Segmentation Using Relative values of RGB in Various Illumination, Circumstances”, International Journal of computers, vol.5(2),pp.252-261, 2011.

[3] Scanlon, M. G., “Computerized video image analysis to quantify colour of potato chips”, American Potato Journal, Vol. 71(11), pp. 717-733, 1994.

[4] F. Mendoza, J .M. Aguilera, “Application of Image Analysis for Classification of Ripening Bananas”, Journal of Food Science. 69(9): pp.E471-E477, 2004.

[5] J. Blasco, N. Aleixos, E., “Molto. Machine Vision System for Automatic Quality Grading of Fruit”, Biosystems Engineering, vol. 85 (4):pp.415-423, 2003.

[6] Ling Mei Chang, Rodney Tan, Gilbert Thio, “Design Of Visual- based Colour Classification System”, Research papers, JASA 2,pp. 30-33, January 2007.

[7] Zhi- yuan Wen, Lu-ming Shen, Hui-png jing, Kui Fang, “Colour and Shape Grading Of Citrus Fruit Based on Machine Vision with Fractel Dimention”, 3rd International Congress on Image and Signal Processing (CISP2010),pp. 898-903,IEEE 2010.

[8] Choi K. H., “Tomato maturity evaluation using colour image analysis”, Transactions of the ASAE, Vol. 38(1), pp. 171-176, 1995.

[9] S. Romani, P. Sobrevilla, E. Montseny, “Obtaining the Relevant Colours of an image through Stability- based Fuzzy Colour Histograms”, The IEEE International Conference On Fuzzy Systems,pp. 914-919,2003.

[10]Allan Hanbury, “Physics based Segmentation of Colour Images in Spherical Coordinates”, Technical Report, pp. 1-2,23rd july, 2003.

[11]Rafael C.Gonzalez, Richard E. Woods and Steven L. Eddins, ”digital image processing using MATLAB”,pp.208-211.

[12]Wikipedia, the free encyclopedia. Color image segmentation, 2010

Perception of 3D Effect Using Integral Imaging Technology

Amanpreet Kaur

ECE Dept

aman_ghuman16@yahoo.com; amnghuman@gmail.com

***Abstract* — This paper describes an integral imaging technology that displays 3-D image with continuous viewing points, full parallax and full colour view to the observers and there is no need to wear special glasses to perceive the 3D effect. This technology is based on the integral photography. It also includes performance factors such as image quality enhancement, resolution and depth of field improvement. In this paper,new approaches in integral imaging technology are also discussed.**

Keywords*— *Three-dimensional display; integral imaging;
integral photography

I. INTRODUCTION

Integral imaging is a three-dimensional imaging technique that works with incoherent light and provides with auto-stereoscopic images that can be observed without the help of any additional devices. It is the natural consequence of applying the modern technology to the old concept of integral photography which was proposed by Lipmann in the beginning of 20th century[1].This technique is based on the reversible principle for the light rays. Integral imaging allows for the capture of both light intensity and direction as the single large aperture is replaced by a multitude of smaller apertures. Because integral imaging uses 2D lens array or pinhole array to define the viewing directions of correspondent elemental images, it is distinguished from stereoscopic or autostereoscopic displays. The technique of integral imaging has many features as follows.

(1) It provides full parallax (vertical as well as horizontal parallax)

(2) It provides quasi-continuous viewpoints, full color and real time images.

(3) It does not require any special apparatus.

(4) Multiple observers can see integrated 3D images within viewing angle.[2]

In general integral imaging system as shown in Fig. 1 consists of two processes: pickup and reconstruction. In the pickup process, a lenslet array is used to capture the 3D object. Each of the lenslets provides different perspective views of the 3D object, which results in a collection of demagnified 2D images, known as an elemental image array (EIA). To record the EIA, a 2D image sensor such as a charge coupled device (CCD) sensor is used. In order to reconstruct the 3D image, rays are reversely propagated through the EIA and a similar lenslet array is used in the pickup process. On the other hand, in reconstruction part, there are two kinds of integral imaging reconstruction techniques. One is based on optical integral imaging reconstruction (OIIR) and the based on computational integral imaging reconstruction (CIIR)[3].

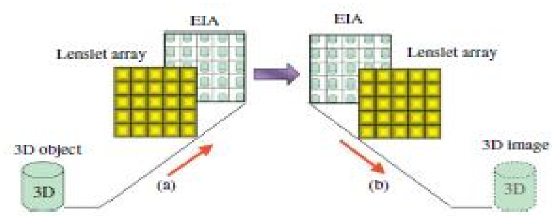


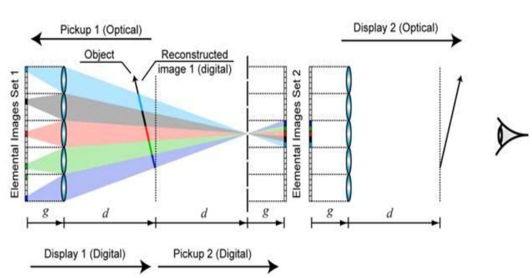
Fig.1:A general integral imaging system(a) pick up(b) reconstruction

II. PERFORMANCE FACTORS OF INTEGRAL IMAGING

One problem encountered with integral imaging for 3D display applications is the pseudoscopic (or depth reversed) nature of the displayed images when the captured elemental images do not receive pixel pre-processing. Therefore, the additional procedures are needed to convert the 3D image from pseudoscopic to orthoscopic form. Therefore ,the additional procedures are needed to convert the 3D image from pseudoscopic to orthoscopic form [3]. Most of these methods might degrade the image quality of the 3D image after the conversion process due to the use of optical devices [4,5].

1) Image quality enhancement

To overcome the problem of poor quality image, a Computat-ional reconstruction method called smart pixelmapping(SPM) was reported which provides real,undistorted and orthoscopic integral images.



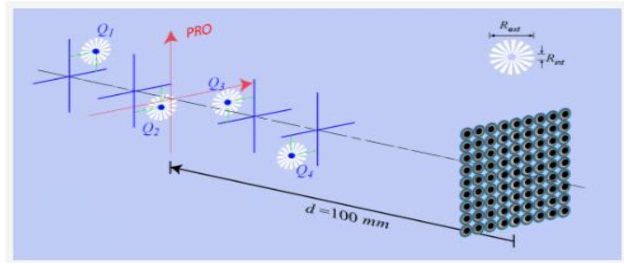
The principle of CIIR, which is based on the pinhole-array model, is that 3D images are digitally reconstructed at the required output planes by superposition of all of the inversely mapped and magnified elemental images. The magnification factor increases in proportion to the distance of required output plane. When 3D object is recorded through a square-shaped lens array in the pickup process, the reconstructed images of CIIR have intensity irregularities with the grid noise. Thus, the visual quality of reconstructed image gets degraded. To solve this problem, behavior of interpolation methods in the CIIR to reduce the intensity irregularity along the distance from the virtual pinhole array and the grid noise in the reconstructed images. The conventional CIIR method employs image magnification before superimposing elemental images. This image magnification is considered to be the zero-order interpolation method, i.e., the pixel replication method. There exist numerous methods of interpolation in the image processing literature. Among image interpolation methods several methods are applied to many applications because of a good tradeoff between the image quality and the implementation cost; the zero-order interpolation method is the simplest method and it is applied to real-time applications although it has the blocking artifact. The linear interpolation method is simple and provides moderate image quality. Thus it is applied to many applications such as image compression, image display, computer graphics, and so on. But it still has the serious blurring problem. To overcome the blurring problem and to obtain a substantial gain in image quality, higher order interpolation methods were developed such as the Keys' cubic convolution interpolation (CCI) method. Although the CCI method provides better image quality than the two previous methods, the computational cost is much higher than those methods. Thus the CCI method is limited to applying to non-real-time applications or to high performance applications[6,7].

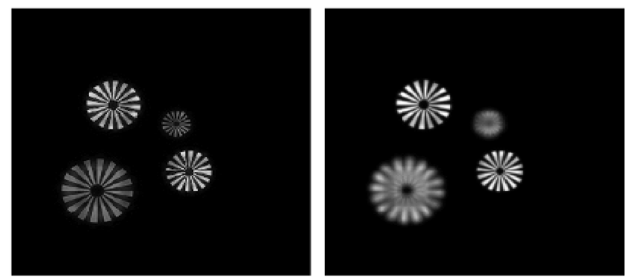
2) Elimination of Faceting effect

The another problem (the faceting effect) results from the from vigneting occurred when the observer looks through the microlenses. What happens is that the observer sees only a small portion of the reconstructed scene through any microlens. The global effect is that the observer sees a kind a puzzle in which any piece is the facet seen through the corresponding microlens. Typical faceting effect occurs when the filling factor of the microlenses is smaller than one. In such case, any gap between microlenses inherently results in empty space between facets in the retinal image.To overcome the problem the use of amplitude modulation masks for changing the transmission properties of the microlenses, is highly non recommendable in the display of Integral Imaging[12].

3) Resolution Improvement

The lateral resolution of captured elemental images is mainly determined by sensor constraints and also by diffraction effects.To improve the lateral resolution of Integral Imaging system in the capture stage of hybrid technique is applied that is based on the use of binary amplitude modulation during the capture process and Wiener filtering of the captured elemental images by computer processing. In this method, the elemental images of four spoke targets for the pickup architecture shown down in the left and in the center, the reconstructed images is obtained with the hybrid method and in the right the images reconstructed with the conventional setup[8,12].





4) Integral imaging with improved depth of field by use of amplitude-modulated microlens arrays.

One of the main challenges in three-dimensional integral imaging is its limited depth of field. Such a limitation is imposed by diffraction, among other factors. The easiest way to improve the depth of field is by reducing the numerical aperture of the microlenses. However, such an improvement is obtained at the expense of an important deterioration in the spatial resolution.however new technique is applied in the context of integral imaging, for improving the depth of field with no deterioration of the spatial resolution. The technique, based on amplitude modulation of the array of phase elements, can substantially improve the figure of merit of the product of depth of the focus and the squared resolution[9].

III. New approaches in integral imaging display

Integral imaging has many issues to be enhanced in viewing parameters such as viewing resolution, viewing angle and image depth, in spite of efforts of many enthusiastic researchers. For higher viewing resolution, many display devices such as LCDs, SLMs or projectors are needed; however it is not a practical option. In regards to viewing angle, it is affected by lens parameters in integral imaging. Although the method of using curved lens array and display devices has been proposed, it is very hard to apply in industrial fields due to display device thickness. Image depth is also an important feature in integral imaging. Because of the presence of CDP in integral imaging, image depth only can be expressed within certain range around CDP. With these features, integral imaging does not have occupied mass market of display industries yet. In addition, the constraints on the limited number of views and the insufficiencies of proper content have been obstructions to capturing the display market. Recently as a part of solution for these features, new technologies in integral imaging have been researched – integral floating display and 2D/3D convertible integral imaging display. In the following subsections, these techniques are described as:

a)Integral floating display

Integral floating display was originally proposed as an effort for enhancement of image depth in integral imaging. Image floating technique uses a large convex lens or a concave mirror to display the image of real object to the observer. Because the 3D image of real object can be located in close distance to the observer, the floating display method can provide an impressive feel of depth. Integral floating display is a fused 3D display scheme which combines image floating display and integral imaging. The integral imaging is adopted as a display part in electro-floating.

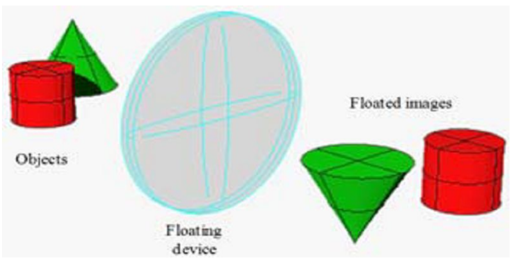
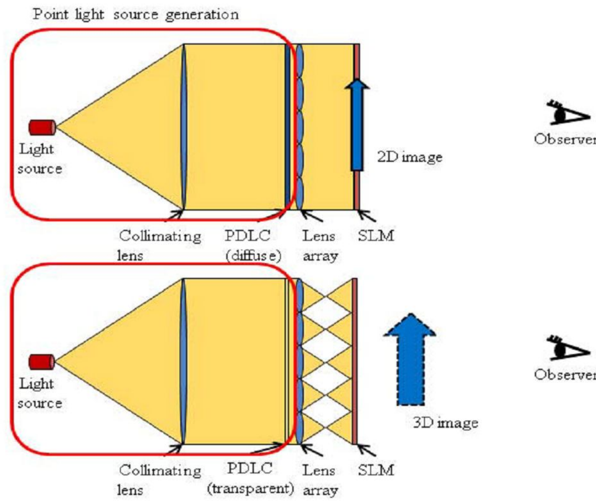


Fig. shows a schematic of floating display

As shown in Fig., the floated image is converted by 180° due to the floating lens. The floating lens can be changed into a concave mirror because the concave mirror is optically equivalent to the convex lens. In the floating display, real objects or display device can be adopted as an image source. When the display device is used for image source in floating display, it is called the electro-floating display because image source can be changed electrically[2,10].

b) 2D/3D convertible integral imaging methods

Many researchers of 3D display systems pointed out that there should be a stepping stone between 2D display era and 3D display era. There is an enormous gap between the amount of information for 2D display and 3D display. The main reason for this is that the resolution of 3D display is worse than that of 2D display. Various types of 2D/3D convertible integral imaging methods have been proposed recently.In the original work of the point light source integral imaging method, collimated light through a lens is used for producing a point light source array[11]. The key of 2D/3D convertible integral imaging method is a point light source array, although there exist other methods of using two display panels as well. 2D/3D convertible operation can be controlled by inscribing a diffusing material such as PDLC in front of lens array.



IV. CONCLUSION

The 3D display hardware technique based on integral imaging has been vigorously researched. Integral imaging is a promising way for realization of 3D display. It can display real-time 3D movies with full colour and parallax with lens array or pinhole array and a commercial 2D display.This technology also improves the quality of the 3D image that can be viewed without any additional devices.

REFERENCES

[1] Lippmann G. La photographic intergrale. C R Acad Sci 1908;146:446–51.

[2] YoungminKim, Hong,Byoungho Lee Recent Researchers based on Integral Imaging Display Method, 3D Research Center and Springer 2010 .

[3] Shin D-H, Kim E-S. Computational integral imaging reconstruction of 3D object using a depth conversion technique. J. Opt. Soc. Korea 2008;12:131–5.

[4] Dong-Hak Shin, Byung-GookLee , Eun-SooKim Modified smart pixel mapping method for displaying orthoscopic 3D images in integral imaging,2009,Elsevier Ltd.,sciencedirect.

[5] Jang J-S, Javidi B. Formation of orthoscopic three-dimensional real images in direct pickup one-step integral imaging. Opt Eng 2003;42:1869–70.

[6] Dong-Hak Shin and Hoon Yoo **,**Image quality enhancement in 3D computational integral imaging by use of interpolation methods,2007 ,Optical Society of America.

[7] W. K. Pratt, Digital Image Processing, (New York: Wiley, 1991).

[8] J.-H. Park, J. Kim, Y. Kim, and B. Lee (2005) Resolution enhanced three-dimensional/two-dimension convertible display based on integral imaging, *Opt. Express*. **13**: 1875-1888.

[9] Manuel Martı´nez-Corral, Bahram Javidi, Rau´ l Martı´nez-Cuenca, and Genaro Saavedra , Integral imaging with improved depth of field by use of amplitude-modulated microlens arrays, 2004,OSA.

[10]S.-W. Min, M. Hahn, J. Kim, and B. Lee (2005) Three-dimensional electro-floating display system using an integral imaging method, *Opt. Express*. **13**: 4358-4369

[11]J.-H. Park, H.-R. Kim, Y. Kim, J. Kim, J. Hong, S.-D. Lee,and B. Lee (2004) Depth-enhanced three-dimensional two-dimensional convertible display based on modified integral imaging, *Opt. Lett*. **29**: 2734-2736

[12] http://www.uv.es/imaging3/lineas/InI.htm

Image Enhancement: In Frequency Domain

Neelakshi Gupta[1]
[1]Assistant Professor, Department of Electronics & Communication Engineering, ACET, Amritsar
Email: erneelakshigupta@gmail.com

Abstract-Image enhancement is the task of applying certain transformations to an input image such as to obtain a visually more pleasant, more detailed, or less noisy output image. The transformation usually requires interpretation and feedback from a human evaluator of the output result image. Therefore, image enhancement is considered a difficult task when attempting to automate the analysis process and eliminate the human intervention. An alternative image representation is based on spatial frequencies of grey value or colour variations over the image plane. This dual representation by a spectrum of different frequency components is completely equivalent to the conventional spatial representation: the direct conversion of a 2D spatial function f(x,y) into the 2D spectrum F(u,v) of spatial frequencies and the reverse conversion of the latter into a spatial representation f(x,y) are lossless, i.e. involve no loss of information. Such spectral representation sometimes simplifies image processing. This paper will discuss two dimensional image filtering in the frequency domain. The reason for doing the filtering in the frequency domain is generally because it is computationally faster to perform two 2D Fourier transforms and a filter multiply than to perform a convolution in the image (spatial) domain. This is particularly so as the filter size increases.

Index Terms - Image Enhancement, Frequency Domain, Spatial Domain

I. INTRODUCTION

The aim of image enhancement is to improve the interpretability or perception of information in images for human viewers, or to provide `better' input for other automated image processing techniques. Image enhancement techniques can be divided into two broad categories: 1. spatial domain methods, which operate directly on pixels [1], and 2. Frequency domain methods, which operate on the Fourier transform of an image. Unfortunately, there is no general theory for determining what `good' image enhancement is when it comes to human perception. If it looks good, it is good! Many different, often elementary and heuristic methods [2] are used to improve images in some sense. The problem is, of course, not well defined, as there is no objective measure for image quality. However, when image enhancement techniques are used as pre-processing tools for other image processing techniques, then quantitative measures can determine which techniques are most appropriate. In this paper, we restrict our study Frequency Domain methods.

II.FILTERING

Low pass filtering involves the elimination of the high frequency components in the image. It results in blurring of the image (and thus a reduction in sharp transitions associated with noise). Edges and sharp transitions in gray values in an image contribute significantly to high-frequency content of its Fourier transform. An ideal low pass filter would retain all the low frequency components, and eliminate all the high frequency components. Hence, an image can be smoothed in the Frequency domain by attenuating the high-frequency content of its Fourier transform. However, ideal filters suffer from two problems: blurring and ringing. These problems are caused by the shape of the associated spatial domain filter, which has a large number of undulations.

An ideal low pass filter with cutoff frequency r_0:

$$H(u,v)=\begin{cases}1, \text{ if } \sqrt{u^2+v^2}\le r_0\\ 0, \text{ if } \sqrt{u^2+v^2}> r_0\end{cases}$$

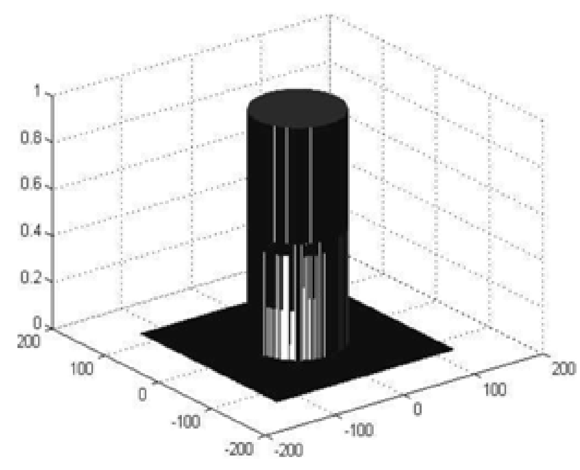


Fig 1: Ideal LPF with $r_0 = 57$

Note that the origin (0, 0) is at the center and not the corner of the image . The abrupt transition from 1 to 0 of the transfer function $H(u,v)$ cannot be realized in practice, using electronic components. However, it can be simulated on a computer.

a) Original Image

b) LPF Image, r_0 = 57

c) LPF Image, r_0 = 26

Fig 2: LPF image with different r_0

Notice the severe ringing effect in the blurred images, which is a characteristic of ideal filters. It is due to the discontinuity in the filter transfer function. The cutoff frequency r_0 of the ideal LPF determines the amount of frequency components passed by the filter. Smaller the value of r_0, more the number of image components eliminated by the filter. In general, the value of r_0 is chosen such that most components of interest are passed through, while most components not of interest are eliminated. Usually, this is a set of conflicting requirements. A useful way to establish a set of standard cut-off frequencies is to compute circles which enclose a specified fraction of the total image power. Ideal LPF presents a Sin function in the spatial domain. The main lobe (centre component) is primarily responsible for blurring. The concentric components are responsible for ringing. The radius of the main lobe is inversely proportional to the cutoff frequency. The density of side lobes is proportional to the cutoff frequency. Due to the multiple peaks of the ideal filter in the spatial domain, the filtered image produces ringing along intensity edges in the spatial domain.

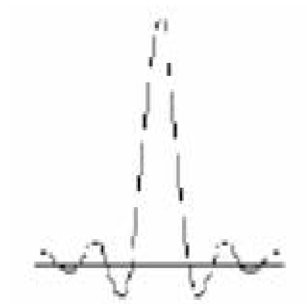


Fig 3: Transfer Function of Ideal Low Pass Filters

Butterworth Low Pass Filters

The Butterworth filter is a type of signal processing filter designed to have as flat a frequency response as possible in the passband. It is also referred to as a maximally flat magnitude filter. A two-dimensional Butterworth lowpass filter has transfer function:

$$H(u,v) = \frac{1}{1 + \left[\frac{\sqrt{u^2 + v^2}}{r_0}\right]^{2n}}$$

n: filter order, r_0: cutoff frequency

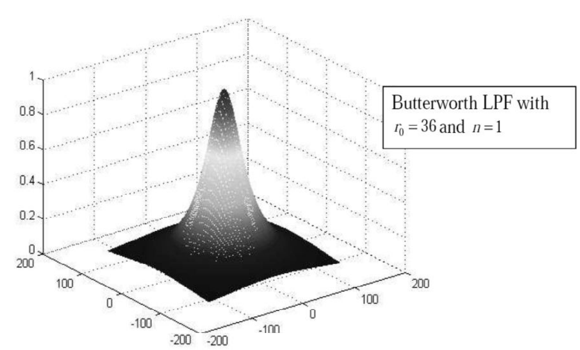


Fig 4: BLPF with r_0 = 36 and n=1

BLPF does not have a sharp discontinuity like ILPF. Frequency response does not have a sharp transition as in the ideal LPF.This is more appropriate for image smoothing than the ideal LPF, since this not introduce ringing.

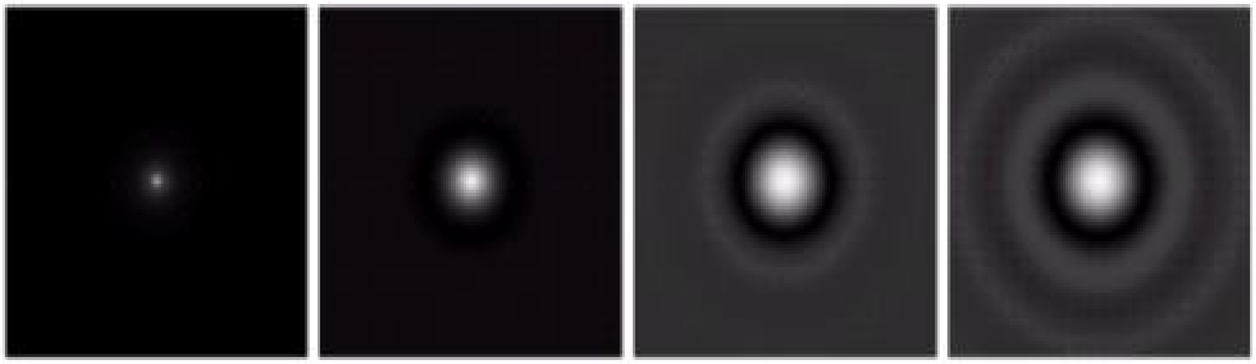


Fig 5:a-d represents BLPS of order 1, 2, 5 and 20 and corresponding gray-level profiles through the center of the filters. It has been observed that ringing increases as a function of filter order. Ringing observed in filters of higher order but not when n = 1 or 2.In this all filters have a cutoff frequency of 5.

a	b	c
d	e	f

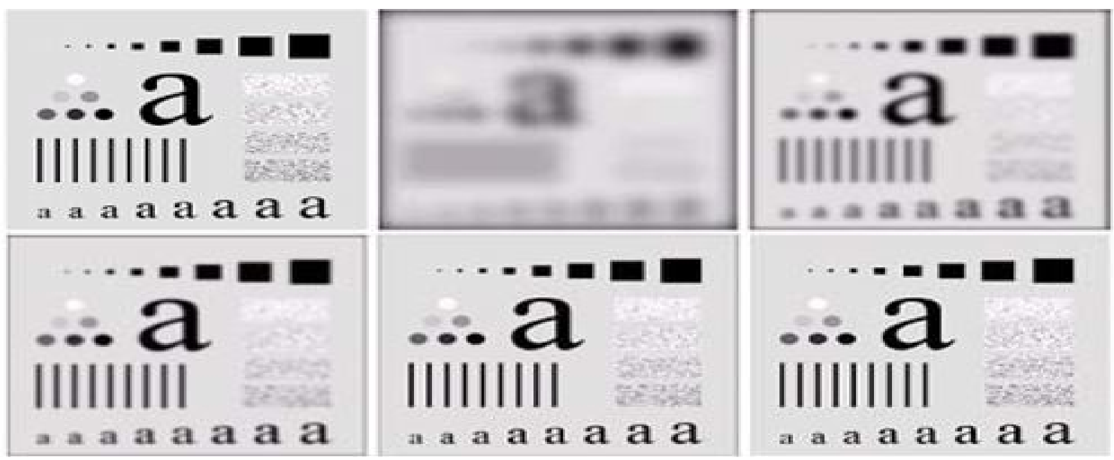


Fig:6 a)Original Image (b)-(f)Results of filtering with GLPF with cut off frequencies set at radii of 5,15,30,8, 230.

High Pass Filtering

Edges and sharp transitions in gray values in an image contribute significantly to high-frequency content of its Fourier transform. Negative images [3] are useful for enhancing white or grey detail embedded in dark regions of an image.Regions of relatively uniform gray values in an image contribute to low-frequency content of its Fourier transform. Hence, image sharpening in the Frequency domain can be done by attenuating the low-frequency content of its Fourier transform. This would be a high pass filter! For simplicity, it is considered only those filters that are real and radially symmetric.

An ideal high pass filter with cutoff frequency r_0:

$$H(u,v) = \begin{cases} 0, & \text{if } \sqrt{u^2+v^2} \le r_0 \\ 1, & \text{if } \sqrt{u^2+v^2} > r_0 \end{cases}$$

Note that the origin (0, 0) is at the center and not the corner of the image. The abrupt transition from 1 to 0 of the transfer function $H(u,v)$ cannot be realized in practice, using electronic components. However, it can be simulated on a computer.

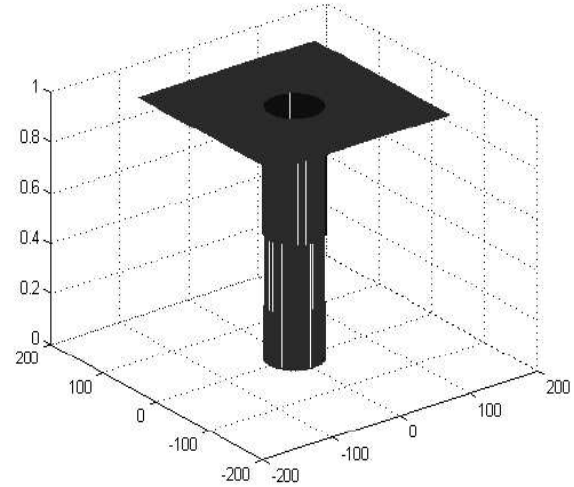
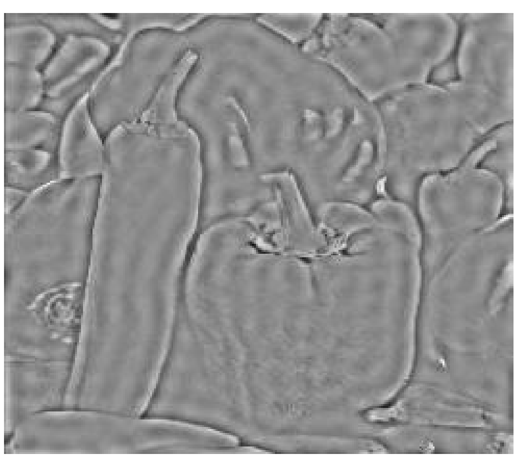
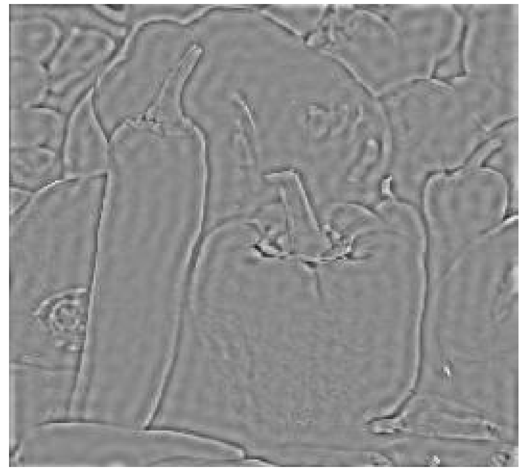


Fig 7: Ideal HPF with $r_0 = 57$

a) Original Image



b) HPF Image, $r_0 = 18$



c) HPF Image, $r_0 = 26$

Fig 8: HPF image with different r_0

Notice the severe ringing effect in the output images, which is a characteristic of ideal filters. It is due to the discontinuity in the filter transfer function.

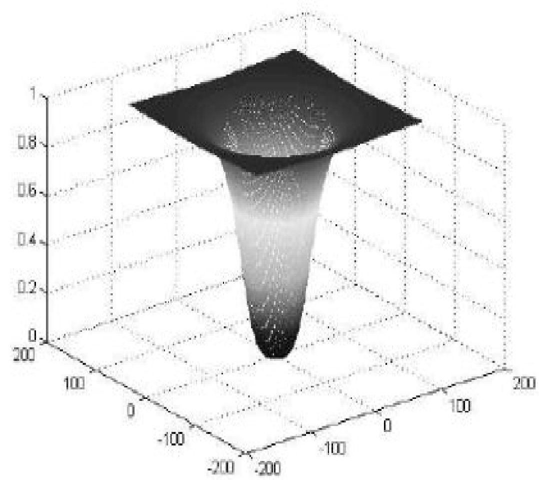


Fig 9: BHPF with r_o= 47 and 2

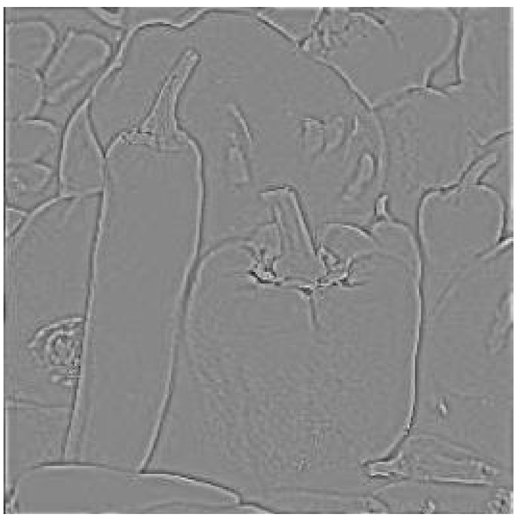
Butterworth High Pass Filter

A two-dimensional Butterworth high pass filter has transfer function:

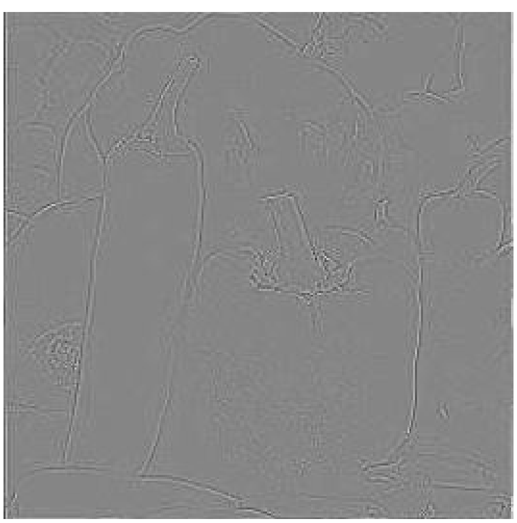
$$H(u, v) = \frac{1}{1 + \left[\frac{r_0}{\sqrt{u^2 + v^2}} \right]^{2n}}$$

Frequency response does not have a sharp transition as in the ideal HPF. This is more appropriate for image sharpening than the ideal HPF, since this not introduce ringing.

a) Original Image



b) HPF Image, r_0 = 36



c) HPF Image, r_0 = 81

Fig 10: BHPF image with different r_0

Gaussian High Pass Filters

The form of a Gaussian low pass filter in two-dimensions is given by

$$H(u,v) = 1 - e^{-D^2(u,v)/2\sigma^2}$$

where D (u,v) $=\sqrt{u}2 + v2$ is the distance from the origin in the frequency plane. The parameter σ measures the spread or dispersion of the Gaussian curve. Larger the value of, larger the cutoff frequency and more severe the filtering.

CONCLUSION

A Frequency domain enhancement based on FFT is reviewed. From this discussion, it is found that there are maximum variations between original and enhanced images; and also the increased number of terminations and decreased number of bifurcations are due to the un-smoothing and noisiness. In High Pass filters and low pass filters, Gaussian Filters are proven to be best filters because they give the smoothen images.

REFERENCES

[1] Bhabatosh Chanda and Dwijest Dutta Majumder, 2002, Digital Image Processing and Analysis.
[2] R.W.Jr. Weeks,(1996). Fundamental of Electronic Image Processing. Bellingham: SPIE Press
[3] A. K. Jain, *Fundamentals of Digital Image Processing.* Englewood Cliffs, NJ: Prentice Hall, 1989.
[4] Rafael C. Gonzalez and Richard E. Woods. Digital Image processing. Addison-Wesley Publishing Company, 1992, chapter 4.
[5] William J. Mitchell. The Recon_gured Eye: Visual Truth in the Post-Photographic Era. MIT Press 1992.

Wireless Optical Communication Technology for Short Range/Indoor Environments

Jaswinder Kaur[1], Anu Sheetal[2]
[1]*ABIET, Pathankot*
[2]*GNDU, Regional Campus,Gurdaspur*
Email: jaswindergoraya@gmail.com, nripanu@yahoo.co.in

***Abstract*— It is commonly agreed that the next generation of wireless communication systems, usually referred to as 4G systems, will not be based on a single access technique but it will encompass a number of different complementary access technologies. The ultimate goal is to provide ubiquitous connectivity, integrating seamlessly operations in most common scenarios, ranging from fixed and low-mobility indoor environments in one extreme to high-mobility cellular systems in the other extreme. Surprisingly, perhaps the largest installed base of short-range wireless communications links are optical, rather than RF, however. Indeed, 'point and shoot' links corresponding to the Infra-Red Data Association (IRDA) standard are installed in 100 million devices a year, mainly digital cameras and telephones. In this paper we argue that optical wireless communications (OW) has a part to play in the wider 4G vision. An introduction to OW is presented, together with scenarios where optical links can enhance the performance of wireless networks.**

***IndexTerms*—Optical-wireless communications, wireless communications**

I. INTRODUCTION AND MOTIVATION

As the third generation mobile communication system (3G) is being deployed, manufacturers and scientific community are increasingly turning their research interests toward future wireless communication systems. It is commonly agreed that the next generation of wireless communication systems, usually referred to as 4G systems, will not be based on a single access technique but it will encompass a number of different complementary access technologies. Future systems will not only connect users and their personal equipment but also access to independent (stand-alone) equipment will be provided. Ultimately one would expect that everybody and will be wirelessly connected. This vision places short-range communications in a position of preponderance, as one could argue that most of the wireless links in future wireless communication networks will be established over relatively short distances. In addition, a significant proportion of these links will be characterized by high data throughputs. Probably the largest portion of practical applications of short-range communications take the form of WLAN, WPAN and WBAN (Wireless Local, Personal and Body Area Networks), covering ranges from a few tens of meters down to sub-meter communications.

In the context of short-range communications, two techniques have received increasingly attention in the last years, namely multicarrier (MC) and Ultra Wideband (UWB). These fundamental technologies for the physical layer have been extensively studied in the literature, for a comprehensive introduction and initial pointers. Application of these techniques in short-range environments will be considered in detail within the framework of WWRF.

In this paper we argue that optical wireless communications (OW) has a part to play in the wider 4G vision. The optical wireless channel has THz of unregulated bandwidth, and characteristics that are distinct from that of radio. It should be noted that our aim is to show that the optical channel has complementary characteristics and can, in certain situations, add to, but certainly not replace the capability of a RF 4G wireless system. Together these media might provide a broad spectrum of channel characteristics and capabilities that radio alone would find it difficult to meet.

The aims of this paper are to;

(i) Introduce OW and the components and systems used

(ii) Summarise the state of the art, and rich research community that exists

(iii) Compare the characteristics of OW with radio

(iv) Identify particular areas where OW can contribute to the 4G vision, and areas of future research

Optical Wireless Communications as a complementary Technology for Short-Range Communications

In this section an overview of Optical Wireless Communication systems is presented. The main emphasis in this paper is put on OW for indoor environments. Another important approach to OW, free-space optics

(FSO), a point-to-point optical connection supporting very high rates in outdoor environments, will not be considered in this paper. We start with a brief introduction and classification of OW systems and then continue with different engineering aspects, including transmitters, receivers, the optical channel and other related issues. Comparisons to conventional radio systems are presented to give the reader a broader perspective of the possible baseband technologies. An up-to-date account of different modulation techniques, practical systems and standards related to OW as well as future research issues complete the paper.

II.BASIC SYSTEM CONFIGURATION

Figure 1 shows a number of different OW configurations. There are two basic configurations; communications channels either use diffuse paths (Figure 1 (a)) or Line Of Sight (LOS) paths (Figure 1(b)) between transmitter and receiver. In a diffuse system an undirected source (usually Lambertian) illuminates the coverage space, much as it would be illuminated with artificial lighting. The high reflectivity of normal building surfaces then scatters the light to create an optical 'ether'. A receiver within the coverage space can detect this radiation, which is modulated in order to provide data transmission. Diffuse systems are robust to blocking and do not require that transmitter and receiver are aligned, as many paths exist from transmitter to receiver. However, multipath interference at the receiver can cause Inter Symbol Interference (ISI) and the path loss for most systems is high.

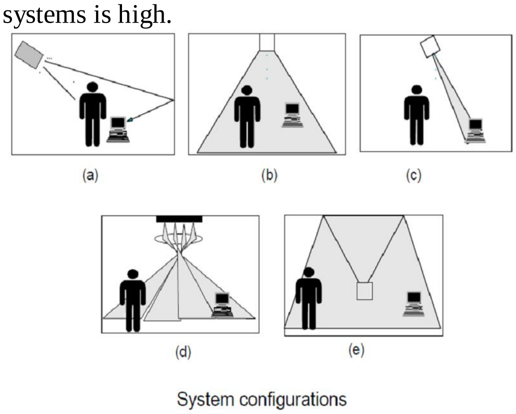


Fig.1(a) Optical wireless configurations.

(a) Diffuse system.

(b) Wide LOS system.

(c) Narrow LOS system with tracking.

(d)Narrow LOS system using multiple beams to obtain coverage.

(e)Quasi diffuse system

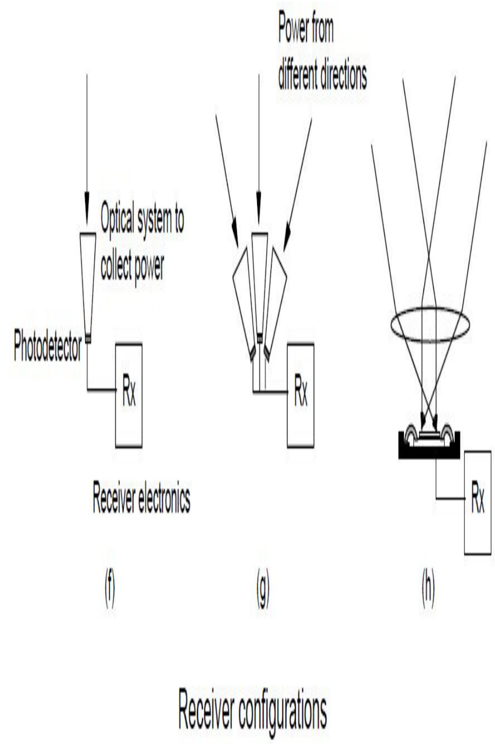


Fig. 1(b)

(f) Receiver configuration: single channel receiver.

(g) Receiver configuration: angle diversity receiver.

(h) Receiver configuration: imaging diversity receiver.

The alternative approach is to use directed Line of Sight paths between transmitter and receiver. Wide LOS systems such as that shown in Figure 1 (b) use ceiling mounted transmitters that illuminate the coverage area, but minimize reflections from walls, ensuring that a strong LOS path exists. The wide beam ensures coverage. As the beams are narrowed Path loss reduces and the allowed bit rate increases, albeit at the cost of coverage. Narrow beam systems therefore either require tracking to allow user mobility (Figure 1 (c)), or some sort of cellular architecture to allow multiple narrow beams to be used (Figure 1 (d)). A third class of system also exists; quasi diffuse systems minimize the number of multipaths by limiting the surface reflections, but allow robust coverage by directing radiation to a number of surfaces so that a suitable receiver may select a path from one surface only (Figure 1 (e)).

III.THE OPTICAL CHANNEL

LOS optical channels are subject to path loss, and this can be modeled using either ray-tracing or analytical techniques. The diffuse channel has both high path loss (>40dB typically) and is subject to multipath dispersion. Both of these characteristics are dependent on the orientation of the source and receiver within the space.

There has been extensive work on predicting the characteristics of the diffuse channel,as well as analytical models of the channel impulse results. Most building materials are found to have a high reflectivity (0.4-0.9) and they can be approximately modeled as Lambertian reflectors. Ray tracing techniques therefore allow generally good predictions of the channel response, even in the presence of chairs and other objects. Depending on the balance of LOS and diffuse paths within a space channels can be modeled as Rician or Rayleigh, with exponential impulse responses. Various measurements have also been made.Recent high-resolution data indicate that transparent 'unlimited' bandwidth diffuse channels are available in particular directions for most diffuse environments.One of the major advantages of the OW channel is that there is no coherent fading, and the channel is therefore extremely stable when compared with its RF counterpart. Even though sources are coherent the size of detectors, and the scattering environment mean that any effects are removed by the spatial integration that occurs at the receiver.

System components

Transmitter

The transmitter consists of a single, or a number of sources, and an optical element to shape the beam and also render it eyesafe if required. The main element of the transmitter is the optical source. Light Emitting Diodes (LED) and laser diodes are employed as the optical radiating element, and their transmission power is limited by eye safety regulation. Most systems use laser diodes, due to their higher modulation bandwidth and efficiency. IR LEDs are also important optical sources being considered for establishing optical links. In addition, there is a small and growing interest in using visible LEDs that would be installed in a building to provide solid-state lighting for optical communications [16]. In such cases multiplexing of the low bandwidth devices might be used to increase data rates.

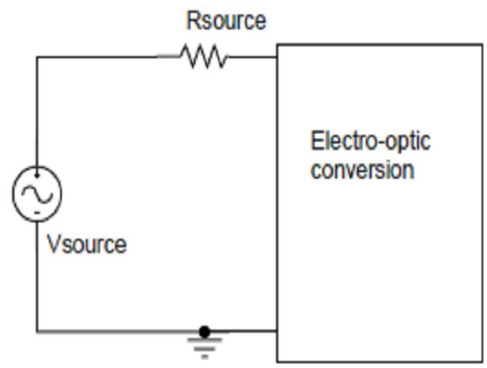


Fig. 2(a) Communications transmitter channel model

Receiver

A typical OW receiver consists of an optical system to collect and concentrate incoming radiation, an optical filter to reject ambient illumination, and a photodetector to convert radiation to photocurrent. Further amplification, filtering and data recovery are then required .

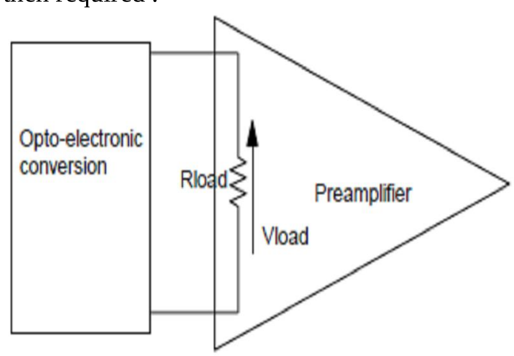


Fig. 2(b) Communications Receiver channel model

Detector/preamplifiers

The detector and preamplifier together are the main determining factor in the overall system performance. Both PIN structures and APDs have been used in single detector systems, whilst array receivers have tended to use PIN devices.

IV.MODULATION SCHEMES

Unlike in conventional RF systems, the optical channel uses intensity modulation and direct detection. The optical power output of the transmitting source is controlled according to some characteristics of the information bearing signal. The transmitted signal is thus always positive and its average amplitude is limited.

Analog and digital optical modulation is possible but, due to the intensity modulation, common modulation schemes employed in the radio frequency domain will perform differently when applied to optical systems.

The changes in optical power produced by intensity modulation are detected by direct detection, that is, a current proportional to the incident optical power is induced in the photodetector. As pointed out in two criteria should be used to evaluate the feasibility of an optical modulation system; the average optical received power required to achieve a given target BER performance and the required receiver electrical bandwidth.

Three basic modulation schemes are usually used in OW systems, namely On-Off Keying (OOK), Pulse-Position Modulation (PPM) and Subcarrier Modulation (SCM). An extensive account of these and other techniques can be found in . Other issues to take into account when considering optical modulation schemes are their robustness to multipath propagation and, in networks, their suitability to multiple access environments. PPM is very well suited to work in low signal-to-noise ratio scenarios, quite typical in optical channels due to blocking effects (shadowing) and ambient noise. However, multipath propagation induces intersymbol interference and PPM is particularly sensitive to these dispersive effects of the optical channel. Spread-spectrum modulation techniques can also be used to combat multipath distortion as well as to reduce the effects of interference, in a similar fashion as they are exploited with radio systems. Direct-sequence techniques are usually used in conjunction with optical links. Since bipolar spreading sequences cannot be used to modulate an always positive optical signal, a unipolar sequence is formed by biasing to the bipolar sequence with a fixed DC offset. This unipolar sequence preserves the correlation properties of the original sequence and it can be correlated with a bipolar sequence at the receiver. Several direct-sequence spread-spectrum approaches specially designed for optical systems have been proposed and studied, including sequence inversion keying modulation (SIK), complementary SIK (CSIK) and M-ary bi-orthogonal keying modulation (MBOK) . Sequences with low auto- and cross-correlation sidelobes are preferred in order to minimize the degrading effects of intersymbol interference.

V.OPTICAL WIRELESS VS. RADIO COMMUNICATIONS

Over the past decade the capacity of an optical fibre link has increased by several orders of magnitude, showing almost ‘Moore’s Law’ growth, largely due to the availability of optical spectrum. At the same time regulation of the RF spectrum limits available bandwidth to several orders of magnitude below this. The vision of a highly connected world is likely to require unaffordable amounts of the already scarce radio frequency-spectrum. OW occupies fully unlicensed spectrum bands, and the possibility of using unregulated and unlicensed bandwidth is one of the most attractive characteristics of OW.

Unlike radio communications, the nature of the optical radiation is such that the transmitted signal is obstructed by opaque objects, and the radiation can have high directivity using sub-millimetre scale beam shaping elements.

Another unique characteristic of wireless optical links is in the channel itself, and it is the fact that these links are not affected by multipath fading. This is because the dimensions of the receiver’s photodetector are many orders of magnitude larger that the wavelength of the optical radiation and thus, the spatial fluctuations in signal strength due to multipath are averaged over the large detector area, which acts as an integrator. For most of the cases, and as an essential advantage, optical components are small in size, low-cost and they have low power consumption. Furthermore, transceivers are relatively simple compared with their radio frequency counterparts.

There are several drawbacks, however; since IR radiation can reach the retina and eventually cause thermal damage, the maximum power that can be transmitted is limited by eye safety regulations and extra optical elements are required to render high power sources safe.

In diffuse optical communication systems, multipath propagation caused by the dispersive optical channel will introduce pulse spreading and intersymbol interference (ISI), much as would be experienced by a radio channel, although the process is due to incoherent, rather than coherent fading. Systems above 50Mb/s or so might typically require some form of equalisation.

Perhaps the major difference for the optical wireless channel is the detection process is usually incoherent, so that the detector responds linearly with power, rather than amplitude, as is the case with a radio receiver. Receiver sensitivity is therefore substantially lower than for radio channels, and therefore systems are more susceptible to path loss, especially in the case of diffuse systems. More complex receiver and transmitter structures can be used to reduce this, and the effect of noise from ambient illumination. As the detection process is incoherent there is no inherent rejection in the detection process, so filtering mechanisms must be introduced, as mentioned earlier.

The wavelength of optical radiation makes directive channels easy to implement, and system design often leads to asymmetric channels. Such directive channels are necessarily subject to blocking, which is again distinct from radio applications.

RADIO COMMUNICATIONS

There are many models of the path loss of a radio link, depending on environment, and on link distance. In this case a simple set of limiting cases is considered.

If the transceivers lie in each others far-field, so that Fraunhofer diffraction can be assumed, and a line of sight exists between transmitter and receiver the link loss can be estimated using the standard Friis' equation. In most indoor environments the antenna will be approximately isotropic, and have transmitter and receiver gains of unity. In this case

the link loss *Llink* can be approximated by;

$$L_{link} = \left(\frac{\lambda}{4\pi}\right)^2 \frac{1}{r^2}$$

where r is the link distance andλ is the wavelength of the radiation. The minimum link distance at which this occurs is the Fraunhofer distance *fd* and can be estimated as;

$$d_f = \frac{2D^2}{\lambda}$$

where is the largest dimension of the antenna. *D*

In a real environment the situation is more complicated however. At distances greater than a reference distance *dref* from the antenna and an $r^{-\gamma}$ loss beyond this.

The link loss then becomes

$$L_{link} = \left(\frac{\lambda}{4\pi}\right)^2 \frac{1}{r^2} \qquad d_f \le r \le d_{ref} \quad (3)$$

And

$$L_{link} = \left(\frac{\lambda}{4\pi}\right)^2 \left(\frac{1}{d_{ref}^{\ 2}}\right)\left(\frac{d_{ref}}{r}\right)^{\gamma} \text{ for } \quad d_{ref} < r$$

Both the position of the break-point and the slope beyond it can therefore vary widely, and the model above is at best an indication of the loss.

VI.OUTLOOK FOR OPTICAL WIRELESS

Short term

The wide range of RF standards and the rapid increase in bit-rate available makes the adoption of large area coverage OW LANs unlikely in the short to medium term. However, within the broad range of strategies for 4G systems there is a common theme of heterogeneous wireless links, and there are situations where OW links can offer higher performance, or lower power consumption, or both than their RF counterparts. This is most likely where there are lines of sight or quasi line of sight. Links are likely to be narrow field of view and short range, or when the geometry can be controlled, such as in optical hotspots.

Medium term

The use of higher frequency RF approaches to obtain bandwidth requires path management, and makes the use of OW in 'managed' situations more likely given that the alternative has a similar requirement, and OW may offer a simpler solution with low power consumption.

Longer term

The results here show that the major long-term challenge for OW is to improve the link budget to that provided by RF systems, so that obtaining LOS geometries is less critical. Coherent systems offer a potential long term solution, albeit with formidable challenges in providing a stable low-cost geometry. Vertical cavity amplifiers using modified laser structures have been demonstrated, although these usually operate with very small etendue, which is unsuitable for OW applications without modification. Substantial parametric gain has been demonstrated over limited bandwidths, (although broadband operation is possible), as well as the use of Avalanche Detectors. Detector geometry and optimized devices also offer potential for increased antenna size and hence link margins.The optimum solution may well be a combination of these techniques, and further work is required to compare each of these approaches and determine the best future approach.

VII.RESEARCH DIRECTIONS AND CONCLUSION

The distinct properties of the OW channel can add to the 4G vision, with the possibility of a future terminal having a number of interfaces, both radio and optical. In order to achieve this work in the following areas is proposed, although this is not an exhaustive list.

Link budget improvement: the major barrier to non-LOS systems is the power required at the receiver, and work to improve this should be a major focus.

More comprehensive performance comparisons between short-range optical systems and their counterpart based on conventional RF approaches; there is a need to understand the properties of both channels between the same points, so that alternative data paths can be modeled, and the performance of a network that chooses the optimum path be determined.

Network modelling: understanding how optical and radio communications might co-exist.

Signal processing: examination of radio processing techniques such as Multisensor (MIMO) optical systems exploiting space and angular diversity. Space-time coding for optical wireless channels. Some work in this field has already been recently introduced by,where space-time codes are designed specifically for optical channels, specifically in the context of free-space optics communications. A MIMO channel model applied to diffuse WOC has been recently presented by.

Hybrid optical-RF systems: determination of the optimum method of using the alternative resources under different conditions, and the resulting performance improvement.

Visible light communications: fundamental capabilities and limitations of communication systems based on visible light (combined with illumination).

Future wireless standards offer a good opportunity for the wider adoption of OW. In particular, as 4G networks will be highly heterogeneous, OW based air interfaces can be incorporated to terminals in addition to the conventional RF based ones. Considerable work is still needed to fully exploit the clear advantages of the optical solutions, as well as developing low-cost subsystems and components to implement them.

REFERENCES

[1] R. van Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*: Artech House, 2000.

[2] J. Foerster, E. Green, S. Somayazulu, and D. Leeper, "Ultra-Wideband technology for short- or medium-range wireless communications," *Intel Technology Journal. 2001; (2):*, 2001.

[3] W. Hirt and D. Porcino, "Pervasive Ultra-wideband Low Spectral Energy Radio Systems (PULSERS)," presented at WWRF7, Eindhoven,, 2002.

[4] D. Porcino and W. Hirt, "Ultra-wideband radio technology: potential and challenges ahead," *IEEE Communications Magazine*, vol. 41, pp. 66-74, 2003.

[5] K. Gosse, "New Radio Interfaces for Short Range Communications," presented at WWRF9, 2003.

Overview of Performance of Coding Techniques in Mobile WiMAX Based System

Gurpreet Singh
Sai Institute of Engg. & Tech. Manawala, Asr.(Lecturer in ECE Deptt.)
Preet.boparai77@gmail.com

***Abstract:* - In this paper, we present the review of the convolution coding (CC) and convolutional product code (CPC) based on mobile WiMAX system. In coding techniques the number of symbols in the source encoded message is increased in a controlled manner in order to facilitate two basic objectives at the receiver one is Error detection and other is Error correction. The amount of error detection and correction required and its effectiveness depends on the signal to noise ratio (SNR). The WiMAX technology based on standard 802-16 wireless MAN is configured in the same way as a traditional cellular network with base stations using point to multipoint architecture to drive a service over a radius up to several kilometers. The range and the Non Line of Sight (NLOS) ability of WiMAX make the system very attractive for users, but there will be slightly higher BER at low SNR. Coding is a technique where redundancy is added to original bit sequence to increase the reliability of the communication. This paper reviews the code used in mobile WIMAX. A detailed description of the coding techniques in WIMAX system is studied.**

KEYWORDS: *WiMAX*, *CPC*, IEEE 802.16

I. INTRODUCTION

WiMAX is introduced by the Institute of Electrical and Electronic Engineers (IEEE) which is designated by 802.16 to provide worldwide interoperability for microwave access. There are fixed (802.16d) and mobile (802.16e) WiMAX. The majority of aspects which make WiMAX technology different from others that can be applied to the same scenario reside in its physical layer. This technology offers a high speed, secure, sophisticate, last mile broadband service, ensuring a flexible and cheap solution to certain rural access zones [1]-[7]. In a fixed wireless communication, WiMAX can replace the telephone company's copper wire networks, the cable TV's coaxial cable infrastructure. In comparison with Wi-Fi and Cellular technology, Wi-Fi provides a high data rate, but only on a short range of distances and with a slow movement of the user. And Cellular offers larger ranges and vehicular Mobility, but it provides lower data rates, and requires high investments for its deployment [4]. WiMAX tries to balance this situation. WiMAX fills the gap between Wi-Fi and Cellular, thus providing vehicular mobility, and high service areas and data rates. WiMAX is a standards based technology for wireless MANs conforming to parameters which enable interoperability[1][2]. WiMAX developments have been rapidly moving forward. The objective of this paper is to done literature review and to provide a critical review of the performance of the different coding techniques using mobile WiMAX. WIMAX transmit data with low bit error rate in the noisy environment for that we apply Forward Error Correction method with different coding techniques. This method is useful to reduce the bit error rate (BER) and increase the efficiency. The paper is organized as follows. In Section I, we give an introduction to the WiMAX system. Section II, the Literature is reviewed. Section III, describes the critical evaluation based on the literature review, we conclude in Section IV, which will summarize the work done in the paper followed by the future work. In last the references are given.

II. LITERATURE REVIEW

The purpose of this literature review is to study the literature of both coding techniques and WiMAX/802.16e.

A. WiMAX/802.16e

The WiMAX technology, based on the IEEE 802.16 Air Interface Standard is rapidly proving itself as a technology that will play a key role in fixed broadband wireless metropolitan areanetworks. In December, 2005 the IEEE ratified the 802.16e amendment [9] to the 802.16 standard. Mobile WiMAX is a broadband wireless solution that enables convergence of mobile and fixed broadband networks through a common wide area broadband radio access technology and flexible network architecture. The Mobile WiMAX Air Interface adopts Orthogonal Frequency Division Multiple Access (OFDMA) for improved multi-path performance in non-line-of-sight environments.

B. CONVOLUTION CODE (CC)

In the Mobile Wi-Max OFDMA part, the CC is the only mandatory coding scheme. Its computations depend not only on the current set of input symbols but on some of the previous input symbols. A trellis description is used for convolution encoding which gives relation how each possible input to the encoder influences the output in shift register. It uses the Viterbi algorithm for decoding. In communication, a **convolution code** [12] is a type of error-correcting code in which

- Each m-bit information symbol (each m-bit string) to be encoded is transformed into an n-bit symbol, where m/n is the code rate ($n \geq m$) and
- The transformation is a function of the last k information symbols, where k is the constraint length of the code.

- There are three parameters which define the convolotional code[11]:

(a) *Rate :* Ratio of the number of input bits to the number of output bits. In this example, rate is 1/2 which means there are two output bits for each input bit.
(b*) Constraint length* : The number of delay elements in the convolutional coding. In this example, with $K = 3$ there are two delay elements.
(c) *Generator polynomial :* Wiring of the input sequence with the delay elements to form the output. In this example, generator polynomial is $[7,5]_8 = [111,101]_2$. The output from the $7_8 = 111_2$ arm uses the XOR of the current input, previous input and the previous to previous input. The output from the $5_8 = 101_2$ uses the XOR of the current input and the previous to previous input.

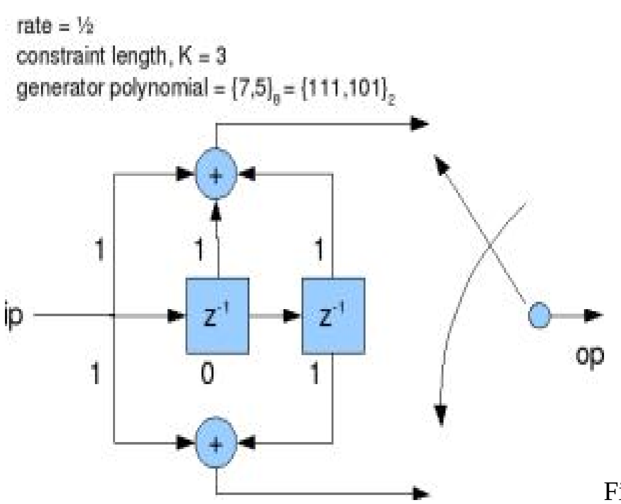


Figure 1: Convolutional code with Rate 1/2, K=3, Generator Polynomial [7,5] octal

C. CONVOLUTIONAL ENCODER

The channel coding scheme, IEEE 802-16, as shown in fig 2 is based on binary non-recursive Convolutional Coding (CC) [4].

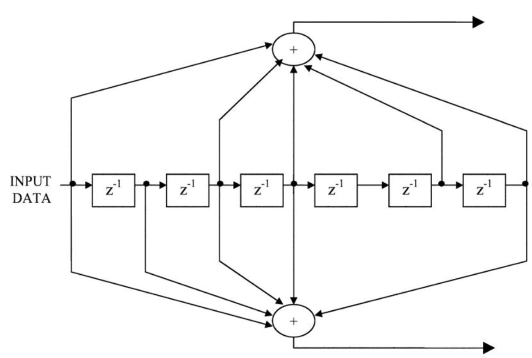


Fig: 2 Convolutional Encoder

The convolutional encoder uses a constituent encoder with constraint length 7, code rate 1/2 and generator polynomials (133,171) octal. Tail biting is used to initialize the encoder by padding each FEC block with 6 zeros.

D. CONVOLUTION PRODUCT CODE (CPC) METHOD

CPC is a new coding method , in which the information bits are placed into two dimensions (2D) Matrix. The rows and the columns are encoded separately by using recursive systematic convolutional encoders. Each row of the matrix is encoded using a convolutional code the same recursive systematic convolutional code is used to encode each row. Once all rows have been encoded, the matrix is sent, if desired, to an interleaver. Our original data matrix dimensions are (nxk), and the encoded data matrix dimensions will be (Znxk). The coded rows matrix is then recoded column by column using the same or different recursive systematic convolutional encoder. CPC uses a recursive systematic convolutional code with rate 1/2 and generator polynomials (1,5/7) octal to encode each row and column. Hence, the overall code rate is 1/4. The coding by CPC will be done in 2 stages. First each column will be independently coded, and then each row of the resulting matrix will be coded by the same generator polynomials. Therefore the following step are done[13]

(1) Dividing the overall matrix produced from CPC into three matrices. Each one has a size (nx4) or (nx6) according to the type of QAM .The reason for using three matrices only is to have a number of message bits equals to bits used in the turbo code method, as a comparison between it and CPC is done.
(2) Applying symbol mapping for each one independently (16QAM or 64QAM).
(3) Inserting the pilot and DC subcarriers for each matrix.
(4) Performing the IFFT independently resulting in three OFDMA symbols.
(5) Applying (cyclic prefix) CP for each symbol.
(6) Sending each symbol independently.

III. CRITICAL EVALUATION

In this section the evaluation of review is done critically. Table 1 contains the tabular format of a summarized review of the literature. What are the problem for performance of coding techniques and WiMAX? & what are the solutions for these problems.

Every author has its own view. The limitations found in the reviewed papers are also mentioned in this table.An analysis of the data communication on the WiMAX has been conducted. Main focus is on the error detection and correction during data transmission in noisy behavior.

TABLE I. SUMMARIZED TABLE OF THE REVIEW

Author	**Summary**	**Problems/ Challenges**	**Solutions**	**Limitations**
Prabhakar Telagarapu, G.B.S.R.Naidu, K.Chiranjeevi	Performance evaluation of physical layer of WiMAX by using Reed-Solomon coding and convolution coding scheme, cyclic prefix and interleaving for different modulation technique with respect to bit-error rate and SNR ratio.	When data is transmitted over a noisy channel than some information are loss and burst error.	The effect of error occur during transmission is reduced by adding redundancy to the data prior to transmission.	Low data rate, limited bandwidth.
Eng. Ahmed Ebian, Dr. Mona Shokair, Prof. Kamal Awadalla	A comparison between the performance of WiMAX using convolutional code and Convolutional Product Code (CPC) is made on the basis of BER.	Different behaviour of the WiMAX system with different coding techniques.	Using different number of sub carrier with different coding techniques.	Bit error rate.
Shraddha Bansal, Raksha Upadhyay	Mobile WiMAX, its physical layer is simulated using Matlab and bit error rate (BER) performance is observed. Performance improvement is achieved using forward error correction codes (FEC) for this convolution code (CC) and LDPC are used.	At small bit rate the behaviour of the system is not appropriate with convolution code.	LDPC perform better than CC in low SNR environment due to its inherent property. WiMAX Performance is improved in presence of FEC.	Not suitable for fading channel.

IV.CONCLUSION

Error detection and correction techniques are essential for reliable communication over a noisy channel. Error detection and correction technique are essential for reliable communication over a noisy channel. The effect of error occur during transmission is reduced by adding redundancy to the data prior to transmission. The redundancy is used to enable decoder in the receiver to detect and correct errors. CPC code is efficient code to correct the errors by using this codes in WiMAX we can reduce the bit-error in noisy environment and useful to provide the efficient data to the subscriber. In this paper, CPC coding and convolution coding method have been discussed. The CPC coding method leads to reduce BER as compare to convolution coding method.

V. FUTURE SCOPE

After studying different coding technique performance and mobile WiMAX and their performance, we came to an end that performance of coding techniques for provided error free communication but still not correct. A lot of error detection and correction techniques should be provided, so future work is needed in this area to error free communication. Further scope of this paper is that performance analysis of WiMAX MC-CDMA based system using CPC coding and convolution coding method. Which may gives better result as compare to WiMAX OFDMA based system using CPC coding and convolution coding method.

REFERENCES

[1] L. Nuaymi, *"WiMAX: Technology for Broadband Wireless Access,"* John Wiley & Sons, 2007.

[2] WiMAX Forum, "*Mobile WiMAX - Part I: A Technical Overview and Performance Evaluation,"* White paper, August 2006.

[3] IEEE 802.162004," *IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems"*, 1 October, 2004.

[4] WiMAX Forum, "*Mobile WiMAX - Part I: A Technical Overview and Performance Evaluation*," white paper, August 2006.

[5] Y. Xiao, "WiMAX-Mobile Fi: Advanced research and technology, "Auerbach publications, 2008.

[6] Changlong Xu,"Soft Decoding Algorithm for RS-CC Concatenated Codes in Wimax System", Vehicular Technology Conference, 2007.

[7] Jeffrey G Andrews, Arunabha Ghosh, and Rias Muhamed, "*Fundamentals of WiMAX "*

[8] "*Air Interface for Fixed Broadband Wireless Access Systems,"* IEEE STD 802.16 – 2004, October, 2004.

[9] "Air Interface for Fixed and Mobile Broadband Wireless Access Systems," IEEE P802.16e/D12, Feb, 2005.

[10] Eng. Ahmed Ebian, Mona Shokair and Kamal Awadalla, "Performance Evaluation of WiMAX System Using Convolutional Product Code (CPC)."

[11] Prabhakar Telagarapu, Prabhakar Telagarapu, K.Chiranjeevi," Analysis of Coding Techniques in WiMAX" *International Journal of Computer Applications Volume 22– No.3, May 2011*

[12] http://en.wikipedia.org/wiki/Convolutional_code.

[13] http://www.dsplog.com/2009/01/04/convolutional-code/

[14] http://www.waset.org/journals/waset/v51/v51-33.pdf

Automation Framework of Browser Based Testing Tool WATIR

Harwant Singh Gill[1] and Manjit Kaur[2]
[1,2]*Lecturer*, Global Institute of Management & Emerging Technologies, Amritsar, India
Email: harwantgill@gmail.com, manjitbhinder8@gmail.com

***Abstract*— This paper presents the automation framework of Browser based testing tool Watir (web application testing in `Ruby). The main advantage of a framework is the low cost for maintenance. If there is change to any test case then only the test case file needs to be updated and the Driver Script and Start-up script will remain the same. So, there is no need to update the scripts in case of changes to the application. Choosing the right framework helps in maintaining lower costs. There are several frameworks available suited for the particular application. Specifically, this paper discusses Keyword-driven framework for Watir and its various aspects in detail that can guide web application testing actively.**

I. INTRODUCTION

With the development of internet technology, the Web application becomes more and more complex and the scale of it changes more and more great. The Web application program testing is much more difficult than the traditional. Meanwhile, software development cycle becomes shorter and shorter; this makes web application testing, especially functional regression testing, more challengeable. Traditional ways of testing can no longer meet the need of the requirements of software development, thus automation testing becomes the only way out. As a frequently used method, the automation scripts created by Capture-Replay technology are too difficult to maintain. Thus developing effective web testing tool and framework becomes a hotspot in software testing research area [1].

II. BACKGROUND

A test automation framework is a set of assumptions, concepts and tools that provide support for automated software testing. Choosing the right framework/scripting technique helps in maintaining lower costs. The approach of scripting used during test automation has effect on costs. Automation Framework is not a tool to perform some specific task, but is an infrastructure that provides a complete solution where different tools work together in a unified manner [2, 3]. The test automation tool vendors market their product as the main feature of the tool is the ability to capture the user actions and later to playback them. Here is the basic paradigm for GUI-based automated regression testing – the so called Record/Playback method (also called as Capture/Replay approach)

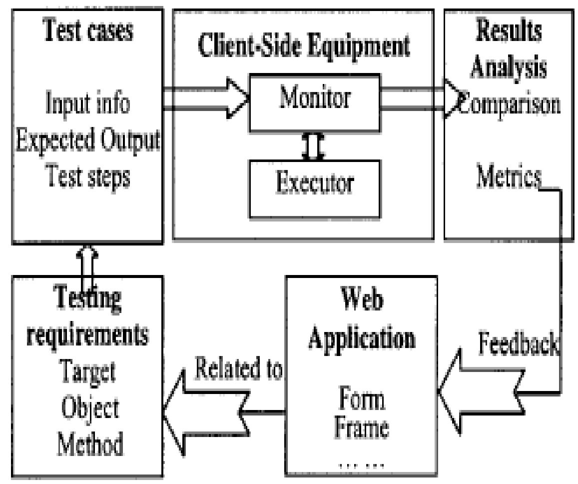


Fig. 1. Testing Framework for Web Applications

The basic drawback in this method is the scripts resulting from this method contain hard-coded values which must change if anything at all changes in AUT. The costs associated with maintaining such scripts are astronomical, and unacceptable. These scripts are not reliable, even if the application has not changed, and often fail on replay (pop-up windows, messages, and other things can happen that did not happen when the test was recorded).

If the tester makes an error entering data, etc., the test must be re-recorded even if the application changes, the test must be re-recorded. All that is being tested are things that already work. Areas that have errors are encountered in the recording process. These bugs are reported, but a script cannot be recorded until the software is corrected. So logically nothing is tested by this approach. This method is fraught with problems, and is the most costly (time consuming) of all methods over the long term.

The record/playback feature of the test tool is useful for determining how the tool is trying to process or interact with the application under test, and can give some ideas about how to develop test scripts, but beyond that, its usefulness ends quickly.

A. Types of Test Automation Frameworks

With the elimination of Record/Playback method, the shift is towards more powerful automation methodologies. The Automation testing framework is responsible for defining the format in which to express expectations, creating a mechanism to hook into or drive the application under test, executing the tests and reporting results. There are several test automation frameworks available, among these the selection is made based on the factors such as reusability of both the scripts and the test assets. The different test automation frameworks available are as follows [4].

- Test Script Modularity
- Test Library Architecture
- Data-Driven Testing
- Keyword-Driven or Table-Driven Testing
- Hybrid Test Automation

III. KEYWORD OR TABLE DRIVEN TEST AUTOMATION

Keyword-driven testing is growing in popularity. It involves the creation of modular, reusable test components that are built by test architects and then assembled into test scripts by test designers. Keyword-driven test looks very similar to manual test cases. In a keyword-driven test, the functionality of the system-under-test is documented in a table as well as in step by-step instructions for each test [5].

It requires the development of data tables and keywords, independent of the test automation tool used to execute them and the test script code that "drives" the application-under-test and the data. In a keyword-driven test, the functionality of the application-under-test is documented in a table as well as in step-by-step instructions for each test. In this method, the entire process is data-driven, including functionality.

It is typically an application-independent automation framework designed to process the tests. These tests are developed as data tables using a keyword vocabulary that is independent of the test automation tool used to execute them.

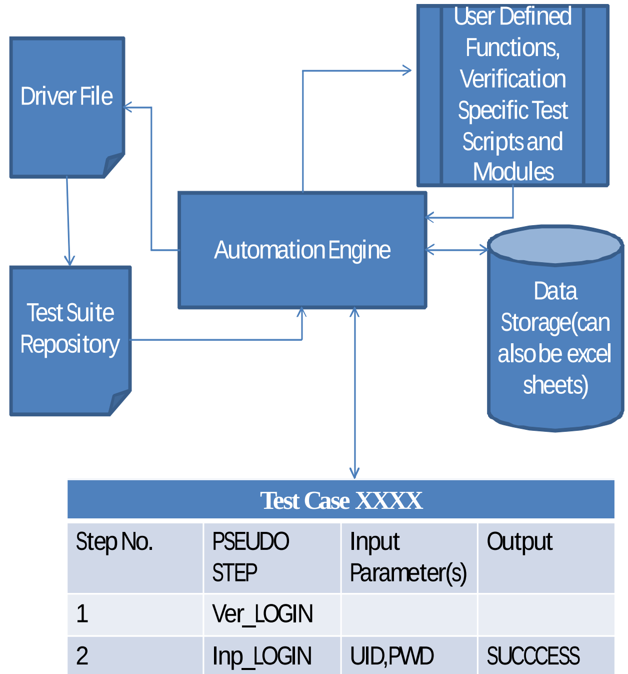


Fig.2 Keyword Driven Testing Framework

A. Action, Input Data, and Expected Result ALL in One Record:

The data table records contain the keywords that describe the actions we want to perform. They also provide any additional data needed as input to the application, and where appropriate, the benchmark information we use to verify the state of our components and the application in general.

For example, to verify the value of a user ID textbox on a login page, we might have a data table record as seen in Table 1:

WINDOW	**COMPONENT**	**ACTION**	**EXPECTE DVALUE**
LoginPage	UserIDTextbox	VerifyValu e	"MyUserI D"

Table 1

Another advantage of the keyword driven approach is that testers can develop tests without a functioning application as long as preliminary requirements or designs can be determined. All the tester needs is a fairly reliable definition of what the interface and functional flow is expected to be like.

Take login example. We do not need the application to construct any login tests. All we have to know is that we will have a login form of some kind that will accept a user ID, a password, and contain a button or two to submit or cancel the request. A quick discussion with development can confirm or modify determinations. We can then complete the test table and move on to another.

The essential guiding principles to be followed when developing overall test strategy are:

- The test design and the test framework are totally separate entities.
- The test framework should be application-independent.
- The test framework must be easy to expand, maintain, and perpetuate.
- The test strategy/design vocabulary should be framework independent.
- The test strategy/design should isolate testers from the complexities of the test framework. [6]

B Example

In order to open a window, the following table is devised, and it can be used for any other application, just it requires changing the window name.

Window	Control	Action	Arguments
Window Name	Menu	Click	File, Open
Window Name	Menu	Click	Close
Window Name	Pushbutton	Click	Folder Name
Window Name		Verify	Results

Once creating the test tables, a driver script or a set of scripts is written that reads in each step, executes the step based on the keyword contained the Action field, performs error checking, and logs any relevant information.

Keyword-driven testing is a tool used in automated testing where pre-defined keywords are used to define actions. These actions correspond to a process related to the application. It is the first step for creating a domain testing language. These keywords (or action words) represent a very simple specification language that non-programmers can use to develop automated tests. If these keywords are extended to emulate the domain language, customers and non technical testers can specify tests that map to the workflow more easily [7].

C. Framework Overview

Definition 1: A keyword identifies an operation or atomic action in test execution. In order to enable the automation test drivers to recognize the keyword, there are some restrictions on naming keyword.

For example, we define “Click”, “SetValue” and “VerifyValue” in GUI application as keywords. Similarly, “Connect” and “Cmd” belong to the keyword sets of database driver.

Definition 2: A command is used to simulate an action of software. In general, a test case can be parsed into sequences of test steps. To model the steps in test cases, we constitute a series of standards in our framework to formalize the steps of test cases into framework-recognizable input. We call them commands. Each step can be parsed into a command, and the command is based on keyword-driven.

A command can be defined as: (*TS, DriverName, Keyword, Arguments, ExpectedResults, And Variable Names*), where:

- $TS \in N+$: The number of test steps in test case;
- *Keyword* is a string which belongs to the pre-defined keyword sets;
- *DriverName* is a string to represent the test driver for executing the command;
- *Arguments* is an string list to define the properties and input of the keyword;
- *ExpectedResults* is a string to specify the expected results. It cannot be NULL if the purpose of the command is to verify a value;
- *VariableNames* is a string to specify the name of Variables. It cannot be NULL if we need to fetch a variable value during the test execution.

Definition 3: A Command Sequence contains a series of commands to be executed in a specific test driver in sequence. It can be defined as: (*DriverName, CommandList*), where:

- *DriverName* is a string to represent the test driver for executing the command;
- *CommandList* is a list of commands for execution in the test driven.

In this framework, the test cases are described in XML, which include the test steps, predefined keywords and expected results, etc. In automation engine layer, it will parse the test cases into the command sequence according to the driver type, and dispatch the command sequence to different test drivers in the test driver layer.

The test drivers will include UNIX Driver, Database Driver, GUI Driver and FTW Driver, etc. to dispatch the commands to the corresponding test applications to generate the test scripts automatically according to the keywords predefined in the commands. During the execution of the test scripts by the test applications under different test environments, the automation engine layer will coordinate the interfaces and results among the test drivers to ensure executing the process automatically [5].

D. Merits of keyword driven testing

The merits of the Keyword Driven Testing are as follows,

- The Detail Test Plan can be written in Spreadsheet format containing all input and verification data.
- If "utility" scripts can be created by someone proficient in the automated tool’s Scripting language prior to the Detail Test Plan being

written, then the tester can use the Automated Test Tool immediately via the "spreadsheet-input" method, without needing to learn the Scripting language.

- The tester needs only to learn the "Key Words" required, and the specific format to use within the Test Plan. This allows the tester to be productive with the test tool very quickly, and allows more extensive training in the test tool to be scheduled at a more convenient time.

E. Demerits of keyword driven testing

The demerits of the Keyword Driven Testing are as follows,

- Development of "customized" (Application-Specific) Functions and Utilities requires proficiency in the tool's Scripting language.
- If application requires more than a few "customized" Utilities, this will require the tester to learn a number of "Key Words" and special formats. This can be time-consuming, and may have an initial impact on Test Plan Development. Once the testers get used to this, however, the time required to produce a test case is greatly improved.

IV. CONCLUSION

The main concern of this paper is to describe the comprehensive framework for automating the testing process for Browser based testing Tool Watir. The framework provides a means for experimenting with different strategies for each testing component, within an integrated and automated framework. It includes how to execute the keyword based test case by using different test applications under different test environments to meet the specific requirements. We propose a novel and adaptive framework, which includes three layers: automation engine layer, test driver layer and test execution layer. Through the interactions among these three layers, the keyword-based commands in the test case can be converted into the test scripts to be executed under different test environments automatically. Additionally, it is important to note that this does not suggest that these tests can be executed automatically as soon as the application becomes available. The test record in a Table may be perfectly understood and executable by a person, but the automation framework knows nothing about the objects in the particular record until we can provide that additional information. That is a separate piece of the framework that can be discussed through application mapping as a future work.

REFERENCES

[1] Zeng Wandan, Jiang Ningkang, Zhou Xubo Design and Implementation of a Web Application Automation Testing Framework. 2009 Ninth International Conference on Hybrid Intelligent Systems, Software institute of East China Normal University Shanghai, China, 978-0-7695-3745-0/09, 2009 IEEE

[2] http:// Test automation framework - Wikipedia, the free encyclopedia.mht

[3] Lei Xu, Baowen Xu, A Framework for Web Applications Testing, Proceedings of the 2004 International Conference on Cyberworlds (CW'04), IEEE

[4] http:// Automation testing - Wikipedia, the free encyclopedia.mht

[5] Jingfan Tang1 and Xiaohua Cao Albert Ma, Towards Adaptive Framework of Keyword Driven Automation Testing, Proceedings of the IEEE International Conference on Automation and Logistics 2008

[6] http://safsdev.sourceforge.net/FRAMESDataDrivenTestAutomationFrameworks.htm

[7] http:// Keyword driven testing /Keyword driven testing.pdf– Wikipedia, the free encyclopedia.mht

Digital Image Watermarking

Shuchi Saini
M.Tech- E.C.E, Amritsar College of Engineering & Technology
saini2608@yahoo.com

***Abstract* — Digital watermarking is a highly evolving field, which involves the embedding of a certain kind of information under a digital object (image, video, audio) for the purpose of copyright protection. This paper includes various digital watermarking techniques and spatial domain (Least Significant Bit(LSB)) as well as transform domain (Discrete Cosine Transform(DCT) and Discrete Wavelet Transform (DWT)) methods.**

Keywords—Digital watermarking, Least Significant Bit, Discrete Cosine Transform, Discrete Wavelet Transform

I. INTRODUCTION

Digital watermarking is defined as a process of embedding data (watermark), into a multimedia object to help to protect the owner's right to that object. The ease with which digital content can be exchanged over the Internet has created copyright infringement issues. Digital image watermarking is modification of the original image data by embedding a watermark containing key information such as authentication or copyright codes [1]. Watermark is perceptible or imperceptible identification code which uniquely identifies ownership of an image. It is permanently embedded into the host image.

There are four essential factors those are commonly used to determine quality of watermarking scheme. They are robustness, imperceptibility, capacity, and blindness.

1) *Robustness:* Watermark should be difficult to remove or destroy. Robust is a measure of immunity of watermark against attempts to image modification and manipulation like compression, filtering, rotation, scaling, collision attacks, resizing, cropping etc.
2) *Imperceptibility*: means quality of host image should not be destroyed by presence of watermark.
3) *Capacity:* It includes techniques that make it possible to embed majority of information.
4) *Blind Watermarking:* Extraction of watermark from watermarked image without original image is preferred because sometimes it's impossible to avail original image.

The embedded watermark may be pseudo-random binary sequence, chaotic sequence, spread spectrum sequence or binary/gray scale image and is mainly divided in two categories spatial domain techniques [2] and frequency domain techniques such as DWT and DCT are used for objective embedding watermark and detection using correlation measures.

II. DIGITAL WATERMARKING TECHNIQUE

Watermark can be considered as a kind of a signature that reveals the owner of the multimedia object. A watermarking algorithm embeds a visible or invisible watermark in a given multimedia object. The embedding process is guided by use of a secret key which decided the locations within the multimedia object (image) where the watermark would be embedded. Once the watermark is embedded it can experience several attacks because the multimedia object can be digitally processed. The attacks can be unintentional (in case of images, low pass filtering or gamma correction or compression) or intentional (like cropping). Hence the watermark has to be very robust against all these possible attacks. When the owner wants to check the watermarks in the possibly attacked and distorted multimedia object, s/he relies on the secret key that was used to embed the watermark.

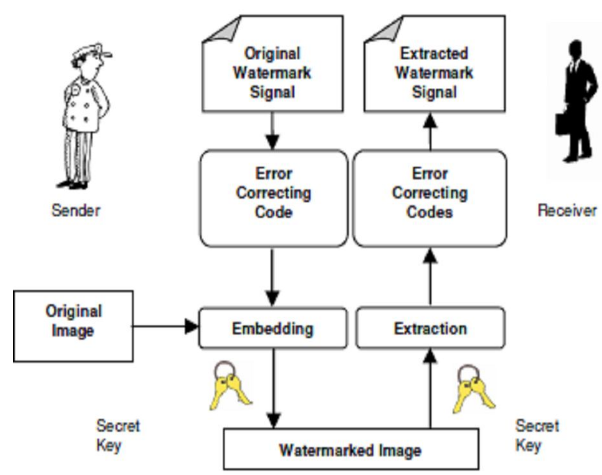


Fig.1 Watermarking Technique

Using the secret key, the embedded watermark sequence can be extracted. This extracted watermark may or may not resemble the original watermark because the object might have been attacked. Hence to validate the existence of watermark, either the original object is used to compare and find out the watermark signal (non-blind watermarking) or a correlation measure is used to detect the strength of the watermark signal from the extracted watermark (blind water marking). In the correlation based detection the original watermark sequence is compared with the extracted watermark sequence and a statistical correlation test is used to determine the existence of the watermark.

III. VARIOUS DIGITAL WATERMARKING TECHNIQUES

Various watermarking techniques are as follows:

1) Based on Working Domain

Two techniques are there: spatial domain and frequency domain techniques. The spatial domain techniques consist of two categories, out of which first one is Least- Significant Bit (LSB). It is used to hide the information and attackers could not destroy the information. Another one is SSM-Modulation-Based Technique. This technique is applied in the water marking algorithms with an linked information and attached to the original image with pseudo noise signal, its modulated by the watermark. The frequency-domain techniques mainly used for watermarking of the human visual system are better captured by the spectral coefficients. The transforms are broadly categorized in two ways: Discrete Cosine Transformation (DCT) and Discrete Wavelet Transformation (DWT).

2) Based on Human Perception

It includes Visible and Invisible watermarking or Transparent. Visible watermarks are ones, which are embedded in visual content in such a way that they are visible when the content is viewed. Whereas Transparent watermarks are imperceptible and they cannot be detected by just viewing the digital content.

3) Robust & Fragile Watermarking

Robust watermarking is a technique in which modification to the watermarked content will not affect the watermark. In contrast, fragile watermarking is a technique in which watermark gets destroyed when watermarked content is modified or tampered with.

4) Public & Private Watermarking

In public watermarking, users of the content are authorized to detect the watermark while in private watermarking the users are not.

5) *Asymmetric &Symmetric Watermarking*

Asymmetric watermarking (also called asymmetrickey watermarking) is a technique where different keys are used for embedding and detecting the watermark. In symmetric watermarking (or symmetric key watermarking) the same keys are used for embedding and detecting watermarks.

6) Steganographic & Non-steganography watermarking

Steganographic watermarking is the technique where content users are unaware of the presence of a watermark. In non-stenographic watermarking, the users are aware of the presence of a watermark.

7) Blind and non-blind watermarking

In case of blind watermarking, receiver does not know the idea of position of watermark in watermarked image. Where as in case of non-blind extraction process receiver have idea of position.

IV. SPATIAL DOMAIN METHOD

The spatial domain is the normal image space, in which a change in position in I directly projects to a change in position in space. Distances in I (in pixels) correspond to real distances (e.g. in meters) in space. This concept is used most often when discussing the frequency with which image values change, means, over how many pixels does a cycle of periodically repeating intensity variations occur. One would refer to the number of pixels over which a pattern repeats (its periodicity) in the spatial domain. Here we use Least Significant bit (LSB) method.

1) Least Significant Bit

One of the simplest technique in digital watermarking is in spatial domain using the two dimensional array of pixels in the container image to hold hidden data using the least significant bits (LSB) method. Note that the human eyes are not very attuned to small variance in color and therefore processing of small difference in the LSB will not noticeable. The steps to embed watermark image are given below:

A. Steps of Least Significant bit

1) Convert RGB image to gray scale image.
2) Make double precision for image.
3) Shift most significant bits to low significant bits of watermark image.
4) Make least significant bits of host image to zero
5) Add shifted version (step 3) of watermarked image to modified (step 4) host image.

B. Limitations of Spatial Domain Watermarking

This method is comparatively simple, lacks the basic robustness that may be expected in any watermarking applications. It can survive simple operation such as cropping, any addition of noise. However lossy compression is going to defeat the watermark. An even better attack is to set all the LSB bits to ‘1’ fully defeating the watermark at the cost of negligible perceptual impact on the cover object. Furthermore, once the algorithm was discovered, it would be very easy for an intermediate party to alter the watermark.

V. TRANSFORM DOMAIN METHOD

The produce of high quality watermarked image is by first transforming the original image into the frequency domain by the use of Fourier, Discrete Cosine Transform (DCT) or Discrete Wavelet transforms (DWT) for example. With this technique, the marks are not added to the intensities of the image but to the values of its transform coefficients. Then inverse transforming the marked coefficients forms the watermarked image. The use of frequency based transforms allows the direct understanding of the content of the image; therefore, characteristics of the human visual system (HVS) can be taken into account more easily when it is time to decide the intensity and position of the watermarks to be applied to a given image.

1) Discrete Cosine Transform Watermarking

The DCT allows an image to be broken up into different frequency bands, making it much easier to embed watermarking information into the middle frequency bands of an image. The middle frequency bands are chosen such that they have minimized they avoid the most visual important parts of the image (low frequency) without over-exposing themselves to removal through compression and noise attacks.

A. Steps of DCT watermarking

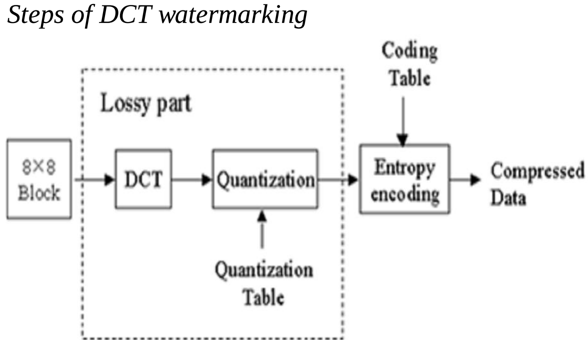


Fig. 2 DCT Watermarking

1) Transforming from the RGB color of the original image into the formation of Gray color
2) To divide the image into 8 × 8 blocks by JPEG standard as below.
3) Transforms original 8 x 8 block into a cosine-frequency domain
a) C(h) = if (h == 0) then 1/sqrt(2) else 1.0
--C(h) is a auxiliary function used in main function F(u,v)
b) F(u,v) = ¼ x C(u) x C(v) Σx=0..7 Σy=0..7 Dxy x cos(π(2u + 1)u/16) x cos(π(2y + 1)v/16)
-Gives encoded pixel at row u, column v
-Dxy is original pixel value at row x, column y
-F(u,v) is new matrix value after DCT apply.

B. Extracting Watermarked Image

1) Perform DCT transform on watermarked image and original host image.
2) Subtract original host image from watermarked image.
3) Multiply extracted watermark by scaling factor to display.

C. Advantages

1) DCT domain watermarking is comparatively much better than the spatial domain encoding since DCT domain watermarking can survive against the attacks such as noising, compression, sharpening, and filtering.
2) It use JPEG compression method to apply DCT watermarking as a parameter. One may use different parameters related to image processing, and these parameters might provide equal or even stronger robustness against various attacks based on image processing.
3) Discrete cosine transform (DCT), where pseudorandom sequences, such as M sequences, are added to the DCT at the middle frequencies as signatures.

2) Discrete Wavelet Transform Watermarking

The basic idea in the DWT for a one dimensional signal is as: A signal is split into two parts, usually high frequencies and low frequencies. The edge components of the signal are largely to the high frequency part. The low frequency part is split again into two parts of high and low frequencies. This process is continued an arbitrary number of times, which is usually determined by the application at hand.

A. Steps of DWT watermarking

1) The first part of the watermarking process is, of course, the encoder. The first step is to decompose the image into four frequency bands using first resolutions of Haar wavelets at first level. In second level, decompose image into seven frequency bands using second resolutions of Haar wavelets. At three level, decompose image into ten frequency bands using third resolutions of Haar wavelets and so on.
2) The next operation is to add a pseudo random sequence N , in fact a Gaussian distribution of mean zero and variance one, to the coefficients of the medium and high frequency bands (i.e. all the bands except the lowest one which is represented by the top left corner in Fig. 3.
3) The normal distribution is used because it has been proven to be quite robust to collusive attacks. We can

further weight the watermark according to the magnitude of the wavelet coefficients.

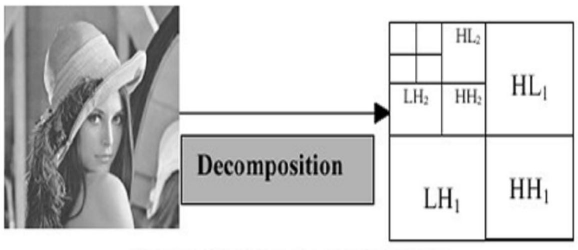


Fig. 3 Resulting Decompose image

B. Advantages

1) The watermarking method has multi resolution characteristics and is hierarchical. It is usually true that the human eyes are not sensitive to the small changes in edges and textures of an image but are very sensitive to the small changes in the smooth parts of an image. With the DWT, the edges and textures are usually to the high frequency sub bands, such as HH, LH, HL etc. Large frequencies in these bands usually indicate edges in an image.

2) The watermarking method robust to wavelet transform based image compressions, such as the embedded zero-tree wavelet (EZW) image compression scheme, and as well as to other common image distortions, such as additive noise, rescaling/stretching, and half toning. This is advantage over DCT.

VI. CONCLUSION

This paper provides review of various watermarking techniques. DCT and DWT domain watermarking is comparatively much better than the spatial domain.

REFERENCES

[1] Lin, S. and C. Chin, 2000. "A Robust DCT-based Watermarking for Copyright Protection," IEEE Trans. Consumer Electronics, 46(3): 415-421.

[2] Deng, F. and B. Wang, 2003. "A novel technique for robust image watermarking in the DCT domain," in Proc. of the IEEE 2003 Int. Conf. on Neural Networks and Signal Processing, vol. 2, pp: 1525-1528.

[3] http://cisjournal.org/archive/vol2no1/vol2no1_5.pdf

[4] http://ebookbrowse.com/gdoc.php?id=2855663&url=8af5267f6d64011df7b56ceaf3a95744

[5] http://www.academypublisher.com/ijrte/vol01/no02/ijrte0102223225.pdf

[6] http://www.digitavid.net/VirtualUniveristy/Presentations/ImagingWatermarking.pdf

[7] http://www.enggjournals.com/ijcse/doc/IJCSE10-02-09-042.pdf

[8] http://www.imada.sdu.dk/~joan/security/reports06/watermarks.pdf

[9] http://www.ijetae.com/files/Volume2Issue5/IJETAE_0512_61.pdf

Energy Harvestation using Wireless based Microwaves (Wireless Mobile Charger)

Ravinder Kaur[1], Anu Sheetal[2]
[1]M.Tech. Student, ECE, BCET, Gurdaspur
[2]GNDU, Regional Campus,Gurdaspur
Email: ravsabharwal07@gmail.com, nripanu@yahoo.co.in

***Abstract*--With mobile phones becoming a basic part of life, the recharging of mobile phone batteries has always been a problem. The mobile phones vary in their talk time and battery standby according to their manufacturer and batteries. All these phones irrespective of their manufacturer and batteries have to be put to recharge after the battery has drained out. The main objective of this current proposal is to make the recharging of the mobile phones independent of their manufacturer and battery make. In this paper a new proposal has been made so as to make the recharging of the mobile phones is done automatically as you talk in your mobile phone. This is done by use of microwaves. The microwave signal is transmitted from the transmitter along with the message signal using special kind of antennas called slotted wave guide antenna at a frequency is *2.45 GHz*. There are minimal additions, which have to be made in the mobile handsets, which are the addition of a sensor, a Rectenna, and a filter. With the above setup, the need for separate chargers for mobile phones is eliminated and makes charging universal. Thus the more you talk, the more is your mobile phone charged? With this proposal the manufacturers would be able to remove the talk time and battery standby from their phone specifications.**

I.INTRODUCTION

A.THE ELECTROMAGNETIC SPECTRUM

To start with, to know what a spectrum is: when white light is shone through a prism it is separated out into all the colors of the rainbow; this is the visible spectrum. So, white light is a mixture of all colors. Black is not a color; it is what you get when all the light is taken away. Some physicists pretend that light consists of tiny particles which they call photons. They travel at the speed of light (what a surprise). The speed of light is about 300,000,000 meters per second. When they hit something they might bounce off, go right through or get absorbed. What happens depends a bit on how much energy they have. If they bounce off something and then go into your eye you will "see" the thing they have bounced off. Some things like glass and Perspex will let them go through; these materials are transparent. Black objects absorb the photons so you should not be able to see black things: you will have to think about this one. These poor old physicists get a little bit confused when they try to explain why some photons go through a leaf, some are reflected, and some are absorbed. They say that it is because they have different amounts of energy. Other physicists pretend that light is made of waves. These physicists measure the length of the waves and this helps them to explain what happens when light hits leaves. The light with the longest wavelength (red) is absorbed by the green stuff (chlorophyll) in the leaves. So is the light with the shortest wavelength (blue).

In between these two colors there is green light, this is allowed to pass right through or is reflected. (Indigo and violet have shorter wavelengths than blue light.)

Well it is easy to explain some of the properties of light by pretending that it is made of tiny particles called photons and it is easy to explain other properties of light by pretending that it is some kind of wave. The visible spectrum is just one small part of the electromagnetic spectrum. These electromagnetic waves are made up of to two parts. The first part is an electric field. The second part is a magnetic field. So that is why they are called electromagnetic waves. The two fields are at right angles to each other.

Designation	Frequency range
L Band	1 to 2 GHz
S Band	2 to 4 GHz
C Band	4 to 8 GHz
X Band	8 to 12 GHz
Ku Band	12 to 18 GHz
K Band	18 to 26 GHz
Ka Band	26 to 40 GHz
Q Band	30 to 50 GHz
U Band	40 to 60 GHz

Table1 Radar bands with their frequency range

MICROWAVE REGION

Microwave wavelengths range from approximately one millimeter (the thickness of a pencil lead) to thirty centimeters (about twelve inches). In a microwave oven, the radio waves generated are tuned to frequencies that can be absorbed by the food. The food absorbs the energy and gets warmer. The dish holding the fooddoesn't absorb a significant amount of energy and stays much cooler. Microwaves are emitted from the Earth, from objects such as cars and planes, and from the atmosphere. These microwaves can be detected to give information, such as the temperature of the object that emitted the microwaves. Microwaves have wavelengths that can be measured in centimeters! The longer microwaves, those closer to a foot in length, are the waves which heat our food in a microwave oven. Microwaves are good for transmitting information from one place to another because microwave energy can penetrate haze, light rain and snow, clouds, and smoke. Shorter microwaves are used in remote sensing. These microwaves are used for clouds and smoke, these waves are good for viewing the Earth from space Microwave waves are used in the communication industry and in the kitchen as a way to cook foods. Microwave radiation is still associated with energy levels that are usually considered harmless except for people with pace makers.

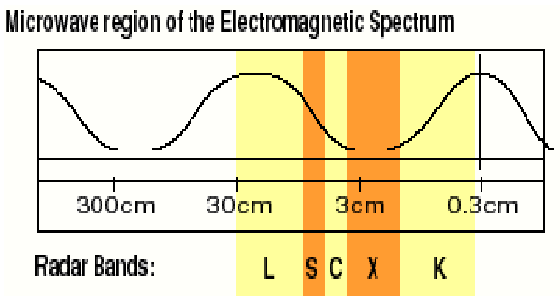


Fig.1 Microwave region

Here we are going to use the S band of the Microwave Spectrum. The frequency selection is another important aspect in transmission. Here we have selected the license free 2.45 GHz ISM band for our purpose. The Industrial, Scientific and Medical (ISM) radio bands were originally reserved internationally for non-commercial use of RF electromagnetic fields for industrial, scientific and medical purposes. The ISM bands are defined by the ITU-T in S5.138 and S5.150 of the Radio Due to variations in national radio regulations. In recent years they have also been used for license-free error-tolerant communications applications such as wireless LANs and Bluetooth: 900 MHz band (33.3 cm) (also GSM communication in India) 2.45 GHz band (12.2 cm) IEEE 802.11b wireless Ethernet also operates on the 2.45 GHz.

II.TRANSMITTER DESIGN

A.MAGNETRON

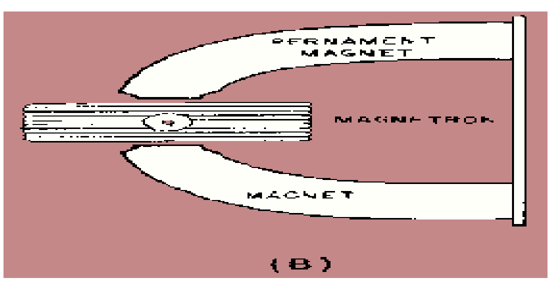


Fig.2 Magenetron

The Magnetron is a self-contained microwave oscillator that operates differently from the linear-beam tubes, such as the TWT and the klystron. Fig.2 is a simplified drawing of the magnetron. CROSSED-ELECTRON and MAGNETIC fields are used in the magnetron to produce the high-power output required in radar and communications equipment. The magnetron is classed as a diode because it has no grid. A magnetic field located in the space between the plate (anode) and the cathode serves as a grid. The plate of a magnetron does not have the same physical appearance as the plate of an ordinary electron tube. Since conventional inductive-capacitive (LC) networks become impractical at microwave frequencies, the plate is fabricated into a cylindrical copper block containing resonant cavities that serve as tuned circuits. The magnetron base differs considerably from the conventional tube base.

The magnetron base is short in length and has large diameter leads that are carefully sealed into the tube and shielded. The cathode and filament are at the center of the tube and are supported by the filament leads. The filament leads are large and rigid enough to keep the cathode and filament structure fixed in position. The output lead is usually a probe or loops extending into one of the tuned cavities and coupled into a waveguide or coaxial line. The plate structure is a solid block of copper.

The cylindrical holes around its circumference are resonant cavities. A narrow slot runs from each cavity into the central portion of the tube dividing the inner structure into as many segments as there are cavities. Alternate segments are strapped together to put the cavities in parallel with regard to the output. The cavities control the output frequency. The straps are circular, metal bands that are placed across the top of the block at the entrance slots to the cavities. Since the cathode must operate at high power, it must be fairly large and must also be able to withstand high operating temperatures. It must also have good emission characteristics, particularly under return bombardment by the electrons. This is because most of the output power is provided by the large number of electrons that are emitted when high-velocity electrons return to strike the cathode. The Cathode is indirectly heated and is constructed of a high-emission material. The open space between the plate and the cathode is called the INTERAC TION SPACE. In this space the electric and magnetic fields interact to exert force upon the electrons.

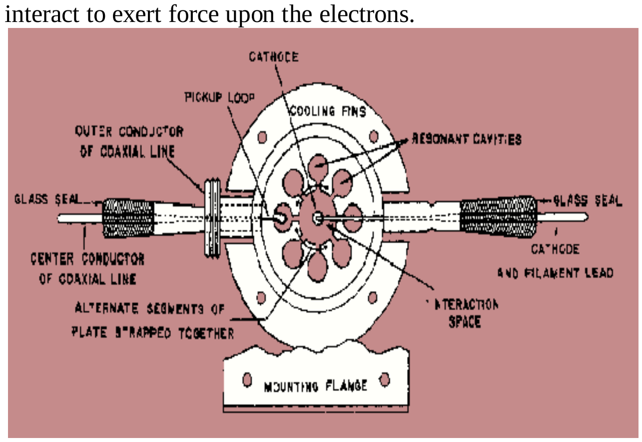


Fig3.Parts of magnetron

III.RECEIVER DESIGN

The basic addition to the mobile phone is going to be the rectenna. A rectenna is a rectifying antenna, a special type of antenna that is used to directly convert microwave energy into DC electricity. Its elements are usually arranged in a mesh pattern, giving it a distinct appearance from most antennae. A simple rectenna can be constructed from a Schottky diode placed between antenna dipoles. The diode rectifies the current induced in the antenna by the microwaves. Rectenna are highly efficient at converting microwave energy to electricity. In laboratory environments, efficiencies above 90% have been observed with regularity. Some experimentation has been done with inverse rectenna, converting electricity into microwave energy, but efficiencies are much lower--only in the area of 1%. With the advent of nanotechnology and MEMS the size of these devices can be brought down to molecular level. It has been theorized that similar devices, scaled down to the proportions used in nanotechnology, could be used to convert light into electricity at much greater efficiencies than what is currently possible with solar cells.

This type of device is called an optical rectenna. Theoretically, high efficiencies can be maintained as the device shrinks, but experiments funded by the United States National Renewable Energy Laboratory have so far only obtained roughly 1% efficiency while using infrared light. Another important part of our receiver circuitry is a simple sensor. This is simply used to identify when the mobile phone user is talking. As our main objective is to charge the mobile phone with the transmitted microwave after rectifying it by the rectenna, the sensor plays an important role. The whole setup looks something like this.

IV.PROCESS OF RECTIFICATION

A rectifying antenna rectifies received microwaves into DC current. A rectenna comprises of a mesh of dipoles and diodes for absorbing microwave energy from a transmitter and converting it into electric power. Its elements are usually arranged in a mesh pattern, giving it a distinct appearance from most antennae. A simple rectenna can be constructed from a Schottky diode placed between antenna dipoles as shown in Fig.3. The diode rectifies the current induced in the antenna by the microwaves. Rectenna are highly efficient at converting microwave energy to electricity. In laboratory environments, efficiencies above 90% have been observed.

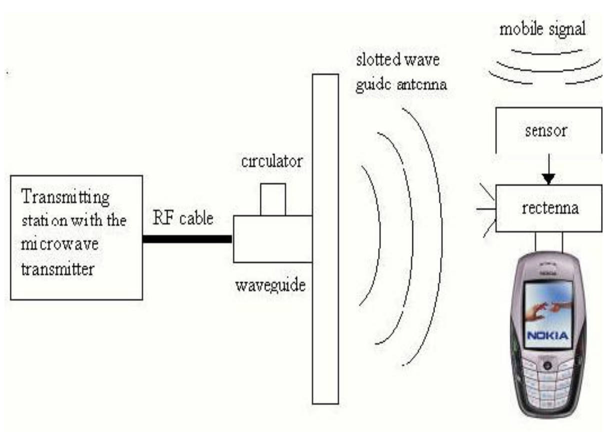


Fig.4 Rectification process

In future rectennas will be used to generate large-scale power from microwave beams delivered from orbiting SPS satellites.

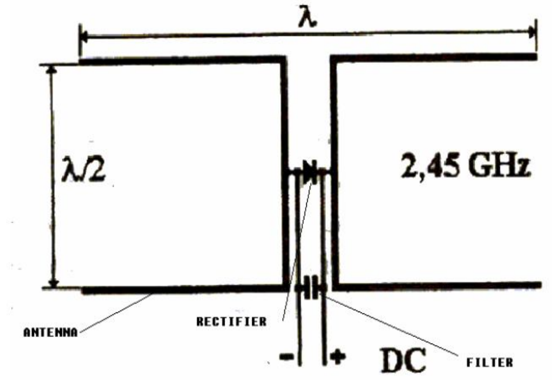


Fig 5.Optical Antenna

BRIEF INTRODUCTION OF SCHOTTKY BARRIER DIODE

A Schottky barrier diode is different from a common P/N silicon diode. The common diode is formed by connecting a P type semiconductor with an N type semiconductor, this is connecting between a semiconductor and another semiconductor; however, a Schottky barrier diode is formed by connecting a metal with a semiconductor. When the metal contacts the semiconductor, there will be a layer of potential barrier (Schottky barrier) formed on the contact surface of them, which shows a characteristic of rectification. The material of the semiconductor usually is a semiconductor of n-type (occasionally p-type), and the material of metal generally is chosen from different metals such as molybdenum, chromium, platinum and tungsten. Sputtering technique connects the metal and the semiconductor. A Schottky barrier diode is a majority carrier device, while a common diode is a minority carrier device. When a common PN diode is turned from electric connecting to circuit breakage, the redundant minority carrier on the contact surface should be removed to result in time delay. The Schottky barrier diode itself has no minority carrier, it can quickly turn from electric connecting to circuit breakage, its speed is much faster than a common P/N diode, so its reverse recovery time (t_{rr}) is very short and shorter than 10 nS. And the forward voltage bias of the Schottky barrier diode is under 0.6V or so, lower than that (about 1.1V) of the common PN diode. So, The Schottky barrier diode is a comparatively ideal diode, such as for a 1 ampere limited current PN interface. Below is the comparison of power consumption between a common diode and a Schottky barrier diode:

P=0.6*1=0.6W

P=1.1*1=1.1W

It appears that the standards of efficiency differ widely. Besides, the PIV of the Schottky barrier diode is generally far smaller than that of the PN diode; on the basis of the same unit, the PIV of the Schottky barrier diode is probably 50V while the PIV of the PN diode may be as high as 150V. Another advantage of the Schottky barrier diode is a very low noise index that is very important for a communication receiver; its working scope may reach20GHz.

SENSOR CIRCUITRY

The sensor circuitry is a simple circuit, which detects if the mobile phone receives any message signal. This is required, as the phone has to be charged as long as the user is talking. Thus a simple F to V converter would serve our purpose. In India the operating frequency of the mobile phone operators is generally 900MHz or 1800MHz for the GSM system for mobile communication. Thus the usage of simple F to V converters would act as switches to trigger the rectenna circuit to on. A simple yet powerful F to V converter is LM2907. Using LM2907 would greatly serve our purpose. It acts as a switch for triggering the rectenna circuitry. The general block diagram for the LM2907 is givenbelow.

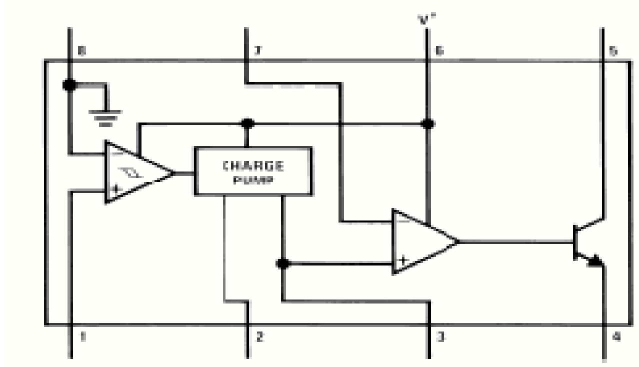


Fig.4 Block diagram of LM2907

Thus on the reception of the signal the sensor circuitry directs the rectenna circuit to ON and the mobile phone begins to charge using the microwave power.

V.CONCLUSION

Thus this paper successfully demonstrates a novel method of using the power of the microwave to charge the mobile phones without the use of wired chargers. Thus this method provides great advantage to the mobile phone users to carry their phones anywhere even if the place is devoid of facilities for charging. A novel use of the rectenna and a sensor in a mobile phone could provide a new dimension in the revelation of mobile phone.

REFERENCES

[1] Hawkins, Joe, et al, "Wireless Space Power Experiment," in Proceedings of the Advanced Space Design Program, June 14-18, 1993.

[2] MW Medley Jr and MW Medley, 'Microwave and RF circuits: analysis, Artech House, Norwood, MA, 1993.

[3] Falcone, Vincent J., "Atmospheric Attenuation of Microwave Power," Journal of microwave Power, 5(4), 1971.

Fingerprint Recognition using Hybrid Matching Method

Megha Kulshesta[1], Dr. Vijay Kumar Banga[2] and Sanjeev Kumar[3]
[1]M.Tech, ECE, ACET, Amritsar, India
[2, 3] Professor and Associate Professor, Department of ECE, ACET, Amritsar, India
Email: sanjeev_be@indiatimes.com, Vijaykumar.banga@gmail.com

***Abstract*—Human fingerprints are reliable characteristics for personnel identification as it is unique and persistence. Fingerprints are the most popular and studied biometrics features. Their stability and uniqueness make thefingerprint identification system extremely reliable and useful forsecurity applications. Fingerprints are the oldest and most widely used form of biometric identification. Everyone is known to have unique, immutable fingerprints. A fingerprint pattern consists of ridges, valleys and minutiae. In this paper we propose Fingerprint Verification based on Gabor Filter algorithm for minutiae feature extraction.**

***Keywords*- Fingerprint Recognition, Minutiae, Gabor Filters, Biometric Accuracy.**

I. INTRODUCTION

The term Biometrics relates to the measurement (metric) of characteristics of a living (Bio) thing in order to identify a person. Biometric recognition is used as an automatic recognition of individuals based on the physiological or behavioral characteristics. A physiological characteristic such as Fingerprint, Face, Iris, Hand geometry and Retina remains same throughout the lifetime of a person. Behavioral characteristics such as signature, gait, voice and keystroke changes with age and mentality of a person. Fingerprint is one of the most widely used biometric because individuals have tried a number of methods to obliterate or remove their fingerprints by abrading or trying to rub off their fingerprints, burning the fingers, trying to dissolve their fingerprints with strong acids, covering their fingertips with superglue, and even having doctors transplant skin from one finger to another unsuccessfully. The fingerprint is formed in third and fourth month of fetal development and unique with epidermal ridges, furrows and patterns. Distinctiveness and persistence are the highly desirable qualities of fingerprint. Factors such as Skin condition, finger pressure and noise detroit the quality of fingerprint images.

The presence of noise gives an image a mottled, grainy, textured or snowy appearance [9]. The images of fingerprint are mistaken during termination; hence it is necessary to investigate the images of fingerprint by using suitable filters. Quality of fingerprint is assessed by Character, Fidelity and Utility. A smoothly flowing pattern formed by alternating crests (ridges) and troughs (valleys) on the palmer aspect of hand is palm print. Formation of a palm print depends on the initial conditions ofthe embryonic mesoderm from which they develop. The pattern on pulp of each terminal phalanx is considered as an individual pattern and is commonly referred to as a Fingerprint.

A fingerprint is believed to be unique to each person, even identical twins have different. It is one of the most mature biometric technologies and is considered legitimate proofs of evidence all over the world, hence used in forensic divisions worldwide for criminal investigations. Fingerprintbased identification has better matching performance than any other existing biometrictechnology. The watermark is embedded into the region of interest in the Fingerprint image to enhance the security of Biometric System. An automatic fingerprint matching is popular in systems which control access to physical locations, computer/network resources, bank accounts or register employee attendance time in company [6].

The matching between the fingerprint pattern are highly sensitive to errors such as various noises, damaged fingerprint areas, the finger being placed in different areas of fingerprint scanner window with different orientation angles and finger deformation during the scanning procedure etc. Minutiae in the fingerprint images are the terminations and bifurcations of ridge lines.

II. FINGERPRINT RECOGNITION

The fingerprint recognition problem can be grouped into three sub-domains: fingerprint enrollment, verification and fingerprint identification. In addition, different from the manual approach for fingerprint recognition by experts, the fingerprint recognition here is referred as AFRS (Automatic Fingerprint Recognition System), which is program-based [5]. Verification is typically used for positive recognition, where the aim is to prevent multiple people from using the same identity.

State of the Art in Fingerprint Recognition

This section provides a basic introduction to fingerprint recognition systems and their main parts, including a brief description of the most widely used techniques and algorithm.

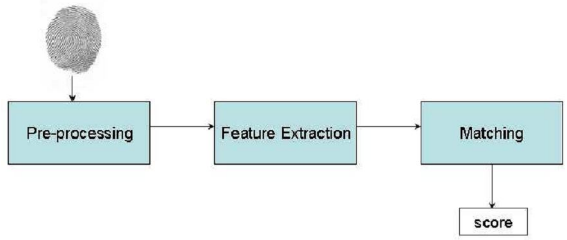


Figure 1 Main modules of a fingerprint verification system

The main modules of a fingerprint verification system (Fig.1) are:

a) fingerprint sensing, in which the fingerprint of an individual is acquired by a fingerprintscanner to produce a raw digital representation;
b) Preprocessing, in which the inputfingerprint is enhanced and adapted to simplify the task of feature extraction;
c) feature extraction, in which the fingerprint is further processed to generate discriminativeproperties, also called feature vectors; and
d) Matching, in which the featurevector of the input fingerprint is compared against one or more existing templates.The templates of approved users of the biometric system, also called clients, areusually stored in a database. Clients can claim an identity and their fingerprints can be checked against stored fingerprints.

III. CLASSES OF FINGERPRINT

Fingerprint classification techniques have been a research topic for more than 30 years. Many classification methodshave been designed. However, most them use the same set of features: ridge line flow, orientation image, singular points,and Gabor filter responses.

Orientation image contain all necessary information to classify fingerprints in five broadclasses [6]. Galton-Henry classification system accounts for more than120 fingerprint classes. The five most common classes are:

- Arch: ridges enter from one side, rise to form a smallbump, and then go down and to the opposite side. No loops ordelta points are present.
- Tented Arch: similar to the arch except that at least oneridge has high curvature, thus one core and one delta points.

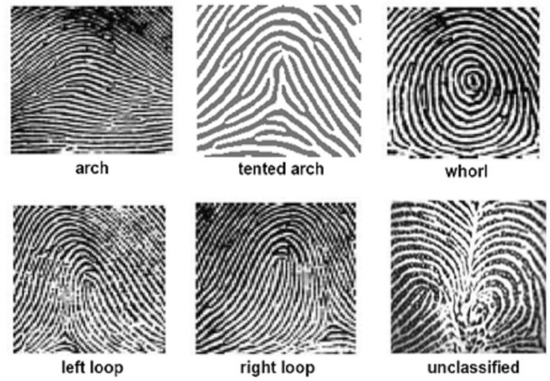


Figure 2 Classes of fingerprint

- Left loop: one or more ridges enter from one side, curve back, and go out the same side they entered. Core and delta are present [8].
- Right loop: same as the left loop, but different direction.
- Whorl: contains at least one ridge that makes a complete360 degree path around the center of the fingerprint. Twoloops (same as one whole) and two deltas can be found.
- Fingerprints in databases are non-uniformly distributed inthese classes.

Basic method of minutiae extraction

The basic method of minutiae extraction is divided in to three part Preprocessing, Minutiae Extraction, Post processing. Our method divides three basic steps in to 7 modules which are given below [7].

A) Step 1: Input-

In this step we take five finger print of a personas input and process them.

B) Step 2: Binarization:

This transform the 8-bit Gray fingerprint image to a 1-bit image with 0- value for ridges and 1-value for furrows.

C) Step 3: Thinning:

Ridge thinning is to eliminate the redundant pixels of ridges till the ridges are just one pixel wide.

1. To get a thinned image we find the location of middle black pixel at each stage of continuation of the curve.
2. In each scan of the full fingerprint image, the algorithm marks down redundant pixels in each small image window (3x3).
3. And finally removes all those marked pixels after several scans.

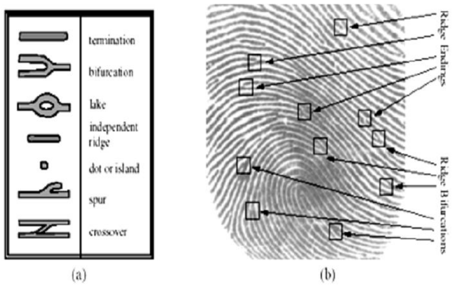


Figure 3 Different minutiae types, (b) Ridge ending & Bifurcation

*D) Step 4:*Minutiae Connect:

This operation takes thinned image as input and producesrefined skeleton image by converting small straight linesto curve to maximum possible extant.

*E) Step 5:*Minutiae Margin:

This increases the margin of endpoints by one pixel of curves of length at least three pixels.

F) Step 6: Minutiae point Extraction:

For extracting minutiae point we compute the number of one-value of every 3x3 window [4]:

- If the centroid is 1 and has only 1 one valued neighbor, then the central pixel is a termination.
- If the central is 1 and has 3 one-value neighbors, then the central pixel is a bifurcation.
- If the central is 1 and has 2 one-value neighbors, then the central pixel is a usual pixel.

G) Step 7: False Minutiae Removal

Procedure for removing false minutiae is given below:

- If the distance between one bifurcation and one termination is less than D and the two minutiae are in the same ridge. Remove both of them. Where D is the average inter-ridge width representing the average distance between two parallel neighboring ridges [5].
- If the distance between two bifurcations is less than D and they are in the same ridge, remove thetwo bifurcations.
- If two terminations are within a distance D and their directions are coincident with a small angle variation. And they suffice the condition that no any other termination is located between the two terminations. Then the two terminations are regarded as false minutia derived from a broken ridge and are removed.
- If two terminations are located in a short ridge with length less than D, remove the two terminations.

If a branch point has at least two neighboring branch points, which are each no further away than maximum distance threshold value and these branch points are closely connected on common line segment than remove the branch points.

And last we do the minutiae matching. Two fingerprint images to be matched, any one minutia is chosen from each image, and then the similarity of the two ridges associated with the two referenced minutia points is calculated. If the similarity is larger than a threshold, each set of minutiae to a new coordination system is transformed, whose origin is at the referenced point and whose x-axis is coincident with the direction of the referenced point. After we get two sets of transformed minutia points, we use the elastic match algorithm to count the matched minutia pairs by assuming two minutiae having nearly the same position and direction are identical.

IV. Pattern Matching Using Gabor Filters

- Quantized co-sinusoidal triplets
- Discrete Fourier transform
- Gabor filters

The different processing steps from pre-processing to matching as the final step of the fingerprint authentication. The first step is the normalization, which results in a better contrast of the fingerprint image. After that, the fingerprint is segmented, which crops areas of the recorded image, which do not contain any relevant information. This is the end of the pre-processing. The last pre-processing step usually consists of a fingerprint enhancement as described in [1]. However, tests have shown that the subsequent reference point detection works on non-enhanced fingerprint images as well as on enhanced. Therefore, any further enhancement is not required for the subsequent processing steps. After that, the fingerprint image is filtered using a Gabor filter. Now, it is possible to create the feature map, which is used as the template. This template is matched in the subsequent matching step with templates of other fingerprints. Theresult of the matching is the matching score, which represents how good two fingerprints resemble each other.

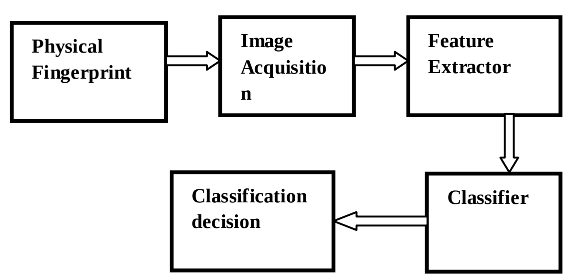


Figure 4. Building blocks for the Gabor approach

Most methods for fingerprint identification use minutiae as the fingerprint features. For small scale fingerprint recognition system, it would not be efficient to undergo all thepreprocessing steps (edge detection, smoothing, thinning..etc), instead Gabor filters will be used to extract features directly from the gray level fingerprint. No preprocessing stage is needed before extracting the features [1].

A. Image Acquisition

A number of methods are used to acquire fingerprints.Among them, the inked impression method remains the most popular one. Inkless fingerprint scanners are also present eliminating the intermediate digitization process. Fingerprint quality is very important since it affects directly the minutiae extractionalgorithm [4]. Two types of degradation usually affect fingerprint images: 1) the ridge lines arenot strictly continuous since they sometimes include small breaks (*gaps*); 2) parallel ridgelines are not always well separated due to the presence of cluttering noise. The resolutionof the scanned fingerprints must be 500 dpi while the size is 300x300.

B. Feature Extractor

Gabor filter based features have been successfully and widely applied to facerecognition, pattern recognition and fingerprint enhancement. The family of 2-D Gaborfilters was originally presented by Daugman (1980) as a framework for understanding theorientation and spatial frequency selectivity properties of the filter. The fingerprint print image will be scanned by a 8x8 window; for each block the magnitude of the Gabor filter is extracted with different values of m (m = 4 and m = 8). The features extracted (new reduced size image) will be used as the input to the classifier.

C. Classifier

The classifier is based on the k-nearest neighborhood algorithm KNN. “Training”of the KNN consists simply of collecting k images per individual as the training set. The remaining images consists the testing set. The classifier finds the k points in the training set that are the closest to x (relative to the Euclidean distance) and assigns x the label shared by the majority of these k nearest neighbors. Note that k is a parameter of the classifier; it is typically set to an odd value in order to prevent ties.

The last phase is the verification phase where the testing fingerprint image:
1) is inputted to the system
2)magnitude features are extracted
3) perform the KNN algorithm
4) Identify the person

V. Measuring Biometric Accuracy

One of the most important factors in the success of a biometric system is its accuracy. This is a measure of how well the system is able to correctly match the biometric information from the same person and avoid falsely matching biometric information from different people. The measurement of biometric accuracy is usually expressed as a percentage or proportion, with the data coming from simulations, laboratory experiments, or field trials. There are four main measures of biometric accuracy [1]:

A. True Acceptance Rate (TAR) / True Match Rate (TMR): this measure represents the degree that the biometric system is able to correctly match the biometric information from the same person. Developers of biometric systems attempt to maximize this measure.

B. False Acceptance Rate (FAR) / False Match Rate (FMR): this measure represents the degree or frequency where biometric information from one person is falsely reported to match the biometric information from another person. Developers attempt to minimize this measure [2].

C. True Rejection Rate (TRR) / True Non-Match Rate (TNMR): this measure represents the frequency of cases when biometric information from one person is correctly not matched to any records in a database because, in fact, that person is not in the database. Developers attempt to maximize this measure.

D. False Rejection Rate (FRR) / False Non-Match Rate (FNMR): this measure represents the frequency of cases when biometric information is not matched against any records in a database when it should have been matched because the person is, in fact, in the database. Developers attempt to minimize this measure.

The accuracy of a fingerprint matching algorithm is measured by:

- FAR: It stands for false acceptance rate. It is defined as the ratio of the number of impostor images considered as authentic by the algorithm to the total number of impostor images.
- FRR: It stands for false rejection rate. It is defined as the ratio of the number of authentic images not considered qualified by the algorithm to the total number of authentic images.

VI. Fingerprint Recognition System Performance Evaluation

Even if fingerprint recognition system is a matured biometric recognition it certainly makes wrong decisions. Some of the reasons are [3]:

• Information limitation: for example due to poor quality finger print images.
• Representation limitation: i.e. features used to represent the fingerprint image.
• Invariance limitation: i.e. images of different fingerprint may look similar.

Figure5 Database of fingerprint

The following method is common for evaluating performance of fingerprint recognition system which is adapted from FVC 2002 report [2]. Let the database contain 8 samples of 10 different fingers. Fs_{ij} is the j^{th} fingerprint sample of the i^{th} fingerprint and its corresponding template Ts_{ij}.

A. The templates are Ts_{ij} computed from the corresponding fingerprint samples and stored in another database.

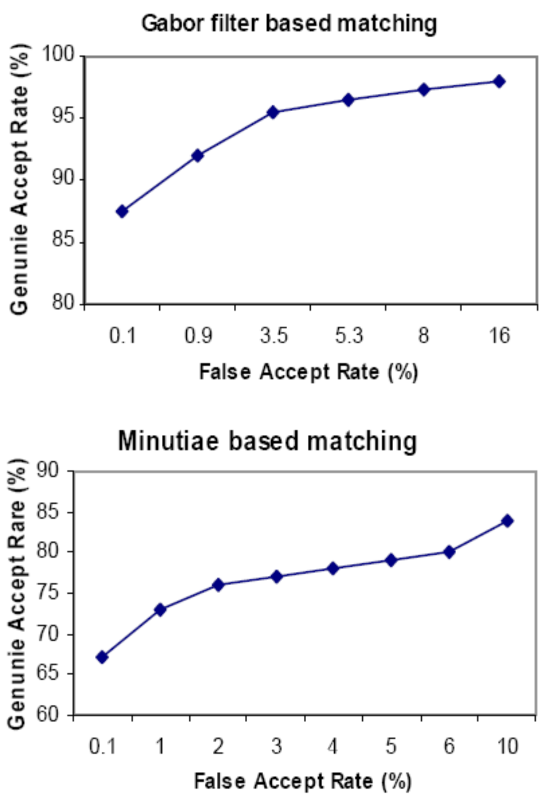


Figure 6 comparing the performance of the gabor filter and minutiae matching.

B. For each finger, each sample is matched against each other (i.e. the samples must be from the same finger). If Ts_{ij} is the template it will be matched with fingerprint Fs_{ik} ($i < k \le 8$) and the scores called genuine matching scores (gms_{ik}) are stored. The number of genuine recognition attempt (NGRA) is $\frac{7\times 8}{2}\times 10 = 280$

C. The first samples from each fingerprints matched against each other (i.e. the samples are from different fingers). If Ts_{i1} is the sample template it will be matched with fingerprint image Fs_{ki} ($i< k \le 10$) and the scores called impostor matching scores (ims_{ik}) will be stored. The number of impostor recognition attempt (NIRA) is $\frac{9\times 10}{2} = 45$.

D. The FMR(t) (False Match Rate, also sometimes called False Acceptance Rate) and FNMR(t) (False Non-Match Rate, also sometimes called False Rejection Rate) curves are computed for threshold t ranging from 0 to 1. Given a threshold t, FMR (t) denotes the percentage of imsik $\le$ t (i.e percentage of the number of samples that are matched even if they are from different fingers), whereas FNMR(t) denotes the percentage of $gms_{ijk} < t$ (i.e. percentage of the number of samples that are not matched even if they are from the same finger).

$$\mathbf{FMR}(t) = \frac{\{ims_{ik} \mid ims_{ik} \ge t\}}{\mathbf{NIRA}},$$

$$\mathbf{FNMR}(t) = \frac{\{gms_{ijk} \mid gms_{ijk} < t\}}{\mathbf{NGRA}}$$

VII. CONCLUSION:

Image quality is related directly to the ultimate performance of automatic fingerprint authentication systems. Good quality fingerprint images need only minor preprocessing and enhancement for accurate feature detection algorithm. For small scale fingerprint recognition system, it would not be efficient to undergo all the preprocessing steps (edge detection, smoothing, thinning etc. as like of minutiae based technique), instead Gabor filters will be used to extract features directly from the gray level fingerprint. The Gabor filter method is widely accepted approach for the fingerprint matching. It is quite suitable for fingerprint matching.Eight Gabor filters are used toextract features from the template and input images. Theprimary advantage of our approach is improved translationand rotation invar

ACKNOWLEDGMENT

I like to extend my thanks to Dr. V. K. Banga HoD ECE ACET who guided me throughout the preparation of the paper and gave me relevant concept to materialize the paper. I would also like give my humble regards to Mr. Sanjeev kumar Professor ACET for helping me to shape out the important findings and necessary reseach for the paper. I wish both them all the success for future.

REFERENCES

1. Federal Bureau of Investigation, The Science of Fingerprints: Classification and Uses, U. S. Government Printing Office, Washington, D. C., 1984.
2. Maio D., Maltoni D., Cappelli R., Wayman J.L. and Jain A.K., “FVC2002: Second Fingerprint Verification Competition,” in Proc. Int. Conf. on Pattern Recognition 16th) , 2002b.
3. Maltoni D., Maio D., Jain A.K. and prabhakar S., “Handbook of Fingerprint Recognition,” Second Edition, Springer , 2009.
4. Manvjeet Kaur, Mukhwinder Singh, Akshay Girdhar, and Parvinder S. Sandhu, “ Fingerprint Verification System using Minutiae Extraction Techn ique”, Proceedings of World Academy of Science, Engineering and Technology vol. 36 , pp. 497-502, (2008).
5. Minutiae Extraction from Fingerprint Images - a Review Roli Bansal, Priti Sehgal and Punam Bedi IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 3, September 2011.
6. http://www.fmrib.ox.ac.uk/~steve/susan/fingerprint/node2.html
7. http://www.fmrib.ox.ac.uk/~steve/susan/minutiae/node2.html
8. Maltoni D., Maio D., Jain A.K. and prabhakar S., “Handbook of Fingerprint Recognition,” Second Edition, Springer , 2009.
9. NTSC Subcommittee on Biometrics, "Fingerprint Recognition", 2000.

Analysis of Hidden Markov Model and Conditional Random Field Approach for Bio-Imaging

SilviGoel[1,] Jaspreet Kaur[2], Sonam[3], Harpreet Kaur[4], Ramanjyot Kaur[5]
Sant Baba Bhag Singh Institute of Engineering & Technology, Jalandhar
Email: goelsilvi@yahoo.com, erharpreetarora@gmail.com, er.ramanjyot@gmail.com

Abstract- "Bioimaging" is related to methods that non-invasively visualize biological processes in real time. Bioimaging aims to interfere as little as possible with life processes. It is often used to gain information on the 3-D structure of the observed specimen without physical interference. Recently, the field of bioinformatics has experienced rapid growth. However, as with other young disciplines, it now faces a host of critical issues such as development of new algorithms and statistics with which to assess relationships among members of large data sets, analyses and interpretation of various types of sequences, domains, and structures, development and implementation of tools that enable efficient access and management of different types of information[2-4]. This paper takes into account the algorithms for efficiently carrying out the process of Bioimaging. The first discussed algorithm is Hidden Markov Model and second is Conditional Random Field Approach.

1. Introduction

"Bioinformatics" is defined by the National Institutes of Health as the "research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data[5]. The exponential growth in the amount of such data has necessitated the use of computers for information cataloging and retrieval, while a more global perspective in the quest for new insights into health and disease and the resulting data mining also underscore the need for bioinformatics.

It finds it usage in protein structure prediction, homology searches, multiple alignment and phylogeny construction, genomic sequence analysis and gene-finding.[1]

2. Methods of Bioimaging

2.1Hidden Markov Model

2.1.1 Introduction

A process is said to be Markovian if the next state is independent of all previous states, given the current state. In bioinformatics a lot of phenomena are perceived to be, or at the very least modelled as, markovian in nature. [7] A few examples include:

- nucleotide substitution models, where the rate of change to a particular nucleotide may depend on the current nucleotide but does not depend on nucleotides previously occupying the position.
- the insertion/deletion model of statistical alignment, where the birth at a link or death of a nucleotide does not depend on the history of the link or nucleotide.
- secondary structure at a particular position in a protein sequence.
- the coding state of a position in a genome, i.e. whether the nucleotide is part of a protein coding gene and the codon position it occupies.
- the network growth by gene duplication.

Markov models can be continuous, both in terms of transition structure and in terms of state space. An obvious example of continuous time Markov models are nucleotide substitution models, where the time at which the next substitution occurs is drawn from an exponential distribution. The Wiener process, also often called Brownian motion, combines a continuous time transition structure with a continuous state space. [7]

2.1.2 State Transitions

We may also want to use special states representing deletions of one or more positions. The deletion states corresponds to no character being observed for an ancestral position, so no character should be emitted from these states. Alternatively we could have added the possibility of no emission from the states matching ancestral or consensus positions. However, quite often we would expect a different transition distribution, or even a different transition structure, when no emission occurs. States with no emissions are generally denoted *silent* states, and all other states are *non-silent* states.

Formally, we describe an HMM M with a tuple (Q, ∑, a, e, start, end) where :
• Q is the set of states in the model.
• ∑ is the set of possible observations
• $a_{p,q}$ describes the transition probability between states p and q in Q. The transitions of M are usually taken to mean the pairs $p \rightarrow q$ with $a_{p,q} > 0$, a state can have a self-transition, i.e. $a_{p,p} > 0$
• $e_{p,\sigma}$ for $p \in Q$ and $\sigma \in \sum$ describes the emission probability of symbol σ when entering state p,
we will usually assume that states can be classified as either silent, where $e_{p,\sigma} = 0$ for all $\sigma \in \sum$, or non-silent, where $\sum_{\sigma \in \sum} e_{p,\sigma} = 1$
• start, end $\in$ Q are designated start and end states of the model.[7]

The transition probabilities are required to specify a probability distribution over outgoing transitions for each state p, i.e. $\sum_{q \in Q} e_p, q_\sigma = 1$, except for the end state for which $a_{end,q} = 0$ for all $q \in Q$. We will also usually assume that $a_{q,start} = 0$ for all $q \in Q$, i.e. that there are no transitions to the start state, and that both start and end are silent states. If this is not the case, we can always add a new start (end) state that only has a transition with probability 1 to (from) the original start (end) state. Finally, we will assume that from any state $q \in Q$ the probability of reaching the end state in a finite number of steps is 1, such that M generates a finite sequence with probability 1. The best way to view an HMM is probably as a generative model: we start in the start state, continue choosing the next state according to the transition probability of the current state, emit a character to an expanding sequence whenever we enter a non-silent state. We continue this process until we enter the end state, at which point we stop and output the constructed sequence. While enjoying wide historical success, standard HMM models have difficulty modeling multiple non-independent features of the observation sequence.[7]

2.2 Conditional Random Model

2.2.1 Introduction

Conditional random fields are undirected graphical models trained to maximize a conditional probability (Lafferty et al., 2001). A common special-case graph structure is a linear chain, which corresponds to a finite state machine, and is suitable for sequence labeling. The CRF approach draws together the advantages of both finite state HMM and discriminative SVM techniques by allowing use of arbitrary, dependent features and joint inference over entire sequences.[6]

2.2.2 State Transitions

In CRFs, state transitions are also represented as features. The feature function $f_k(y_{t-1},y_t,x,t)$ is a general function over states and observations. Different state transition features can be defined to form different Markov-order structures. We define four different state transitions features corresponding to different Markov order for different classes of features. Higher order features model dependencies better, but also create more data sparse problem and require more memory in training.[6]

a. First-order: Here the inputs are examined in the context of the current state only. The feature functions are represented as $f(y_t,x)$. There are no separate parameters or preferences for state transitions at all.

b. First-order+transitions: Here we add parameters corresponding to state transitions. The feature functions used are $f(y_t,x)$, $f(y_{t-1}, y_t)$.

c. Second-order: Here inputs are examined in the context of the current and previous states. Feature function are represented as $f(y_{t-1}, y_t,x)$.

d. Third-order: Here inputs are examined in the context of the current, two previous states. Feature function are represented as $f(y_{t-2},y_{t-1}, y_t,x)$.[6]

3. Comparison of HMM and CRF

We first report the overall results by comparing CRFs with HMMs. The results we obtained with CRFs use second order state transition features, layout features, as well as supported and unsupported features. Feature induction is used in experiments on dataset **R**; (it didn't improve accuracy on **H**). The results we obtained with the HMM model use a second order model for transitions, and a first order for observations. The results on SVM is obtained from (Han et al., 2003) by computing F1 measures from the precision and recall numbers they report.[6]

From Table (1, 2), one can see that CRF performs significantly better than HMMs. CRFs increase the performance on nearly all the fields. The overall word accuracy is improved from 92.9% to 98.3%, which corresponds to a 78% error rate reduction. However, as we can see word accuracy can be misleading sinceHMMmodel even has a higher word accuracy than SVM, although it performs much worse than SVM in most individual _elds except *abstract*. Interestingly, HMM performs much better on *abstract* field (98% versus 93.8% F-measure) which pushes the overall accuracy up. A better comparison can be made by comparing the field-based F-measures. Here, in comparison to the SVM, CRFs improve the F1 measure from 89.7% to 93.9%, an error reduction of 36%.

	HMM		CRF	
Overall acc.	93.1%		98.3%	
Instance acc.	4.13%		73.3%	
	Acc.	F1	Acc.	F1
Title	98.2	82.2	99.7	96.5
Author	98.7	81.0	99.8	97.2
Affiliation	98.3	85.1	99.7	93.8
Address	99.1	84.8	99.7	94.7
Note	97.8	81.4	98.8	81.6
Email	99.9	92.5	99.9	91.7
Date	99.8	80.6	99.9	90.2
Abstract	97.1	98.0	99.6	93.8
Phone	99.8	53.8	99.9	92.4
Keyword	98.7	40.6	99.7	88.5
Web	99.9	68.6	99.9	92.4
Degree	99.5	68.8	99.8	70.1
Pubnum	99.8	64.2	99.9	89.2
Average F1		75.6		89.7

Table1: Extraction results for paper headers on H

	HMM		CRF	
Overall acc.	85.1%		95.37%	
Instance acc.	10%		77.33%	
	Acc.	F1	Acc.	F1
Author	96.8	92.7	99.9	99.4
Booktitle	94.4	0.85	97.7	93.7
Date	99.7	96.9	99.8	98.9
Editor	98.8	70.8	99.5	87.7
Institution	98.5	72.3	99.7	94.0
Journal	96.6	67.7	99.1	91.3
Location	99.1	81.8	99.3	87.2
Note	99.2	50.9	99.7	80.8
Pages	98.1	72.9	99.9	98.6
Publisher	99.4	79.2	99.4	76.1
Tech	98.8	74.9	99.4	86.7
Title	92.2	87.2	98.9	98.3
Volume	98.6	75.8	99.9	97.8
Average F1		77.6%		91.5%

Table2: Extraction results for paper references on R

References

[1] D. Searls, “Grand Challenges in Computational Biology,” *Computational Methods in Molecular Biology*, S. Salzberg, D. Searls, and Simon Kasif, eds., Elsevier Science, 1998.

[2] Attwood, Teresa K. "Genomics. The Babel of Bioinformatics." *Science* 290, no. 5491 (2000): 471-473.

[3] Gibson, Greg and Spencer V. Muse. A Primer on Genome Science. Sunderland, MA: Sinauer Associates, Inc., Publishers, 2002.

[4] Thornton, Janet M. "From Genome to Function." Science 292, no. 5524 (2001): 2095-2097.

[5] Available at the National Center for Biotechnology Information's Web site at www.ncbi.nlm.nih.gov.

[6] P.Fuchun, M.Andrew, “Accurate Information Extraction from Research Papers using Conditional Random Fields”,2004

[7] L.Rune, “Hidden Marlov Models”, 2012

Diagnosis of Stress Involving Acquisition of Biological Signals For Example Heart Rate, Electrocardiogram, Electromyography Signals.

DR. V.K. BANGA[1], ER. SATISH PAWAR[2]
[1]Professor, HOD(ECE/EEE), Amritsar College of Engineering & Technology, India
[2] M.TECH (ECE), Amritsar College of Engineering & Technology, India
Email: vijaykumar.banga@gmail.com, satishpwr75@gmail.com

Abstract - Diagnosis of stress is important because it can cause many diseases e.g., heart disease, headache, migraine, sleep problems, irritability etc. Diagnosis of stress in patients often involves acquisition of biological signals for example heart rate, electrocardiogram (ECG), electromyography signals (EMG) etc. Stress diagnosis using biomedical signals is difficult and since the biomedical signals are too complex to generate any rule an experienced person or expert is needed to determine stress levels. Also, it is not feasible to use all the features that are available or possible to extract from the signal. So, relevant features should be chosen from the extracted features that are capable to diagnose stress. Electronics devices are increasingly being seen in the field of medicine for diagnosis, therapy, checking of stress levels etc. The research and development work of medical electronics engineers leads to the manufacturing of sophisticated diagnostic medical equipment needed to ensure good health care. Biomedical engineering combines the design and problem solving skills of engineering with medical and biological sciences to improve health care diagnosis and treatment.

Keywords: Mental stress, heart rate, heart rate variability

I. INTRODUCTION

Stress is popularly known as the state when a person fails to react properly to the emotional or physical threats whether imaginary or real. It shows symptoms like exhaustion, headache, adrenaline production, irritation, muscular tension and elevated heart rate. When our brain appraises stress, the Sympathetic Nervous System (SNS) prepares our brain to respond to stress. The beat to beat intervals of the heart tend to vary. This gives rise to Heart Rate variability (HRV) which is mostly regulated by the sympathetic and parasympathetic Autonomic Nervous Systems (ANS). Thus the state of the ANS is reflected in HRV. This has made it an increasingly popular tool to investigate the state of the ANS, which can be further used to explain various physiological activities of the body. Heart rate variability (HRV) is one of the popular parameters to analyze the activities of the Heart and the Autonomic Nervous System (ANS) in humans. The state of the ANS is reflected in the HRV. For this reason we chose HRV as one of the key criteria to diagnose stress. The research and development work of medical electronics engineers leads to the manufacturing of sophisticated diagnostic medical equipment needed to ensure good health care. Biomedical engineering combines the design and problem solving skills of engineering with medical and biological sciences to improve health care diagnosis and treatment. Chronic stress, such as experienced in working situations, can lead to a chronic activation, overload and eventually exhaustion of the hormonal, cardiovascular, neural and muscular systems due to insufficient recovery and repair. Long term consequences include, for example, an impairment of immune systems, delay of healing processes and musculoskeletal overload. With this study, we aim to investigate the interaction between HR, HRV and mental stress on group level and for individual changes. In this test ECG was measured during rest and during a mental task. We analyzed HRV to provide insight into how the heart reacts to a mental task. Linear HRV measures were included and finally neural network has been implemented in this paper. The diagnosis is mainly based on the expert's experience but the number of expert is also insufficient. In such a scenario a case base reasoning (CBR) system can be considered as a way of stress diagnosis. In CBR the nature of new problem is identified and solved respectively by reusing the previous problems and their solutions that are categorized and solved by expert. The HRV can be analyzed using both time domain and frequency domain attributes. It is very important to choose features which vary with the changes of the stress levels and show relatively reliable behavior. We collect as many cases as possible in our case library for which the expert has defined the stress level

2. Objectives

The main objective of the research paper is to construct a network using HRV features to diagnose stress and analyze how accurate the result is compared to the expert's diagnosis. A central theme is to improve patients' self efficacy. Self-efficacy refers, for instance, to the patient having the correct information, skills and motivation to do self-care. Thus, the emphasis is moving from improving patients' compliance to improving patients' self-efficacy. The patient-centric care has been shown to be beneficial, for example, in terms of changes in the health status, patients' self-efficacy and health care resource utilization. Patient's Stress classification and detection helps the medical system to reduce the serious problem affect of many people of different professions life situations, and age groups. High stress rates of the patients suffer from the Heart disease and other physiological outcomes, Psychological disease and social and behavioral changes wireless. This thesis will explain different methods for classification of stress levels and help the patients to overcome from that stress. So the main focus of the thesis is to create a conjugate gradient network to find out the stress level of the patients. All the simulations are performed using MATLAB Software.

3. Sensors

In biomedicine and biotechnology, sensors are analytical device which converts a biological response into an electrical signal. Sensors are used to collect the data from patients for classification of stress. The activities at Biosensors Group aim at the integration of nanotechnology methods, tools and materials into low cost, user friendly and efficient sensors and biosensors with interest for several fields such as diagnostics, food analysis, environment monitoring and other industries.

3.1 Transmitter and Receiver sensor

Sensor transmitters are measurement or signal conditioning packages that provide standard, calibrated outputs from sensors or transducers. Outputs types include current loops, variable voltage levels, frequency or pulse signals, timers or counters, relays, and variable resistance outputs.

Fig. 1 Transmitter and Receiver sensor.

Fig. 2 Receiver port of sensor attached.

Fig.2 shows the receiver port of sensor which is attached with laptop for collection of patient's data. The power consumption is critical, especially in the case of long-term monitoring using wearable sensors. Traditionally, most of the studies on automatic health monitoring in uncontrolled environments focus on health monitoring at home. These solutions use mostly embedded sensors that are available in the home environment, either as separate devices or embedded into structures such as furniture. More recently, with advances in electronics and communication technology, wearable sensors have attracted increased attention.

4. Experiment Setup

My aim in this research is the detection of mental stress, as physical stressors occur far less frequently in the context of human-computer interaction. Therefore, in order to elicit mental stress at controlled intervals a computerized "Conjugate gradient network" was used. The development of this recognition system involved three stages: experiment setup for physiological sensing, signal preprocessing for the extraction of affective features and affective recognition using a learning system.

4.1 Data collection from sensors

The purpose of the data collection was to use different wearable sensors during mental and physical load and rest and to identify the most useful sensors, feature extraction methods and classification methods for automatic recognition of those activities. This study was conducted in real-life settings. The data collection equipment consisted of ten hardware units: 1) wireless sensor node; 2) HR monitor. Data from activity monitor worn on the wrist, movement sensor in bed, HR monitor, and wireless sensor node with temperature and illumination sensors were sent.

Fig.3 Data collection of a patient

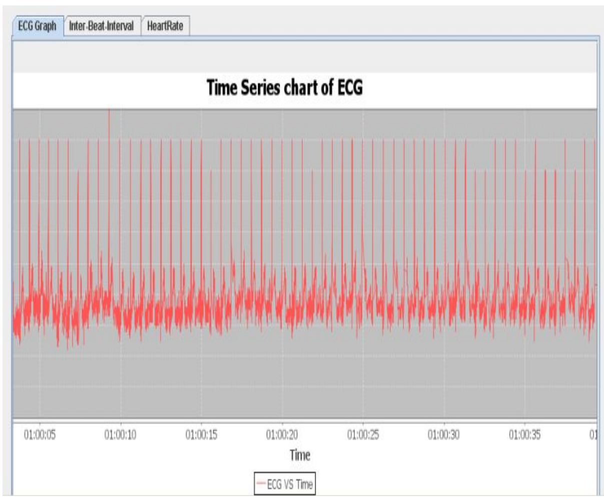


Fig.4 Online graph of ecg system

This is the online graph of a ecg system obtained after collection of data using Matlab for classification of stress.

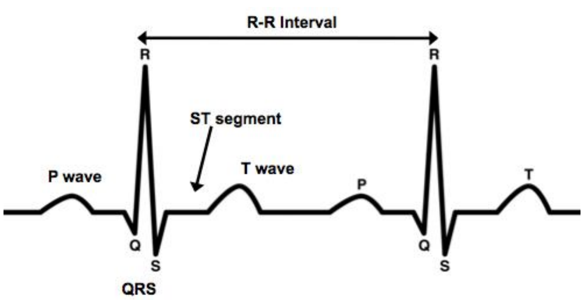
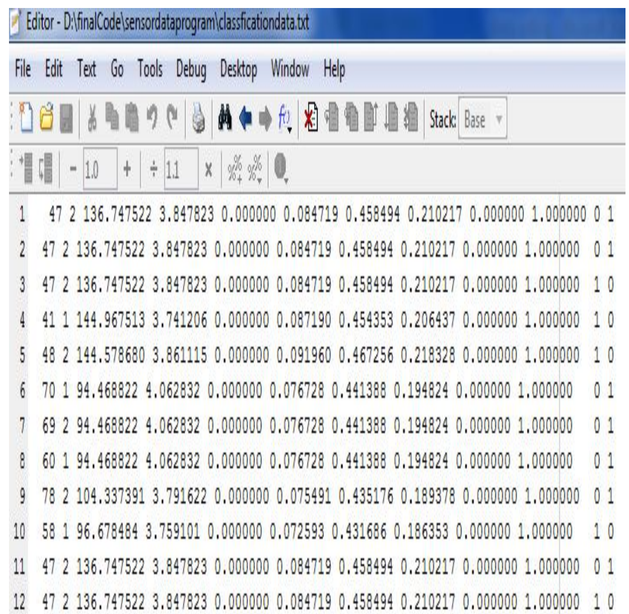


Fig.5 Heart Beat

The first little hump is known as the P wave. It occurs when the atria depolarize (i.e. trigger). The next three waves constitute the QRS complex. They represent the ventricles depolarizing. These three are lumped together because a normal rhythm may not have all three. Many times, you'll only see a R and an S. This is not abnormal. If there are less than three, how do we know which one is which? Well, the R wave is the first wave ABOVE the isoelectric line. You then name the waves in relation to the R wave. If it falls before the R wave, it is called the Q wave; after the R wave is the S wave.

4.2 Make data base of the patients

For collecting the data base of the patients we have to make an formula p1 = s (7200*2:7200*3) for generating data base of the patient. No of participants were monitored, some of men and women with mean age of 22 (±1.96) and an average body mass index of 22.2 (±0.43). Participants were students of lovely professional university. Measurements were recorded for two conditions for each subject: with and without a mental task. HR was recorded throughout the test for each subject. HRV was calculated as the variance in time between two consecutive R-peaks. Different measures in time and frequency domain were used to express HRV.



1	47	2	136.747522	3.847823	0.000000	0.084719	0.458494	0.210217	0.000000	1.000000	0	1
2	47	2	136.747522	3.847823	0.000000	0.084719	0.458494	0.210217	0.000000	1.000000	0	1
3	47	2	136.747522	3.847823	0.000000	0.084719	0.458494	0.210217	0.000000	1.000000	1	0
4	41	1	144.967513	3.741206	0.000000	0.087190	0.454353	0.206437	0.000000	1.000000	1	0
5	48	2	144.578680	3.861115	0.000000	0.091960	0.467256	0.218328	0.000000	1.000000	1	0
6	70	1	94.468822	4.062832	0.000000	0.076728	0.441388	0.194824	0.000000	1.000000	0	1
7	69	2	94.468822	4.062832	0.000000	0.076728	0.441388	0.194824	0.000000	1.000000	0	1
8	60	1	94.468822	4.062832	0.000000	0.076728	0.441388	0.194824	0.000000	1.000000	0	1
9	78	2	104.337391	3.791622	0.000000	0.075491	0.435176	0.189378	0.000000	1.000000	0	1
10	58	1	96.678484	3.759101	0.000000	0.072593	0.431686	0.186353	0.000000	1.000000	1	0
11	47	2	136.747522	3.847823	0.000000	0.084719	0.458494	0.210217	0.000000	1.000000	0	1
12	47	2	136.747522	3.847823	0.000000	0.084719	0.458494	0.210217	0.000000	1.000000	1	0

Fig. 6 Screenshots of the data collection

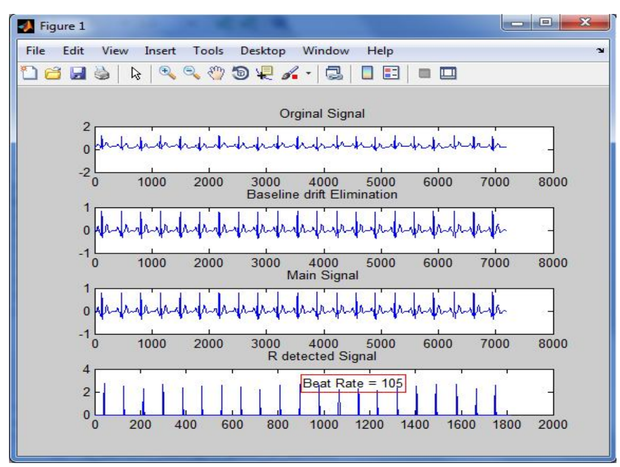


Fig. 7 Heart beat rate of 1 patient

% 1 for Female
% 2 for Male
Age= 89 ;
Sex Code=1;
MakeClassificationData1(Age,SexCode,beat_Rate,high,lowR,meanR,stdR,varR,mediaR,covR);

4.3 Data Analysis

Average heart beat period in ms (Mean RR) was calculated for each subject for the two conditions derived from the raw RR-interval. The Pan-Tompkins algorithm was used to detect the QRS-complexes in the ECG-signal from which we could determine the RR-intervals. HRV analysis is possible in the time or frequency domain. Time domain analysis is the easiest way, calculated directly from the raw RR-interval. The frequency domain shows us the variability of the RR-signal over time by looking at the proportion of the frequencies relative to the original RR-signal. Frequently misused clip class parameters are link and standardized divagation (SD) of RR, think and SD of HR, RMS. It utilised spectral measures are apex frequency and index of real low cardinal bands (VLF: 0 to .04 Hz), low frequency bands (LF: .04 to .15 Hz) the ratio of LF/HF, which is interpreted as a measure of sympathovagal balance. . We have to analyse the data in the given format. Make Classification Data (Age,SexCode, beat_Rate,high,lowR,meanR, stdR ,varR, mediaR, covR). For generation of patients data we have to use the given formula, p1 = s (7200*p1:7200*p2). We calculated LF and HF components for each subject in each condition, and the LF/HF ratio was determined for further analysis. After that we have to use neural network on whole data of patients. We use neural network because it is a kind of machine learning in which patients data from various resources can be compiled in one learning experience and can be utilized to aid doctor and proper identification of high and low stress people.

4.4 Feature Extraction

Feature extraction aims at extracting such characteristics of the input patterns that enable their classification into distinct classes. In the analysis of personal health monitoring data, the features are computed from the content of a sliding window. Feature extraction methods commonly used to study signal characteristics include 1) time-domain features and 2) frequency domain features.

4.5 Stress Classification

We used the WEKA machine learning engine to train classifiers using various learning methods, including the J48 Decision Tree, Bayes Net, and support vector machine (SVM) for stress inference . We divided the training data into two different sets in order to evaluate how activity information may influence the results of stress inference.

5. Results

The system was evaluated with maximum no of case representations. The cases were classified in to 2 stress levels viz. 1 and 2. The levels of stress 1 and 2 are categorized as low and high stress resp

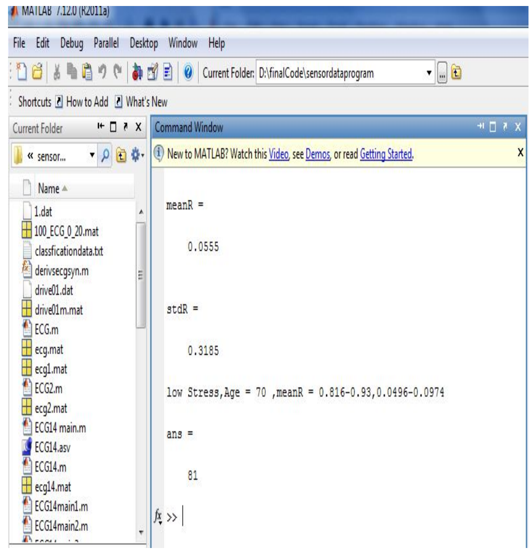


Fig8. Output data of low stress

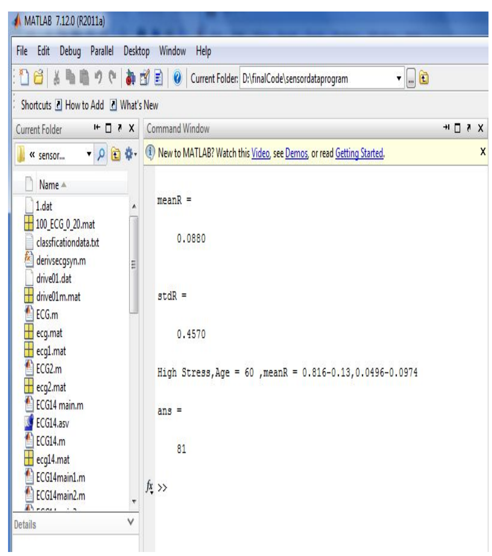


Fig.9 Output data of high stress.

The diagnosis of stress is mainly done by on careful analysis on biological signals like heart rate, finger temperature, electrocardiogram (ECG), electromyography signal (EMG), skin conductance signal (SC) etc. The experts use their experience and human reasoning system to find out a pattern from the signal to diagnose the stress and the number of such kinds of experts is very less. So, a decision support system will be useful to diagnose stress instead of expert. At last confusion matrix has been generated.

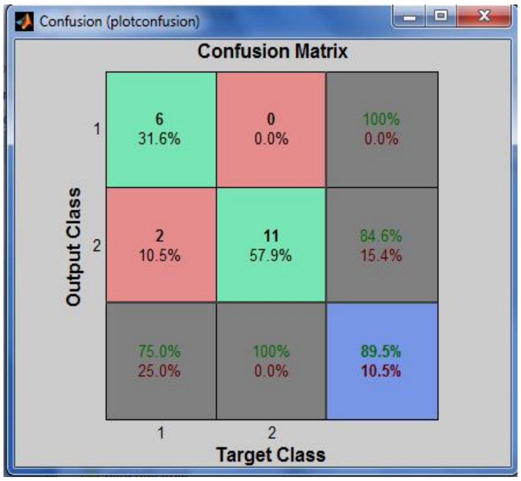


Fig.10 Confusion Matrix

The confusion matrix after personalization shows an improved accuracy. The accuracy after Personalization is 88%. Much of our classifiers' confusion seen in the results can be explained with transitions from one activity to another. The annotator was not given the choice to annotate "transition," but he had to switch from one activity to another instantly at some point during the transition. The resulting inaccuracy is especially visible in the recognition of stress level, which should be detected almost perfectly from the conjugate gradient network. In the field of artificial intelligence, a confusion matrix is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one (in unsupervised learning it is usually called a matching matrix). Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. The name stems from the fact that it makes it easy to see if the system is confusing two classes (i.e. commonly mislabeling one as another). Outside artificial intelligence, the confusion matrix is often called the contingency table or the error matrix. . Low and High stress cases are classified by the expert system based on the matlab and on the following parameters i.e age, sex code, beat rate, meanR, stdR, mediaR, covR, Stress. Finally all cases are classified after all these signals taking in consideration. The classification are shown in table 1.The use of the word validation here should not be confused with the high stress and low stress. The validation set is used to further refine the neural network construction. The testing set is then used to determine the performance of the neural network by computation of an error metric. This training-validating-testing approach is the first, and often the only, option system developers consider for the assessment of a neural network. The assessment is accomplished by the repeated application of neural network training data, followed by an application of neural network testing data to determine whether the neural network is acceptable.

5.1 Best validation performance

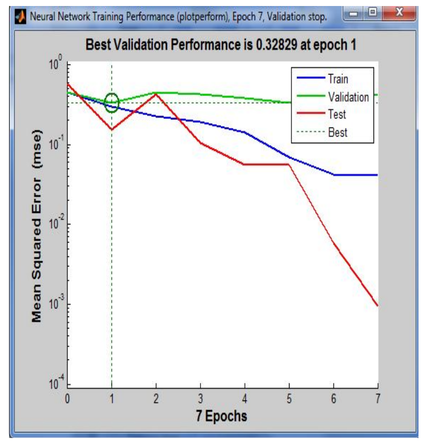


Fig. 12 Neural Network Traning Network

5.2 Error Histogram

Histograms are used to plot density of data, and often for density estimation: estimating the probability density function of the underlying variable. The total area of a histogram used for probability density is always normalized to 1. If the length of the intervals on the x-axis is all 1, then a histogram is identical to a relative frequency plot. There is no "best" number of bins, and different bin sizes can reveal different features of the data.

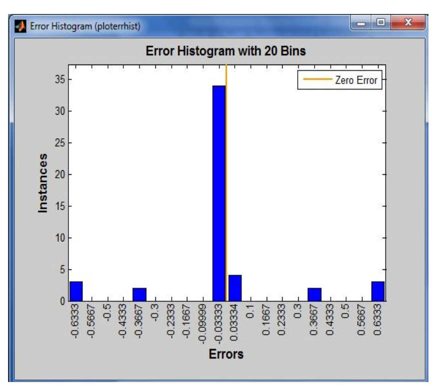


Fig. 13 Error Histogram

Age	Sex Code	Beat Rate	High R	Low	meanR	stdR	stdR	mediaR	covR	Stress
41	1	144.967513	3.741206	0.000000	0.087190	0.454353	0.206437	0.000000	1.000000	1
48	2	144.578680	3.861115	0.000000	0.091960	0.467256	0.218328	0.000000	1.000000	1
70	1	94.468822	4.062832	0.000000	0.076728	0.441388	0.194824	0.000000	1.000000	2
89	2	99.047452	2.719985	0.000000	0.055332	0.316915	0.100435	0.000000	1.000000	2
58	1	96.678484	3.759101	0.000000	0.072593	0.431686	0.186353	0.000000	1.000000	1
78	2	104.337391	3.791622	0.000000	0.075491	0.435176	0.189378	0.000000	1.000000	2
99	1	144.967513	3.888144	0.000000	0.086690	0.456807	0.208673	0.000000	1.000000	2

Fig. 11 Table 1

Stress
Low stress – 1
High stress - 2

And

Sex code
Female- 1
Male- 2

Conclusion

This new technology has potential to offer a wide range of benefits to patients, medical personnel, and society through continuous monitoring in the ambulatory setting, early detection of abnormal conditions, supervised rehabilitation, and potential knowledge discovery through data mining of all gathered information. To formulate rule for stress diagnosis by using ECG signal is hard because of the non stationary and uncertain characteristics of the ECG signal. So, the stress diagnosis is mostly dependent on the expert knowledge but the number of available experts is also insufficient. The features selection is done according to the study where a survey was done on most of the papers to find out the relative and important features. Most of the works in the related field are done on the basis of time and frequency domain. Both time and frequency domain features, the range of some power spectral that are features and the normalization equations were selected carefully by the study.

Future scope

There are many other classifier for diagnosis of stress i.e. support vector machine, Bayesian network etc which can also be tested on the similar work which can yield better accuracy and research. The database acquisition was based on psychological experiments carried out by expert psychologists. These experiments ensure that stressing situations are provoked on an individual, validating posterior HR and HRV acquisitions. This paper provides a decision system able to detect stress with an accuracy of 92.5% using Neural network. In future we have to detect maximum accuracy using other networks. The results show moderate, but significant overall correlations to daily stress level on working days and pateint. Even moderate correlations can be considered important in case of these kinds of data . However, it must be kept in mind that many of the variables (e.g., BP variables) are not independent. Thus, when the value of one variable changes, the dependent variable changes as well. Strong correlations can be found at individual level, but the correlations of different individuals can even be conflicting, which weakens the overall correlations. Thus, based on the results of this study, it can be said that stress is an individual phenomenon. Different people react to stress in different ways. One person may react with blood pressure, while another reacts with disturbed sleep, etc. Thus, finding very specific variables that always indicate stress, when it is present, is a demanding task for future research.

REFERENCES

[1] Klein CL & Corwin EJ Seeing the unexpected: how sex differences in stress responses may provide a new perspective on the manifestation of psychiatric disorders. Current Psychiatry Reports 4:441-448, Department of Biobehavioral Health, Pennsylvania State University, 2002.

[2] Joad Luiz Azevedo De Carvalho, Adson Ferreria da , "Compression of electromyographic signals using image compression techniques", international conference on technology in Biomedicine, British Columbia, Canada, pp. 333-338, 10-15 Oct. 2002.

[3] S.Michie and S.Williams, "Reducing work related psychological ill health and sickness absence: a systematic literature review" IEEE Transactions on Information Technology in Biomedical, Department of Psychology, University College London ,pp 791-795, Jan. 2003.

[4] Martin Kusserow, Oliver Amft, "Proposes a wearable talk assistant to monitor mental stress and provide relaxation feedback during public speaking", fifth international conference on image processing and its application, Wearable Computing Lab., ETH Zurich, Switzerland, Vol 14 , pp. 447 – 450 , 2003.

[5] J. Taelman, S. Vandeput , "Influence of Mental Stress on Heart Rate and Heart Rate Variability" submitted a paper to a Katholieke Universiteit Leuven, Department of Biomedical Kinesiology, Leuven, Belgium., Vol. 13, No. 2, Nov. 2003

[6] Juha Parkka, Miikka Ermes, Panu Korpipa, "Activity Classification Using Realistic Data From Wearable Sensors" IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 10, NO. 1, JANUARY 2006.

[7] E. Jovanov, A. Milenkovic, C. Otto, P. De Groen, B. Johnson, S. Warren, and G. Taibi, "A WBAN system for ambulatory monitoring of physical activity and health status: applications and challenges," Proc. 27th Ann. IEEE Conf. Eng. Med. Biol., Shanghai, China, pp. 3810– 3813, Sep. 1–4, 2005.

[8] Liao D et al, "Age, race and sex differences in autonomic cardiac function measures by spectral analysis of heart rate variability", ARIC study. Am J Cardiol 76:906-912, 2006.

[9] Alireza Seyedi, Biplab Sikdar, "Energy efficient transmission strategies for Body Sensor Networks", Transportation Research part C 6 (2006).

[10] Aleksandar Milenkovic , Chris Otto, "Wireless sensor networks for personal health monitoring: Issues and an implementation" international conference on clinical diagnosis, The University of Alabama in Huntsville, vol. 50, pp. 1019 – 1025, 2006.

[11] K. Srinivasan and P. Levis, "RSSI is under appreciated," in Proc. 3rd Workshop Embedded Netw. Sensors, Cambridge, MA, May 30–31, 2006.

[12] F. W. Beltman, M. A. Joore, D. E. M. Brunenberg, C. Dirksen, T. Thien, and P. W. de Leeuw, "Selfmeasurement of blood pressure at home reduces the need for antihypertensive drugs," Hypertension, vol. 50, pp. 1019–1025, 2007.

[13] S. N. Haynes and D. T. Yoshioka, "Clinical assessment applications of ambulatory sensors," Psychol. Assess., Department of Psychology, University of Tampere, Tampere, Finland vol. 19, no. 1, pp. 44–57, 2007

[14] M. Kumar, M. Weippert, R. Vilbrandt, S. Kreuzfeld, R. Stoll. "Fuzzy Evaluation of Heart Rate Signals for Mental Stress Assessment", IEEE Transactions on Fuzzy Systems, 2007.

[15] E. Miluzzo, X. Zheng, K. Fodor, and A. T. Campbell, "Radio characterization of 802.15.4 and its impact on the design of mobile sensor networks," in Proc. 5th Eur. Conf. Wireless Sensor Netw., Bologna, Italy, Jan. 30/Feb.1, 2008.

[16] M. U. Ahmed, S. Begum, P. Funk, N. Xiong, B. V. Schéele, "Case-based Reasoning for Diagnosis of Stress using Enhanced Cosineand Fuzzy Similarity", Transactions on Case-Based Reasoning for Multimedia Data, IBaI Publishing, 2008.

[17] E. D. Ubeyli, "Decision support systems for time-varying biomedical signals: EEG signals classification", Expert Systems with Applications, , Kahramanmaras Sutcu Imam University 2008.

[18] S. Begum, M. U. Ahmed, P. Funk, N. Xiong, B. V. Schéele, "A Case-based Decision Support System for Individual Stress Diagnosis Using Fuzzy Similarity Matching", Computational Intelligence, 2009.

[19] M. Rouhani, R. Soleymani, "Neural Networks based Diagnosis of heart arrhythmias using chaotic and nonlinear features of HRV signals", 2009 International Association of Computer Science and Information Technology - , Islamic Azad Univ., Gonabad, Iran Spring Conference, 2009.

[21] E. D. Ubeyli, "Combined neural network model employing wavelet coefficients for EEG signals classification, Digital Signal Processing, Kahramanmaras Sutcu Imam University, 2009.

[22] D. C. Hoang, Y. K. Tan, H. B. Chng, and S. K. Panda, "Thermal energy harvesting from human warmth for wireless body area network in medical healthcare system," in Proc. 8th Int. Conf. Power Electron. Drive Syst., Taipei, Taiwan, , pp. 1277–1282, Nov. 2–5, 2009.

[23] Juho Merilahti, Elina M. Mattila, "Identification of work-related stress symptoms and early intervention could reduce individual suffering and improve the working productivity and creativity." , vol. 07, no. 8, pp. 642 – 652, 2009.

[24] A. Pantelopoulos and N. G. Bourbakis, "A survey on wearable sensorbased systems for health monitoring and prognosis," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., INDONESIA. vol. 40, no. 1, pp. 1–12, Jan. 2010.

[25] M. Chen, S. Gonzalez, A. Vasilakos, H. Cao, and V. C. M. Leung, "Body area networks: A survey", ACM/Springer Mobile Networks and Applications (MONET), Western Michigan University, USA DOI: 10.1007/s11036-010-0260-8, Feb. 2011.

POWERLINE COMMUNICATION: A brief review of architecture, protocols and challenges

Swati Sharma
Lecturer, Departmentof ECE, Model Institute of Engg.&Technology, Kot Bhalwal,Jammu, India
swati.719@gmail.com

Abstract— **The power system has been used for communication purposes for many decades, although it was mainly the power utility companies that used the power lines for low bit rates for control and monitoring purposes. In the last few years, however, the deregulation of the power and telecommunication markets has spurred the idea of using and commercializing the power networks for a range of new communication applications and services. Although there are only a few practically available standards for power line communication (PLC), still these are used and implemented successfully in respective fields. The paper explains the power line communication system architecture and summarizes the protocols and applications in context of PLCC.**

Keywords—**Powerline Communication, Broadband over Powerline, PLC Standards, Challenges.**

I. INTRODUCTION

Power line Communication is the technology that enables the transmission of data over power line that carries and supplies electric power. Power lines are attractive for communications purposes because they have an omnipresence that reaches most homes and businesses, even in the most rural areas. The idea of using the power network for communication purposes is almost a century old. It is the power utility companies that developed and used this capability for many decades, transmitting data at low rates, essentially for monitoring and control purposes. With the start of the deregulation of the electricity and telecom markets in the early 90's, and motivated by the persisting monopoles of large telecom companies, the old idea of using the existing power network wiring for broadband data transmission got a new chance to be explored as a potential medium for providing telecom services to end users [6]. The power network is a large infrastructure covering most inhabited areas. The idea of powerline communication has been developed and implemented into both, narrowband and broadband systems, which are defined in terms of the operation frequency band.

- Narrow-band PLC system. They use frequencies ranging from 3-148.5 kHz in Europe, with the upper frequency extending up to 500 kHz in the United States and Japan. In Europe, this frequency range is standardized by CENELEC Standard EN 50065. It allows data transmission rates of the order of 150 kbit/s.
- Broadband PLC system. The used frequency range is 1-30MHz and a bit rate of 200Mbps; 1-15MHz for outdoor systems and 15-30MHz for indoor systems. In this frequency range, the standardization situation is still not clear and there exist no regulations.

The power network is typically divided into three sections, with different voltage levels:

- High voltage (HV): (110-380KV)
- Medium voltage (MV): (10-30KV)
- Low voltage (LV): (0.4KV)

From a communications point of view, not all parts of the network are of equal interest. The PLC system uses different parts of power network, MV and LV cablings for outdoor applications and building cabling for indoor applications [3].

II. LAYERED ARCHITECTURE

The architecture of PLC networks is comparable in many aspects to that of wired networks but also to that of Wi-Fi networks. The electrical wiring is the medium for data transmission, corresponding to the physical layer in the OSI model. Unlike other physical communication media like UTP (Ethernet cable), coaxial cable, fiber optic cable, and so forth, this role supporting data transmission is not the main function of the electrical wiring. Data transport is therefore a complementary function to delivery of electrical power (approximately 110 V/60 Hz in the United States and Japan and 220 V/50 Hz in Europe) by the wiring that powers electrical devices from the public electrical network.

The OSI (open systems interconnection) layered model provides a common base for the description of any data network. This model is composed of seven layers, each describing an independent protocol that furnishes a service to the layer above it and requests services from the layer below it.

In the context of this model, PLC networks correspond to layers 1 (physical) and 2 (data link), supplying an Ethernet connection service to the layers above.

Layer 1 (physical) is materialized by the electrical wiring that carries the PLC signal. The PLC equipment provides a terminal (typically a PC) with an Ethernet connection service corresponding to layer 2 (data link), using a MAC protocol and RJ-45 connectors.

The terminal uses PLC network services to access services in higher layers (IP, TCP, HTTP, and so forth).

A.The Physical Layer

The physical layer of PLC technologies is materialized by electrical wiring and, more generally, by electrical networks. In order to transport the PLC signal on this medium, the line frequency (for example, 220 V/50 Hz) of the electrical circuit is supplemented by a modulated signal of low amplitude around a center frequency (carrier frequency).

The physical layer therefore consists of this low amplitude modulated signal, transported on electrical wiring at a frequency determined by the PLC technology employed and the applicable regulations.

B. MAC layer

A MAC protocol specifies a resource sharing strategy applied to a multiple access scheme. It accommodates multiple users in the sharing of the network transmission capacity.

The various algorithms used are generally based on either a carrier sense technique or token passing mechanism. However, when those algorithms are not transferable to PLC network, the more attention should be paid to design of MAC protocol for the PLC technology because of the following reasons. "Carrier Sensing problem": on the power line there is insufficient communications reliability to distinguish between noise and signal. This makes carrier sense difficult.

"Hidden-node problem": Since the power line characteristics can be remarkably different for each node, there is a high probability that a node will not necessarily listen to all the transmissions on the power line. In carrier sense, a node may thus incorrectly sense that the channel is quiet and start transmitting in the middle of another transmission. The above two problems make carrier sense multiple access with collision detection (CSMA/CD) hard to be implemented for the power line environment ,so some other methods need to be implemented[4].

III. TYPES OF PLC SYSTEMS

Depending on the frequency band, the systems over powerlines can be narrowband or broadband. Narrowband systems use a frequency band limited to a fraction of a megahertz and their potential as interferers is low. Although the operating frequencies employed by these systems are specific to each country, they are in general in the CENELEC bands, ranging from 3-148.5 kHz in Europe and going up to 500 kHz in United States and Japan. Some narrowband systems use proprietary, patented modulation schemes but many of them use plain frequency shift keying (FSK). Broadband systems use the frequency band of 1-30MHz, dividing the 1 to 30 MHz used frequency band into outdoor, for which the lower half of frequency band is used and indoor, for which the rest of the spectrum is used. Broadband systems are still not standardised, although the HomePlug vendor alliance has advanced its own specifications mostly in the United States.

The frequency range between 1 and 30 MHz appears to be of practical interest for broadband services using the low voltage (LV) distribution network as a transmission channel. The frequencies below 1 MHz cannot be used due to a high noise level produced by electrical appliances and frequencies above 30 MHz have relatively low capacity for data transmission over long distances due to high channel attenuation. For home networking, which involves small distances, however, higher frequencies e.g. up to 60MHz or even 80 MHz may be considered [1][5].

The medium voltage (MV) network is considered as a medium for the backbone connection in a PLC system. Indeed, the backbone connection is of vital importance for the competitiveness of data-over-powerline systems since most medium voltage to low voltage (MV/LV) substations are not connected to existing telecom infrastructure [5][6].

Numerous studies and field tests have shown that the channel capacity of typical distribution networks on the medium voltage and low voltage levels allows data rates up to several hundred Mbits/s for an operation frequency bandwidth of about 20 MHz.

Although the PLC technology is promising and is potentially easy to install and inexpensive, it is still not fully developed and standardized.

Due to asymmetries in the power network and the equipment connected to it, PLC signals are not confined to the network's guiding conductors but some part is also radiated. Depending on the level of the radiated PLC signals, these may interfere with the reception of radio signals or other services in nearby receivers, so tradeoff solutions will have to be worked out for frequency allocation and the determination of level limits [4][6].The mentioned lack of standardization in PLC concerns both the technology specification itself and the electromagnetic compatibility (EMC) aspects.

IV. STANDARDS USED FOR POWERLINE COMMUNICATION

Various protocols have been developed for use for communication on power line. They differ in their modulation techniques, channel access echanisms and the frequency bands they use. A brief overview of the most popular protocols is presented here.

A. X10 home automation

X10 technology is a communications standard for sending control signals to home automation devices via the power line (120 V or 220 V, at 50 Hz or 60 Hz).

It uses a form of amplitude shift keying (ASK) technique for transmission of information. Although it was originally unidirectional (from controller to controlled modules) some bi-directional products have also been implemented. X-10 controllers send their signals over the power line to simple receivers that are used mainly to control lighting and other appliances. A 120 KHz amplitude modulated carrier, 0.5-watt signal, is superimposed into AC power line at zero crossings to minimize the noise interference. Information is coded by way of bursts of high frequency signals. The standard includes addressing mechanisms to individually identify appliances. The presence of a 120Khz signal burst at zero crossing indicates the transmission of a binary '1', whilst the absence of the 120Khz signal indicates a binary '0'. In order to control specific devices,modules are assigned an address, which consists of a house and unit code. A typical X-10 transmission would include a start code, house address, device address, and then function code (such as ON, OFF, etc…). [2][4].

B. Intellon CEBus

The CEBus (Consumer Electronics Bus) standard is an open standard that provides separate physical layer specifications for communication on power lines and other media.

Data packets are transmitted by the transceiver at about 10 kbps, using spread spectrum technology. The CEBus protocol uses a peer-to-peer communications model so that a network node can access the media only when the network is available. It uses a Carrier Sense Multiple Access/Collision Detection and Resolution (CSMA/CDCR) protocol to avoid data collisions. CEBUS is a commercially owned protocol, and thus attracts registration fees.

C. Echelon LONWorks

Echelon, like Intellon, provides a peer-to-peer communication protocol, via Carrier Sense Multiple Access (CSMA) techniques . The protocol provides a set of communication services that allow the applications in a device to send and receive messages over the network without the need to know the topology of the network or the names, addresses or functions of other devices. All communications consists exchange of one or more packets between devices. Each packet is a variable number of bytes in length and contains a compact representation of the data required for each of the 7 layers of the OSI model. The addressing algorithm of this protocol defines how packets are routed from a source device to one or more destination devices. Packets can be addressed to a single device, to any group of devices or to all devices. LONWorks address types include physical, device, group, and broadcast addresses. Every packet transmitted over the network contains the device address of the transmitting device (the source address) and the address of the receiving device (destination address) that can either be a physical, a device, a group or a broadcast address. It is possible for two or more independent LONWorks systems to coexist on the same physical channel, as long as each system has a unique domain ID. Devices in each system respond only to those packets corresponding to their domain ID and do not know about packets addressed with other domain IDs. Each domain in a system using this protocol can have up to 32,835 devices. There can be up to 256 groups in a domain and each group can have any number of devices assigned to it, except that end to end acknowledgement is required. There can be up to 255 subnets in a domain and each subnet may have up to 127 devices.Echelon offers a 10 Kbps power line chip based on spread spectrum technology. The LONWorks has just been passed ANSI/EIA standard proecess and now can be known as ANSI/EIA 701.9-A-1999[7].

D. HomePlug 1.0 standard

HomePlug Power line alliance has worked towards a common standard in the United States for high-speed home power line network. It is a non-profit corporation formed to provide a forum for the creation of an open standard and specification for home powerline networking products and services and to accelerate the demand for products based on these standards worldwide through the sponsorship of market and user education programs. In July 2001, the HomePlug Specification 1.0 was ready. Now, it is endorsed by about 30 vendor companies. While HomePlug 1.0 was designed mainly to distributed broadband internet access in the home, the objective of HomePlug AV standard is to distribute Audio/Video content within the house, as well as data.

E. IEC 61334

Power line communication systems for MV and LV, are denoted as distribution line communication (DLC) systems by the International Electrotechnical Commission (IEC). DLC communication systems are being standardized by IEC, Technical Committee No 57 (Power System Control And Associated Communications), Working Group 9 (Distribution Automation Using Distribution-Line-Carrier Systems). IEC 61334 describes the structure of distribution systems for both medium and low-voltage levels and presents the architecture for a distribution automation system using distribution line carrier systems and use frequencies below 150 kHz.

WG 9 has developed IEC 61334-4 series, a distribution Power Line Carrier (PLC) communication standard called Distribution Line Message Specification (DLMS). DLMS provides two-way communications and can be used on medium and low voltage networks. While DLMS can be used to access a wide variety of devices (switches, meters, lighting control, and other load controlling devices). Currently, it is primarily used for retrieving metering information using the IEC 62056 metering standard [5].

V. Power line carrier challenges

A. Noise

The major sources of noise on power line are from electrical appliances, which utilize the 50 Hz electric supplies and generate noise components, which extend well into the high frequency spectrum. Apart from these induced radio frequency signals from broadcast, commercial, military, citizen band and amateur stations severely impair certain frequency bands on power line. The primary sources of noise in residential environments are universal motors, light dimmers and televisions. This noise can be classified as:

50 Hz periodic noise

Noise synchronous to the sinusoidal power line carrier can be found on the line. The sources of this noise tend to be silicon-controlled rectifiers (SCRs) that switch when the power crosses a certain value, placing a voltage spike on the line. This category of noise has line spectra at multiples of 50 Hz.

Single-event impulse noise

This category includes spikes placed on the line by single events, such as a lightning strike or a light switch turn on or off. Capacitor banks switched in and out create impulse noise.

Periodic impulsive noise

The most common impulse noise sources are triaccontrolled light dimmers. These devices introduce noise as they connect the lamp to the AC line part way through each AC cycle. These impulses occur at twice the AC line frequency as this process is repeated every ½ AC cycle.

Continuous Impulsive noise

This kind of noise is produced by a variety of series wound AC motors. This type of motor is found in devices such as found in vacuum cleaners, drillers,electric shavers and many common kitchen appliances. Commutator arcing from these motors produces impulses at repetition rates in the several kilohertz range. Continuous impulsive noise is the most severe of all the noise sources.

Non-synchronous periodic noise

This type of noise has line spectra uncorrelated with 50 Hz sinusoidal carriers. Television sets generate noise synchronous to their 15734 Hz horizontal scanning frequency. Multiples of this frequency must be avoided when designing a communications transceiver. It was found that noise levels in a closed residential environment fluctuate greatly as measured from different locations in the building.

Noise levels tend to decrease in power level as the frequency increases; in other words, spectrum density of power line noise tends to concentrate at lower frequencies. This implies that a communications carrier frequency would compete with less noise if its frequency were higher.

Background Noise

This is what every subscriber sees as already present on the line, and not caused by subscriber’s appliances. Typically, this originates from the Distribution Transformer, public lighting systems etc.

B. Attenuation

Attenuation is the loss of signal strength as the signal travels over distance. For a transmission line the input impedance depends on the type of line, its length and the termination at the far end. The characteristic impedance of a transmission line (Zo) is the impedance measured at the input of this line when its length is infinite.

Power line networks are usually made of a variety of conductor types and cross sections joined almost at random. Therefore a wide variety of characteristic impedances are encountered in the network. Unfortunately, a uniform distributed line is not a suitable model for PLC communications, since the power line has a number of loads (appliances) of differing impedances connected to itfor variable amounts of time. Channel impedance is a strongly fluctuating variable that is difficult to predict. The overall impedance of the low voltage network results from a parallel connection of all the network’s loads, so the small impedances will play a dominant role in determining overall impedance.

Overall network impedances are not easy to predict either. The most typical coaxial cable impedances used are 50 and 75-ohm coaxial cables. A twisted pair of guage-22wire with reasonable insulation on the wires measures at about 120 ohms. Clearly, channel impedance is low. This presents significant challenges when designing a coupling network for PLC communications. Maximum power transfer theory states that the transmitter and channel impedance must be matched for maximum power transfer. With strongly varying channel impedance, this is tough. We need to design the transmitter and receiver with sufficiently low output/input impedance (respectively) to approximately match channel impedance in the majority of expected situations.

VI. ECONOMIC CHALLENGES

BPL investors face economic challenges for full-scale deployment. It is challenging to assess if BPL can bring good returns on investments. This depends on adequate market penetration. This in turn depends on the economic success of BPL vendors [3]. It is challenging to judge and to bring down the significant acquisition costs (SAC) involved in setting up the BPL in customer premises. SAC includes marketing, sales, installation and authorization. Investors face special economic challenges for rural deployments. More repeaters are needed to improve the signal strength for rural areas. For example, one repeater is needed for every 1000 ft to 2500 ft, which means on the average 30 to 100 repeaters are used for every 20 miles [1].

Reliability of network is also a concern since failure of any one of the repeaters in network could result in interruption of service for downstream subscribers [10]. Field trials carried out up to date have been for distances of less than a mile and therefore, there are no experimental results available for networks that incorporate 30-100 repeaters in series. Sparsely populated rural areas results in assignment of one distribution transformer to possibly one subscriber leading to high investment costs on transformer bypass equipment, couplers [1].

VII. CONCLUSION

Broadband over power lines continues to have technical challenges including interference, signal attenuation, noise, lack of standards, and threat to data security for full-scale deployment and some economical challenges too. Rural BPL deployment continues to pose economic challenges. Although the available measures alleviate the challenges, they do not completely eliminate them. The aforementioned technical and economic areas need to be addressed before full-scale deployment of BPL is possible.

In conclusion, PLC could provide a bi-directional, broadband communications platform capable of delivering real-time data, in order to be applied in a wide range of applications for the utility. With the development of high speed PLC network, the power utilities can benefit from PLC technology as it could increase its operating efficiency and infrastructure usage and incorporate into its portfolio value added services in different areas.

Nevertheless, further research and development would be required to ensure the QoS of PLC Communication architectures and to confirm PLC technology as the most adequate one for the targeted applications.

REFERENCES

[1] Gilbert Held,"Understanding Broadband over Power Line",ISBN 0-8493-9846-0 .

[2]http://www.nationwidesatellite.com/HughesNet/service_plans/HughesNet_plans.asp ~ Online

[3]M.Zimmermann,K.Dostert,"AnAnalysisOfTheBroadband Noise Scenario in Power line Networks", ISPLC 2000 [4] OPERA project website, ttp://www.ist-opera.org/

[4]T.Masamitsu,"Technical Trends in High-Speed Power Line Communication", IEEE Trans on Commu, Vol E88-B, No 8,Aug 2005.(online)

[5]K.Dostert,Powerlinecommunications, Prentice Hall, 2001

[6]H. Hrasnica, A. Haidine, R. Lehnert, Broadband Powerline Communications, Chichester ; Hoboken, N.J. : Wiley, 2004.

[7]http://www.echelon.com/support/documentation/manuals/general/078-0183-01A.pdf

QOS PROVISIONING AND TWOFOLD QOS SOLUTIONS IN MANETs

Tanu Preet Singh[1], Vikrant Das[2], and Srishti Maheshwari[3]
[1,2,3] Department of Computer Science & Engineering, ACET, Amritsar

***Abstract* — An ad hoc network is a collection of mobile hosts forming a temporary network on the move, without using any fixed infrastructure. It is characterized by the uniqueness, self-deployment and self-organization. Hence it is classified as a dynamic network. Characteristics of adhoc networks such as lack of central coordination, mobility of hosts, dynamically varying network topology, and limited availability of resources make QoS provisioning very challenging in such networks. In this paper we have provided a 3-tier classification of the existing QoS solutions. QoS parameters that decide the throughput of the network are bandwidth, end-to-end delivery, jitter (delay variance) and energy of the packets.**

Index Terms – MANETs, Adhoc Networks, Quality of Service, QoS provisioning, QoS Parameters

I. INTRODUCTION

Quality of service (QoS) is the performance level of a service offered by the network to the user. The goal of QoS is to achieve a more deterministic network behavior, so that information carried by the network can be better delivered and network resources can be better utilized. A network or a service provider can offer different kinds of services to the users. Here, a service can be characterized by a set of measurable pre-specified service requirements such as bandwidth, delay, delay variance (jitter), and packet loss rate. After accepting a service request from the user, the network has to ensure that service requirements of the uses flow are met, as per the agreement, throughout the duration of the flow (a packet stream from the source to the destination). In other words, the network has to provide a set of service guarantees while transporting a flow.

Quality of service (QoS) provisioning is becoming a critical issue in designing wireless ad hoc networks due to the necessity of providing multimedia applications in such networks. These applications are typically delay-sensitive and have high bandwidth requirements. Providing QoS in wireless ad hoc networks is challenging because the wireless channel is shared among adjacent hosts, and network topology can change as hosts move. Decoupling routing from QoS provisioning can result in the selection of inefficient routes and thus reduce the likelihood of meeting the QoS requirement(s) of arriving communication requests.

The goals of QoS routing are twofold:

- Selecting paths that can satisfy given QoS requirements of arriving communication requests,
- Achieving global efficiency in resource utilization.

Typical routing metrics for providing QoS include

- Delay
- Jitter (delay variance),
- Bandwidth allocation
- Packet loss rate.

QoS routing plays an important role in providing QoS aware services in wireless ad hoc networks. Here, we give some key design considerations in QoS routing.

- First, a designed routing protocol should scale well with respect to the network size, and with respect to computation, communication and storage overhead.
- Second, it is difficult to guarantee an initial QoS contract with a session that has specific QoS requirement(s), due to network dynamics caused by node mobility. There may exist transient time when the required QoS is not guaranteed due to path break or network partition.

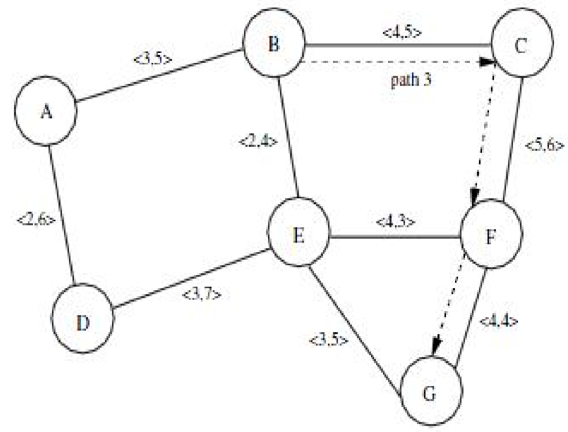


Figure 1: An example of QoS routing in mobile adhoc networks

- Third, a designed protocol should be able to effectively absorb routing information inaccuracy.
- Fourth, it is necessary to prevent QoS traffic from starving best effort traffic in networks wherein the two traffic types coexist [5].

II. QoS (Quality of Service) Parameters

As different applications have different requirements, the services required by them and the associated QoS parameters differ from application to application.

For example, in case of multimedia applications, bandwidth, delay jitter, and delay are the key QoS parameters, whereas military applications have stringent security requirements.For applications such as emergency search and rescue operations, availability of network is the key QoS parameter. For applications such as group communication in a conference hall require that the transmissions among nodes consume as minimum energy as possible. Hence battery life is the key QoS parameter here.

Unlike traditional wired networks, where the QoS parameters are mainly characterized by the requirements of multimedia traffic, in MANETs the QoS requirements are more influenced by the resource constraints of the nodes. Some of the resource constraints are battery charge, processing power, and buffer space.

III. QoS ISSUES IN ADHOC NETWORKS

Providing QoS support in mobile adhoc networks is an active research area. MANETs have certain unique characteristics that pose several difficulties in provisioning QoS.

Characteristics of MANETs that affect the QoS provisioning are listed below.

- Dynamically varying network topology.
- Imprecise state information.
- Lack of central coordination.
- Error-prone shared radio channel
- Hidden terminal problem
- Limited resource availability
- Insecure medium.

3.1 Dynamically varying network topology:

The nodes in the ad hoc networks are dynamic in nature and the network topology changes periodically. Due to this the QoS session that are set up may suffer due to frequent path breaks which intron requires rigorous sessions of path re-establishment.

This delay which is experienced in constant re-establishment of session leads to an increased overhead and will cause some packets of that session to be dropped and miss their targets which is not at all acceptable for the applications that have stringent QoS requirements.

3.2 Imprecise state information:

In most cases, the nodes in an ad hoc wireless network maintain both the link-specific state information and flow-specific state information. The link-specific state information includes bandwidth, delay, delay jitter, loss rate, error rate, stability, cost, and distance values for each link.

The flow specific information includes session ID, source address, destination address, and QoS requirements of the flow (such as maximum bandwidth requirement, minimum bandwidth requirement, maximum delay, and maximum delay jitter).

The state information is inherently imprecise due to dynamic changes in network topology and channel characteristics. Hence routing decisions may not be accurate, resulting in some of the real-time packets missing their deadlines.

3.3 Lack of central coordination:

Unlike wireless LANs and cellular networks, ad hoc networks do not have central controllers to coordinate the activity of nodes. This further complicates QoS provisioning in wireless adhoc networks.

3.4 Error prone shared radio channel:

The radio channel is a broadcast medium by nature. During propagation through the wireless medium the radio waves suffer from several impairments such as attenuation, multi-path propagation, and interference (from other wireless devices operating in the vicinity).

3.5 Hidden terminal problem:

The hidden terminal problem is inherent in adhoc networks. This problem occurs when packets originating from two or more sender nodes, which are not within the direct transmission range of each other, collide at a common receiver node. It necessitates retransmissionof packets, which may not be acceptable for flows that have stringent QoS requirements

3.6 Limited resource availability:

Resources such as bandwidth, battery life, storage space, and processing capability are limited in ADHOC NETWORKSs. Out of these, bandwidth and battery life are very critical resources, the availability of which significantly affects the performance of the QoS provisioning mechanism. Hence efficient resource management mechanisms are required for optimal utilization of these scarce resources.

3.7 Insecure medium:

Due to the broadcast nature of the wireless medium, communication through a wireless channel is highly insecure. Hence security is an important issue in ADHOC NETWORKSs, especially for military and tactical applications. Adhoc Networks are susceptible to attacks such as eavesdropping, spoofing, denial of service, message distortion, and impersonation. Without sophisticated security mechanisms, it is very difficult to provide secure communication guarantees [8].

IV. DESIGN CHOICES FOR QoS SUPPORT

Some of the design choices for providing QoS support are described below:

4.1 Hard state v/s soft state resource reservation:

QoS resource reservation is one of the very important components of any QoS framework(a QoS framework is a complete system that provides required/promised services to each user or application) [1]. It is responsible for reserving resources at all intermediate nodes along the path from the source to the destination as requested by the QoS session.

QoS resource reservation mechanisms can be broadly classified into two categories, *hard state and soft state reservation mechanisms*.

In hard state resource reservation schemes, resources are reserved at all intermediate nodes along the path from the source to the destination throughout the duration of the QoS session. If such a path is broken due to network dynamics, these reserved resources have to be explicitly released by a de-allocation mechanism. Such a mechanism not only introduces additional control overhead, but may also fail to release resources completely in case a node previously belonging to the session becomes unreachable.

Due to these problems soft state resource reservation mechanisms, which maintain reservations only for small time intervals, are used. These reservations get refreshed if packets belonging to the same flow are received before the timeout period. The soft state reservation timeout period can be equal to packet inter-arrival time or a multiple of the packet inter-arrival time. If no data packets are received for the specified time interval, the resources are de-allocated in a decentralized manner without incurring any additional control overhead.

Thus no explicit tear down is required for a flow. The hard state schemes reserve resources explicitly and hence at high network loads, the call-blocking ratio will be high, where as soft state schemes provide high call acceptance at a gracefully degraded fashion.

4.2 Stateful v/s stateless approach:

In the stateful approach, each node maintains either global state information or only local state information, while in the case of stateless approach no such information is maintained at the nodes.State information includes both the topology information and the flow-specific information [4].

If global state information is available, the source node can use a centralized routing algorithm to route packets to the destination. The performance of the routing protocol depends on the accuracy of the global state information maintained at the nodes. Significant control overhead is incurred in gathering and maintaining global state information.

On the other hand, if mobile nodes maintain only local state information (which is more accurate), distributed routing algorithms can be used. extremely difficult.

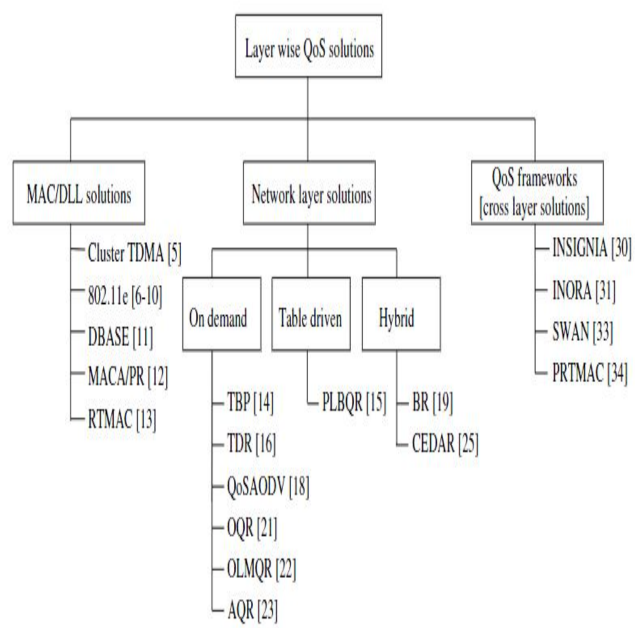


Figure 2 Classification of QoS approaches

Even though control overhead incurred in maintaining local state information is low, care must be taken to obtain loop-free routes. In the case of stateless approach, neither flow-specific nor link specific state information is maintained at the nodes. Though the stateless approach solvesthe scalability problem permanently and reduces the burden (storage and computation) on nodes, providing QoS guarantees becomes

4.2 Hard QoS v/s soft QoS approach:

The QoS provisioning approaches can be broadly classified into two categories, hard QoS and soft QoS approaches. If QoS requirements of a connection are guaranteed to be met for the whole duration of the session,the QoS approach is termed as hard QoS approach.

If the QoS requirements are not guaranteed for the entire session, the QoS approach is termed as soft QoS approach [3].Keeping network dynamics of ADHOC NETWORKSs in mind, it is very difficult to provide hard QoS guarantees to user applications. Thus, QoS guarantees can only be given within certain statistical bounds. Almost all QoS approaches available in the literature provide only soft QoS guarantees.

V. QoS SOLUTIONS

The QoS solutions can be classified in two ways. One classification is based on the QoS approach employed, while the other one classifies QoS solutions based on the layer at which they operate in the network protocol stack.

5.1 Classification on the basis of QoS Approaches

The QoS approaches can be classified based on the interaction between the routing protocol and the QoS provisioning mechanism, based on the interaction between the network and the MAC layers, or based on the routing information update mechanism. Based on the interaction between the routing protocol and the QoS provisioning mechanism, QoS approaches can be classified into two categories,coupled and decoupled QoS approaches. In the case of the coupled QoS approach, the routing protocol and the QoS provisioning mechanism closely interact with each other for delivering QoS guarantees.

If the routing protocol changes, it may fail to ensure QoS guarantees. But in the case of decoupled approach, the QoS provisioning mechanism does not depend on any specific routing protocol to ensure QoS guarantees.

Similarly, based on the interaction between the routing protocol and the MAC protocol, QoS approaches can be classified into two categories, independent and dependent QoS approaches. In the independent QoS approach, the network layer is not dependent on the MAC layer for QoS provisioning [4].The dependent QoS approach requires the MAC layer to assist the routing protocol for QoS provisioning.

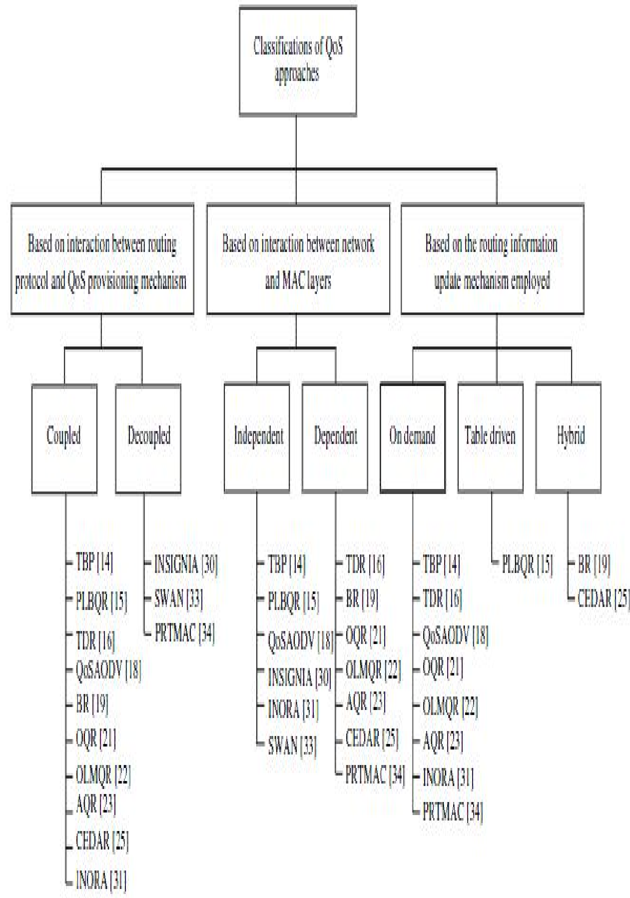


Figure 3 Layer Wise QoS Solutions

Finally, based on the routing information update mechanism employed, QoS approaches can be classified into three categories viz., table-driven, on-demand, and hybrid QoS approaches. In the table-driven approach, each node in the network maintains a routing table which aids in forwarding packets. In the on-demand approach, no such tables are maintained at the nodes, and hence the source node has to discover the route on the fly. The hybrid approach incorporates features of both the table-driven and the on-demand approaches.

5.2 Layer-wise classification of existing QoS Solutions

The existing QoS solution s can also be classified based on which layer in the network protocol stack they operate in. Fig. 3 gives a layer-wise classification of QoS solutions. The figure also shows some of the cross-layer QoS solutions proposed for MANETs.

5.3 MAC layer solutions

The MAC protocol determines which node should transmit next on the broadcast channel when several nodes are competing for transmission on that channel. Some of the MAC protocols that provide QoS support for applications in ADHOC NETWORKSs are described below.

5.6 Network layer solutions

The bandwidth reservation and real-time traffic support capability of MAC protocols can ensure reservation at the link level only, hence the network layer support for ensuring end-to-end resource negotiation, reservation, and reconfiguration is very essential. This section describes the existing network layer solutions that support QoS provisioning.

5.7 QoS frameworks:

A framework for QoS is a complete system that attempts to provide required/promised services to each user or application. All components within this system cooperate together in providing the required services. The key component of any QoS framework is the QoS model which defines the way user requirements are met. The key design issue here is whether to serve users on a per session basis or on a per class basis. Each class represents an aggregation of users based on certain criteria.

The other key components of the framework are, QoS routing which is used to find all or some of the feasible paths in the network that can satisfy user requirements, QoS signaling for resource reservation, QoS medium access control, call admission control, and packet scheduling schemes.

The QoS modules should react promptly to changes in the network state (topology changes) and flow state (change in the end-to-end view of the service delivered).

VI. CONCLUSION

In this paper we have discussed about quality of service of an adhoc network and various parameters related to it.

Initially the issues and challenges involved in the QoS provisioning in adhoc networks were identified. A detailed description of each approach was discussed. The QoS approaches are classified according to the criteria which includes interaction betweenrouting protocol and resource reservationsignaling, interaction between network and MAClayer, and routing information update mechanism. Layer wise solutions were discussed briefly.

REFERENCES

[1] Vishal Garg, Chandan Kapoor, “A Survey of QoS parameters through reactive routing in MANETs”,International Journal of Computational Engineering & Management, Vol. 13, July 2011.

[2] Dr. Shuchita Upadhayaya and Charu Gandhi, Quality Of Service Routing In Mobile Ad Hoc Networks Using Location And Energy Parameters, International Journal of Wireless & Mobile Networks (IJWMN), Vol 1, No 2, November 2009.

[3] M.Uma, Dr.G.Padmavathi,“A Comparative Study And Performance Evaluation Of Reactive Quality Of Service Routing Protocols In Mobile Adhoc Networks”, Journal of Theoretical and Applied Information Technology

[4] Dr.R.Asokan, “A Review of Quality of Service (QoS) Routing Protocols for Mobile Ad hoc Networks”, IEEE , 2010.

[5] Hanzo (II.) and R. Tafazolli, “A Survey of QoS Routing Solutions for Mobile Ad hoc Networks”, 2010

[6] *Baoxian Zhang and Hussein T. Mouftah,”* QoS Routing for Wireless Ad Hoc Networks: Problems, Algorithms, and Protocols*” IEEE, 2005*

[7] Dmitri D. Perkins and Herman D. Hughes, “A survey on quality-of-service support for mobile ad hoc Networks”

[8] VinodKone and Sukumar Nandi, “QoS Constrained Adaptive Routing Protocol For Mobile Adhoc Networks”, 2009.

Secure Steganographic Image using Digital Water Marking

Vishal Gupta[1],Sukhmeet Singh[2]
[1]M.Tech(ECE), Panipat Institute of engineering technology, Panipat
[2]ECE DEPTT, Technical Education Research Integrated Institute, Kurukshetra, Haryana
Email:v.vishu22@gmail.com, sukhmeetreck@gmail.com

Abstract- The illegal use image is increasing day by day. In order to make our image secure steganography is used. Watermarking is one of the most recent proposed systems to observe the authentication of licensed user over e-commerce applications and finds its uses in illegal applications like copying the multimedia data e.g. video, audio, image. The watermark indicates that data is containing copyright or not. To propose a measure against the illegal use of the images different available watermarking standards are studied. If any effort is done to copy or download the image in any unauthentic way i.e. without availability of any license or the Private Key issued by the owner the designed software damages the content of that image file so that the image looses its commercial value.

Keywords – Watermarking, Steganography, LSB, MSB

I. INTRODUCTION

Digital watermarking can be a form of steganography, in which data is hidden in the message without the end user's knowledge. Image watermarking is a technique which allows an individual to add hidden copyright notices or other verification messages to digital audio, video, or image signals and documents. A watermark can be classified into two sub-types: visible and invisible. Visible watermarks change the signal altogether such that the watermarked signal is totally different from the actual signal, e.g., adding an image as a watermark to another image. Invisible watermarks do not change the signal to a perceptually great extent, i.e., there are only minor variations in the output signal. An example of an invisible watermark is when some bits are added to an image modifying only its least significant bits. Invisible watermarks that are unknown to the end user are steganographic. Another application is to protect digital media by fingerprinting each copy with the purchaser's information. If the purchaser makes illegitimate copies, these will contain his name. Fingerprints are an extension to watermarking principle and can be both visible and invisible. [5]

The rest of the paper is organized as follows. Proposed embedding and extraction algorithms are explained in section II. Experimental results are presented in section III. Concluding remarks are given in section IV.

II. PROPOSED ALGORITHM

Let us take the existing work for encrypting images. In Photo shop we can easily encrypt the image and save for security purposes. If we have a situation that we have to encrypt our image at remote web sever then what we have to do we have an alternative we do this by encrypting in Photoshop or other related software the image is encrypted. Problem arises, when we need original and encrypted image at web server. When both the images are available at the web server the encrypted image unwontedly occupy the web space which is costly that is more is the space more is the cost and also we have to consume extra time to encrypt the image and upload into web server which is tedious and time wastage. To over come this problem we design watermark engine at remote web server which is automatically encrypts the images without creating the copy of the original image thus totally work has been performed automatically at remote web server.[2]

Watermark stored in a data file may refer to a method for ensuring data integrity which combines aspects of data hashing and digital watermarking. Data hashing and digital watermarking are useful for tamper detection, but at the same time these techniques have associated disadvantages. For example, a typical data hash will process an input file to produce an alphanumeric string unique to the data file. If one or more bit changes occur within this original file, thereby resulting in a modified data file, the same hash process on the modified file will produce a completely different alphanumeric. In this manner, if a trusted source calculates the hash of the original data file, subscribers can verify the integrity of the data. The subscriber simply compares a hash of the received data file with the known hash from the trusted source. If the hash results are the same, they can assign an appropriate degree of confidence to the integrity of the received data. On the other hand, if the hash results are different, they can conclude that the received data file was altered. A disadvantage of this hash process is that no indications exist as to the extent or location of changes within the received data file. It's an all or nothing process; either the entire received data file is trustworthy or none of it is. Watermarking is distinctly different from data hashing, with associated advantages and disadvantages. Digital watermarking is the process of altering the original data file, allowing for the subsequent recovery of embedded auxiliary data referred to as a watermark. A subscriber, with knowledge of the watermark and how it is recovered, can determine, to some extent, whether or not significant changes occurred within the data file.

Depending on the specific method used, recovery of the embedded auxiliary data can be robust to post-processing (e.g. lossy compression). For example, if the data file to be retrieved is an image, the provider can embed a watermark for protection purposes. In this case, the process allows tolerance to some change, while maintaining an association with the original image file. Researchers have also developed techniques that embed components of the image within the image. This can help identify portions of the image that may contain unauthorized changes and even help in recovering some of the lost data. A disadvantage of digital watermarking is that a subscriber cannot significantly alter some files without sacrificing the quality or utility of the data. This can be true of various files including image data, audio data, and computer code. [1]

A. *STEGANOGRAPHY –*

Steganography comes from a Greek word that means covered writing (*stego* = covered + *graphy* = writing). Examples can be thought as messages exchanged between drug dealers via emails in encrypted forms, or messages exchanged by spies in covert communication. Steganography hides the fact that the communication ever occurred as shown in Figure 1. [6]

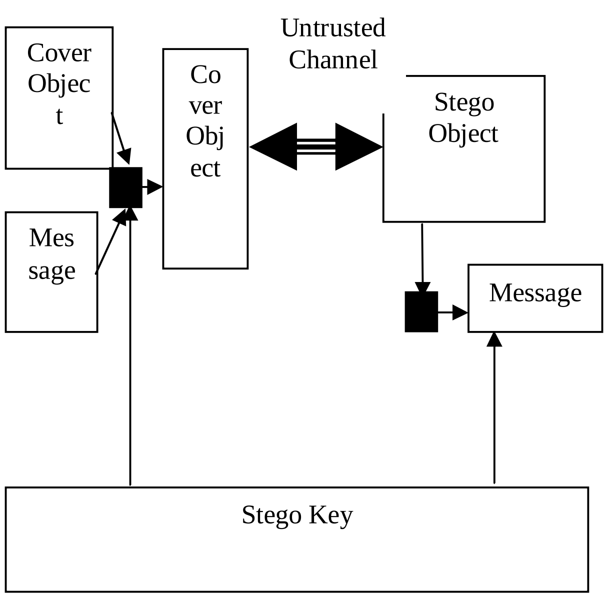


Figure 1: Use of Steganography for protection of text messages

B WATERMARKING –

Watermarking, as opposed to steganography, has an additional requirement of robustness against possible attacks. An ideal steganographic system would embed a large amount of information perfectly securely, with no visible degradation to the cover object. An ideal watermarking system, however, would embed an amount of information that could not be removed or altered without making the cover object entirely unusable.

As a side effect of these different requirements, a watermarking system will often trade Capacity and perhaps even some security for additional robustness. The working principle of the watermarking techniques is similar to the steganography methods. A watermarking system is made up of a watermark embedding system and a watermark recovery system. The system also has a *key* which could be either a public or a secret key. The *key* is used to enforce security, which is prevention of unauthorized parties from manipulating or recovering the watermark. The embedding and recovery processes of watermarking are shown in Figure 2 and Figure 3

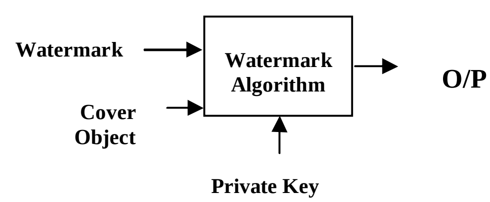


Figure 2: General Watermarking Block Diagram

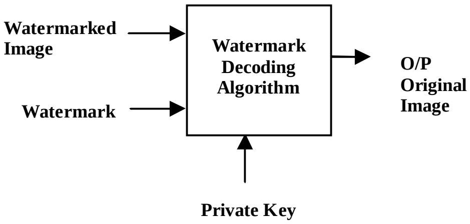


Figure 3: General Watermarking Decoding to recover Original Image

The efficiency of a digital watermarking process is evaluated according to the properties of perceptual transparency, robustness, computational cost, bit rate of data embedding process, false positive rate, recovery of data with or without access to the original signal, the speed of embedding and retrieval process, the ability of the embedding and retrieval module to integrate into standard encoding and decoding process etc.

C CONTRAST MASKING

Masking refers to the effect of one stimulus on the delectability of another stimulus. For instance, in the audio case, a strong noise can hide a weaker signal such as the talking between two people. In the image case, masking refers to a decrease in the visibility of an image component because of the presence of another. The experiments about the contrast masking are conducted in. The subjects are asked to discriminate the superposition of a+b of two sinusoidal grating from b presented alone. Grating b is called the masker and grating a is called the signal. Its contrast is varied to find the threshold of visibility. The configuration for the experiments is illustrated in Figure 4 & Figure 5. In figure 5 the difference become very clear with increase in signal amplitude. [3]

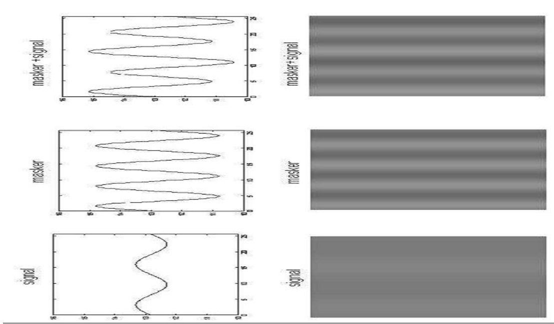


Figure 4: Effect of masker on signal

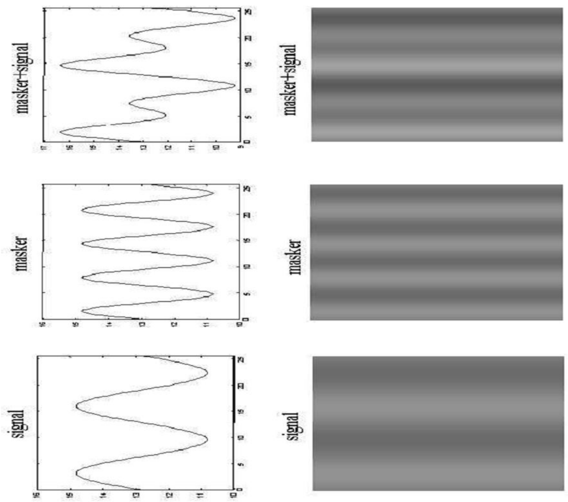


Figure 5: Effect on Contrast with variation of masker

D VISIBILITY REDUCTION TECHNIQUE –

Visibility of the watermark on the picture can be reduced by using no. of techniques. The simplest technique used for hidden watermarking is to hide the message bits in the Least Significant Bits (*LSB*) of the cover object. The advantage with this method is that even if a part of the stego image is cropped the receiver can still get the required message, as the message is embedded a number of times. The message for this case is considered to be very small as compared to the cover object.

For example, for an 8 bit file, each pixel is represented by 8 bits: 10001100.The most significant bits (*MSB*) are to the left and the least significant bits (*LSB*) are to the right. If you change the *MSB* it will have a big impact on the color, however, if you change the *LSB*, it will have minimal effect.

Now to take this method a step further, if we change only 1 or 2 least significant bits in the image, it will have a minimal effect because the human eye can only detect around 6 it’s of color. In other words, the human eye could not tell the difference of the last 2 bits being changed. For example, if we take 100011**00** and change it to 100011**11** **f**or 100011**10**, it will all seem like the same color to the human eye. So we would only embed data in those bits. An example of this is:

If the message converted to binary is 1101 0010, the first 8 pixels will be modified as follows:

- 1100 0101 becomes 1100 01**11**
- 1111 0010 becomes 1111 00**01**

- 1010 1111 becomes 1010 11**00**
- 0010 0010 becomes 0010 00**10**

III. Experiment and Result

PROGRAM FOR LIGHT COLOUR WATER MARK

```
float[][] colorMatrixElements = {
new float[] {1.0f,  0.0f,  0.0f,  0.0f, 0.0f},
new float[] {0.0f,  1.0f,  0.0f,  0.0f, 0.0f},
new float[] {0.0f,  0.0f,  1.0f,  0.0f, 0.0f},
new float[] {0.0f,  0.0f,  0.0f,  0.1f, 0.0f},
new float[] {0.0f,  0.0f,  0.0f,  0.0f, 1.0f}};
```

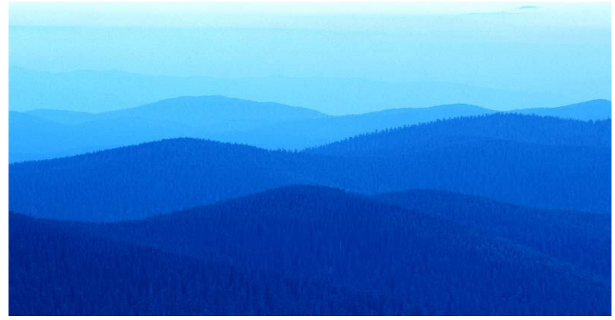
PROGRAM FOR MEDIUM COLOUR WATER MARK

```
float[][] colorMatrixElements = {
new float[] {1.0f,  0.0f,  0.0f,  0.0f, 0.0f},
new float[] {0.0f,  1.0f,  0.0f,  0.0f, 0.0f},
new float[] {0.0f,  0.0f,  1.0f,  0.0f, 0.0f},
new float[] {0.0f,  0.0f,  0.0f,  0.5 f, 0.0f},
new float[] {0.0f,  0.0f,  0.0f,  0.0f, 1.0f}};
```

PROGRAM FOR DARK COLOUR WATER MARK

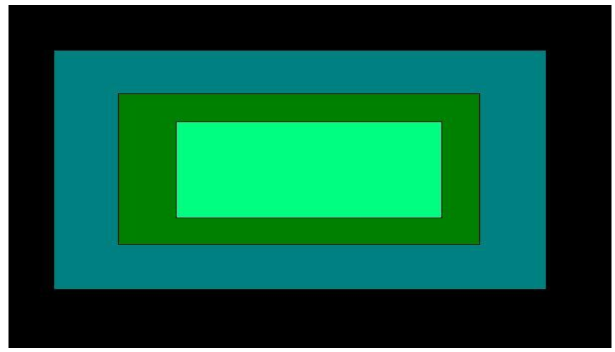
```
float[][] colorMatrixElements = {
new float[] {1.0f,  0.0f,  0.0f,  0.0f, 0.0f},
new float[] {0.0f,  1.0f,  0.0f,  0.0f, 0.0f},
new float[] {0.0f,  0.0f,  1.0f,  0.0f, 0.0f},
new float[] {0.0f,  0.0f,  0.0f,  0.9 f, 0.0f},
new float[] {0.0f,  0.0f,  0.0f,  0.0f, 1.0f}};
```

E IMAGE WITHOUT WATER MARK –



Original Image

F WATER MARK IMAGE –



Water Mark Image

G *OUTPUT IMAGE –*

The software made using C# tends to damage the image characteristics by application of semi visible watermark that damage the commercial value of the Image.

Output Image Distorted by Watermark

H *STEGANOGRAPHIC IMAGE –*

CONCLUSION

The proposed algorithm results has been studied and written in following respects and compared with the previous algorithm:

Actually, the robustness results are expected, since more watermarking energy is embedded into the periphery regions of the image. Due to this reason, the overall watermark energy embedded into the image increases and obviously, detecting the watermark having more energy becomes easier. However, the subjective quality still does not change. [4]

REFERENCES

[1] Shannon C. E., "Communication theory of secrecy systems," Bell Syst.Tech. J., vol. 28, 1949, pp. 656–715.

[2] Boyd C. "Modern data encryption," Electronics & Communication Engineering Journal October 1993, pp 271 – 278.

[3] Schneir Bruce, "Designing Encryption Algorithms for Real People," Comm. of ACM, Vol. 38 no. 3. October 1998

[4] Free encyclopedia "Image Encrypytography" available on line http:// en.wikipedia.org/wiki/ Image-Encrypytography.

[5] Miles E. Smld and Dennis K. Branstad, member, IEEE "Types of Water Mark" proceedings of the IEEE, VOL. 76, NO. 5, May 1988, pp550-559.

[6] M. E. Hellman, "Stegenography," IEEE Trans. Inf. Theory, vol. 26, no. 4, Jul. 1980, pp. 401–406.

CIRCUIT MINIMIZATION IN VLSI

[1]Gaurav Soni ,[2]Rajdeep Singh

[1]Associate Professor, Department of Electronics & Communication Engineering
Amritsar College of Engineering and Technology, Amritsar, India
[2]Lecturer, Department of Electronics & Communication Engineering
College of management and Engineering, kapurthala

***Abstract*— Circuit partitioning is the more critical step in the physical design of various circuit in VLSI. In this partitioning main objective is to minimize the number of cuts. For this PSO algorithm is proposed for the optimization of VLSI inter connection (net list) bipartition. Meanwhile, the corresponding evaluation function and the operators of crossover and mutation are designed. The algorithm is implemented to test various benchmark circuits. Compared with the traditional genetic algorithm (GA) with the same evaluation function and the same genetic operators concerned the hybrid PSO and GA algorithm will give better results.**

Keywords: Partioning, Particle Swarm ptimization, Genetic Algorithm, Hybrid Algorithm.

I. INTRODUCTION

Circuit partitioning/clustering is an important aspect of VLSI design. It consists of dividing a circuit into parts, each of which can be implemented as a separate component (e.g., a chip) that satisfies certain design constraints [1] [2]. One such constraint is the area of the component. The limited area of a component forces the designer to lay out a circuit on several components. Since crossing components incurs relatively large delay, such a partitioning could greatly degrade the performance of a design if not done properly. There has been a large amount of work done in the area of circuit partitioning and clustering [3] [4]. In circuit partitioning, the circuit is divided into two (bi-partitioning) or more (multi-way partitioning) parts. In circuit clustering, the circuit is built up cluster by cluster. To partition designs of large size bottom-up clustering is often combined with top down partitioning. The classical objective of partitioning is to minimize the cut-size, i.e. number of nets spanning two or more parts [5].

The different objectives that may be satisfied by partitioning are:

The minimization of the number of cuts: The number of interconnections among partitions has to be minimized. Reducing the interconnections not only reduces the delay but also reduces the interface between the partitions making it easier for independent design and fabrication. It is also called the min cut problem.

To improve the fitness function is the objective of circuit portioning.Fitness function denotes the improvement in the parameters of the circuit. The more is the fitness function the better is the result of portioning. Area of each partition is also used as a constraint to reduce the fabrication cost with minimum area or as a balance constraint so that partitions are of almost equal size.

Number of partitions appears as a constraint as more number of partitions may ease the design but increase the cost of fabrication and number of interconnections between partitions [6].

II. CIRCUIT MINIMIZATION TECHNIQUES

Particle swarm optimization (PSO):

The basic idea of PSO stems from the behavior of birds, in which each particle or bird keeps track of its coordinates in the solution space which are associated with the best solution that is achieved so far by that particle is called as personal best position (p best) and the another best value obtained so far by any particle is called as global best position (g best). Each particle tries to modify its position using the concept of velocity. The salient features of PSO are:

1. PSO method is based on researches on swarms such as fish schooling and bird flocking.

2. It is a history based algorithm such that in each step the particles use their own behavior associated with the previous iterations.

3. It is easy to implement. Therefore the computation time is less.

GA (Genetic Algorithm):

All genetic algorithms work on a population or a collection of several alternative solutions to the given problem. Each individual in the population is called a string or chromosome, in analogy to chromosomes in natural systems. The population size determines the amount of information stored by the GA. The GA population is evolved over a number of generations. All information required for the creation of appearance and behavioral features of a living organism is contained in its chromosomes. GAs are two basic processes from evolution: inheritance, or the passing of features from one generation to the next, and competition, or survival of the fittest, which results in weeding out the bad features from individuals in the population. The objective of the GA is then to find an optimal solution to a problem .Since GA's are heuristic procedures, modeled as function optimizers, they are not guaranteed to find the optimum, but are able to find very good solutions for a wide range of problems [10].A GA based evolutionary approach for circuit partitioning giving a significant improvement in result quality. Comparative evaluation of genetic algorithm and simulated annealing was done with genetic algorithm giving better results [11] [12]. A new hyper-graph partitioning algorithm was proposed.

Hybrid PSO and GA introduction:

Hybridization of evolutionary algorithms with local search has been investigated in many studies [11] Such a hybrid model is often referred to as a mimetic algorithm. Two global optimization algorithms GA and PSO are combined. Since PSO and GA both work with a population of solutions combining the searching abilities of both methods seems to be a good approach.

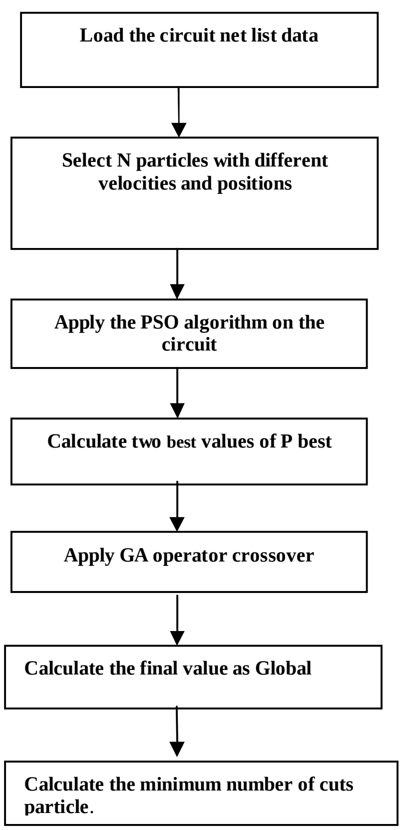


Figure 1 Flowchart of Hybrid PSO_GA Algorithm

Originally PSO works based on social adaptation of knowledge and all individuals are considered to be of the same generation. On the contrary, GA works based on evolution from generation to generation so the changes of individuals in a single generation are not considered. Through GA & PSO have their specific advantage have their specific advantage when solving different algorithm, it is necessary to obtain both their individual feature by combining the two algorithms. The performance of algorithm is described as follow:

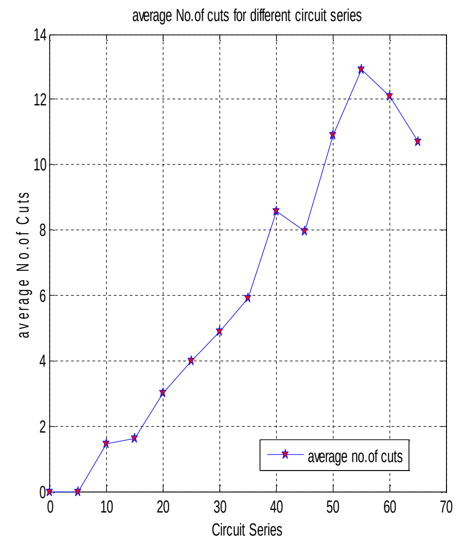


Table 2 Average time of iterations for different netlist

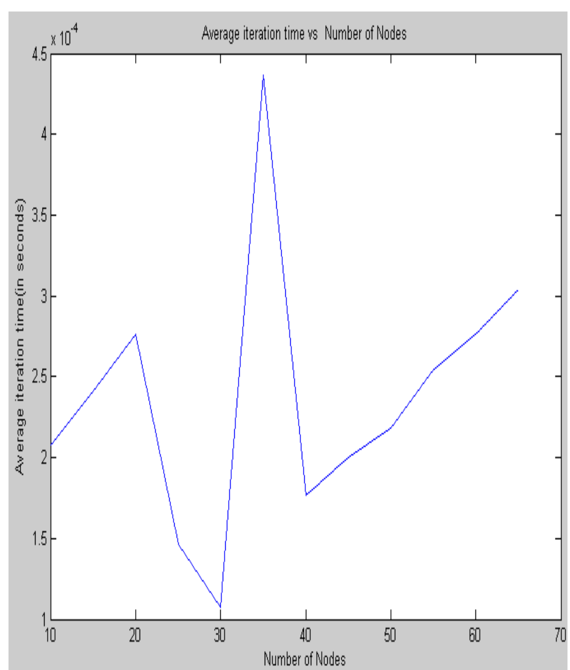


Figure 3. Average iteration time vs Number of Nodes

Table 1 Average Results of partitioning with Hybrid PS0 and Genetic Algorithms with different iterations:

Circuit Series of different net list	**Number of Nodes**	**Number of Files**	**Minimum number of interconnections**
SPP N-10 Series	10	483	1.471074
SPP N-15 Series	15	184	1.630435
SPP N-20 Series	20	121	3
SPP N-25 Series	25	107	3.990654
SPP N-30 Series	30	52	4.884615
SPP N-35 Series	35	31	5.903225806
SPP N-40 Series	40	41	8.585366
SPP N-45 Series	45	28	7.964286
SPP N-50 Series	50	24	10.91667
SPP N-55 Series	55	20	12.9
SPP N-60 Series	60	9	12.11111
SPP N-65 Series	65	7	10.71429

Table 2 Average time of iterations for different netlist

Circuit Series of different net list	**Number of Nodes**	**Number of Files**	**Average iteration time(in seconds)**
SPP N-10 Series	10	483	0.000208
SPP N-15 Series	15	184	0.000241
SPP N-20 Series	20	121	0.000276
SPP N-25 Series	25	107	0.000146
SPP N-30 Series	30	52	0.000107
SPP N-35 Series	35	31	0.000437
SPP N-40 Series	40	41	0.000177
SPP N-45 Series	45	28	0.000200
SPP N-50 Series	50	24	0.000218
SPP N-55 Series	55	20	0.000255
SPP N-60 Series	60	9	0.000276
SPP N-65 Series	65	7	0.000304

III. EXPERIMENTAL RESULT

The Proposed Algorithm is tested on 11 net lists to demonstrate the effect of iteration by using hybrid PSO and GA algorithm on partitioning. The result of partitioning with PSO and GA is given in table 1 Here the no of particles are taken as 5. Fig 2 shows the plot between the no of cuts and netlist series. Fig 3 shows the average time taken by 11 net lists. The figure 3shows graph between circuit series and average no of time for the best of first 50 iteration.

CONCLUSION

In this Paper, Hybrid PSO and GA algorithm is applied to VLSI circuit partitioning problem. By applying the Hybrid PSO and GA algorithms, we are getting an sum of average number of cuts 84.06 which is better by as compared to the results of GA algorithm which was an average of 106.3. So, by this paper, we have concluded that the Hybrid PSO and GA method is better than GA for minimizing the number of cuts in the different circuit series

REFERENCES

[1] N. Sherwani, Algorithms for VLSI Physical Design Automation. Norwell, MA: Kluwer.

[2] C.-W. Yeh , C.-K. Cheng and T.-T. Y. Lin, “Circuit clustering using a stochastic flow injection method,” IEEE Trans. Computer-Aided Design, vol. 14, pp. 154–162, 1995.

[3] Naveed Sherwani, “Algorithms for VLSI Physical Design and Automation”, third edition, Springer (India) Private Limited.

[4] J. Kennedy, R.C. Eberhart, “Particle Swarm Optimization Proc. IEEE Int. Conf. Neural Networks, Vol. 4, pp. 1942-1948 , 1995.

[5] Y. Shi, R. Eberhart, “A modified particle swarm optimizer”, Proc. IEEE Int. Conf. Evol. Comput. Anchorage, AK, 1998, pp. 69-73, 1998.

[6] Kennedy, R.C. Eberhart, Y. Shi, “Comparison between genetic algorithms and particle swarm optimization”, Proc. IEEE Int. Conf. Evol.Comput., Anchorage, pp. 611-616, 2001.

[7] W. Tan, et al., “An efficient multi-way algorithm for balance partitioning of VLSI Circuits”, IEEE International Conference on Computer Design, pp 608-613, 1997.

[8] James Cane, Theodre Manikas, “Genetic Algorithms vs Simulated Annealing: A comparison of approaches for solving circuit partitioning problem”, Technical report, University of Pittsburgh.

[9] George Karypis, Rajat Aggarwal, Vipin Kumar, and Shashi Shekhar, “Multilevel Hyper graph Partitioning: Applications in VLSI Domain”, IEEE 34th Design Automation Conference, Anaheim California, United States, ACM Press New York, NY, USA, pp526-529,1997.

[10] Rehab F. Abdel-Kader,” Genetically Improved PSO Algorithm for Efficient Data Clustering :Second International Conference on Machine Learning and Computing ,pp 71-75,201

[11] S. Naka , T. Yura, Y. Fukuyama, “A hybrid particle swarm optimization for distribution state estimation”, IEEE Trans. on Power Systems, Vol.18, No. 1, pp. 60-68 , 2003.

DIGITAL IMAGE STEGANOGRAPHY

Itender pal Singh
M.Tech (ECE), Amritsar College of Engineering and Technology, Amritsar
Email: itenderpalsingh@gmail.com

Abstract **—Ever since the man started to communicate non-verbally, the need of hiding some secret data has been of great demand. For this purpose several techniques have been proposed from time to time. Steganography is the latest of this series of novel techniques. Better than cryptography, it is able to completely hide the presence of any secret information from everyone. Steganography is of great importance in defense. Sometimes the antinational elements are able to track down the communication frequencies of the army but if the data is sent after hiding it in some image then even if someone who was not intended to have the image gets his hands on it won't be able to do anything. Introduction of extra features like a security password for decoder further enhances the security of the entire system.This paper provides a state-of-the-art review and analysis of the different existing methods of steganography along with some common standards and guidelines drawn from the literature. This paper intends to give an overview of image steganography, techniques, and its algorithms.**

Keywords: Information hiding; Steganography;Steganalysis; LSB.

I. INTRODUCTION

When a steganographic system is developed, it is important to consider what the most appropriate cover Work should be, and also how the stegogramme is to reach its recipient. With the Internet offering so much functionality, there are many different ways to send messages to people without anyone knowing they exist. For example, it is possible that an image stegogramme could be sent to a recipient via email. Alternatively it might be posted on a web forum for all to see, and the recipient could log onto the forum and download the image to read the message. Of course, although everyone can see the stegogramme, they will have no reason to expect that it is anything more than just an image. In terms of development, Steganography is comprised of two algorithms, one for embedding and one for extracting. The embedding process is concerned with hiding a secret message within a cover Work, and is themost carefully constructed process of the two. A great deal of attention is paid to ensuring that the secret message goes unnoticed if a third party were to intercept the cover Work. The extracting process is traditionally a much simpler process as it is simply an inverse of the embedding process, where the secret message is revealed at the end. Two inputs are required for this process. Secret message - usually a text file that contains the message you want to transfer. Cover Work - used to construct a stegogramme that contains a secret message.By now, we are well acquainted with the way MATLAB deals with the images. It treats them like a matrix, which can be two dimensional or three dimensional depending on the format of the image. The elements of the matrix represent the scaled numerical value of the pixel elements of which the image is made. To alter the elements of the matrix wholly or partially means to alter or modify the image. This forms the very basic of steganography. Under the topic of image steganography, we have tried to illustrate four methods by which information can be concealed in an image safely. Comparative analysis at the end in terms of histogram comparison of the "cover image" and "stego image" for various methods of the steganography is also done to understand the way in which the quality of gets effected during information hiding and the degree of invisibility produced.

2.0 This paper aims at illustrating following different types of digital imagesteganography methods:

- **Pixel steganography**
- **LSB steganography(text and image)**
- **Modified LSB steganography**
- **Image in an image steganography.**

2.1 Pixel Steganography

Pixel steganography is the most simple and most crude form of image steganography. MATLAB treats images as two dimensional matrices with matrix element representing the intensity values of the corresponding pixel elements of the image. This method is aimed towards replacing the pixel values of the image with that of the data. Let us suppose that we have an image of the dimension "m×n". Then the image has all in all m*n pixels available for substitution. Let the message that is to be hidden inside the image is denoted by variable "msg" which is "l" characters long.

- Encoding

In the encoding process, the characters of the message are taken one-by-one and they are placed in the image instead of pixels. For this the image as well as the data to be embedded in the image, both are converted into 2 dimensional matrixes. After this, the matrix elements of image are replaced by the elements of the data that needs to be hidden. The proposed algorithm is as under:

Step 1: Read the cover image "I" in MATLAB. This will produce a two dimensional (m×n) matrix. If the image is an RGB image the extract one of the color matrices for data embedding.

Step 2: Read the data as well.

Step 3: For the message "msg" to be l characters long and the image matrix "I" to be of dimension m×n

For a = 1 to l
For b = 1 to m
For c = 1 to n
I(b, c)= msg(a)

This loop will continue to run replacing the matrix elements of image with that of data, row-by-row until the entire data is hidden in the image. The final result will be a "stego image" that contains the invisible data. Although the actual programming is a bit different as we have to run the loops in a proper manner and also define the total dimensions of the data matrix and hide it inside the cover image inside a desired location from which it can be retrieved during decoding process

- Decoding

The decoding process is essentially a reversal of the encoding process. In the decoding process the image is analyzed and the value of the data is extracted from the image. Proposed algorithm

Step 1: The stego image "I" is read generating a 2 dimensional matrix (m×n).

Step 2: The green matrix of the image is extracted.

Step 3: We define any variable say "data" in which secret message will be filled from the matrix. The length of the message or the dimension of the matrix (in case the data is two dimensional) is also extracted from the stego image. Let the length be „l".

Step 4:

For i = 1 to l
For x= 1 to m
For y = 1 to n data(i) = I(m,n)

The process will continue and the data gets extracted into the variable "data". Steganography by pixel substitution is easy to understand and implement. The program doesn't complicate in any language and it yields fruitful results when done on only one matrix (either red or green or blue) of large images. The method has however one drawback. Substituting the entire pixel with data value produces noticeable changes in case large data is embedded in the image. The visibility of the data gets even more prominent for white color. This is because white color is produced for a specific combination of red, green and blue color only and altering any component will produce color that is not white. For other colors, there will be a change in the shade of the color. Images with dark colors serve as good cover images for this method.

2.2 LSB STEGANOGRAPHY

LSB steganography is a modification over the pixel steganography as has less mean square error and more signal to noise ratio. The idea behind this steganography is to embed the characters in the LSB of the image pixel values. Let us consider an example to understand how this method is effective. Consider a number say A= 70. Then A in bit format becomes A= 01000110. The LSB i.e. the least significant bit in this case is 0. LSB is the rightmost bit in any number. As the name suggests this bit has least significance on the value of the data. Now if we change the LSB value from 0 to 1 in this case, A becomes equal to 71, clearly not a lot of deviation from the original value. In terms of colors, deviation to 71 from 70 will produce negligible change in shade of the color which can't be detected by naked eye. Human eye has a limit when it comes to color resolution and in terms of the example given above, the color that the eye will perceive will be decided by the 4 MSBs from the left side itself. So if we convert the data to be hidden into binary format and replace the LSBs of the image pixels with this data, we will get a stego image that is very accurately similar to the original image but has hidden data in it. One of the various algorithms proposed for LSB steganography is "Hide and Seek Approach" which works as follows:

- **Encoding**

Step 1: The image is read and converted into the corresponding matrix. We choose RGB images as it provides three matrices whose concatenation gives us the final image and manipulation in only one matrix would generate very good results.

Step 2: Convert the data into the binary format.

Step 3: Replace the LSBs of the pixel values of the image with those of the data bits and obtain the stego image

- **Decoding**

It is the exact reverse of the encoding process.

Step 1: The stego image is read and the corresponding matrix is obtained.

Step 2: The pixels of the image are read one by one and the LSB of the pixel binary values are extracted and added together to form the hidden data.

Also most famous algorithm is using changing LSB :

Note : pixel consist of 24 bit (3 bytes) the first byte for blue, second for green and the third for red

.

.

.

The original image

Notice the LSB of each byte of image stream bytes.

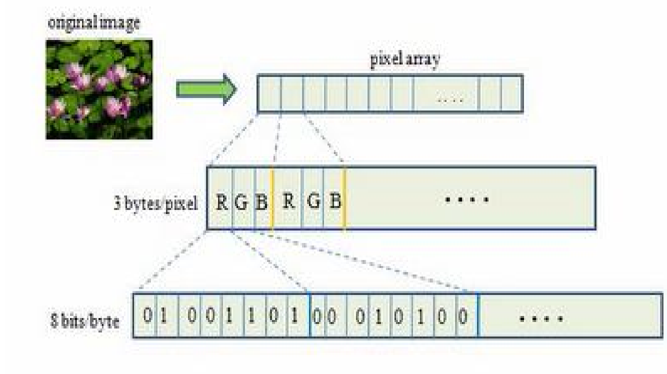


Fig.1 The text which will be hide in the image

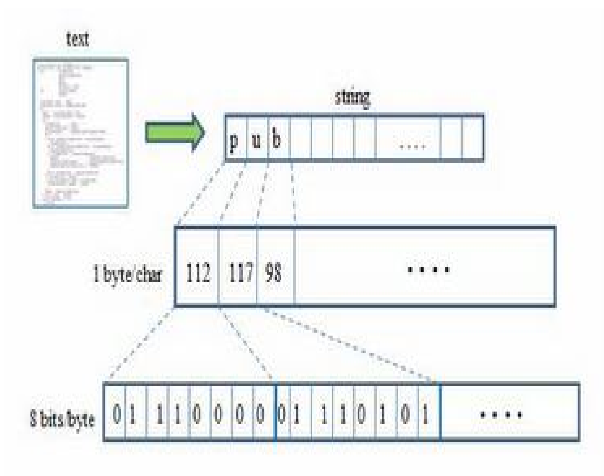


Fig.2 Hiding text into image

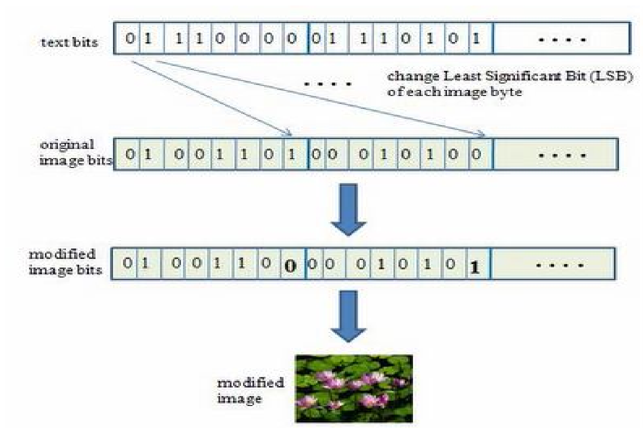


Fig.3 To extract text from image

Humans eyes may not notice the difference between two images.

The method however poses problem in terms of amount of data that can be hidden in one image. As one can clearly conclude from the above discussion that it takes three pixels (i.e. three red, three green and three blue image pixels leading to a total of nine pixels) to hide one character using this method, the actual space available for hiding the data is comparatively less. Though the superb invisibility that the method gives can't be ignored, this much invisibility is not required when it comes to the human eye. Therefore in a trade between the security of the message and the amount of data space available for hiding the secret data, we have developed a modified method of LSB steganography.

2.3 MODIFIED LSB STEGANOGRAPHY

The modified method of LSB steganography is similar to the basic LSB steganography method. The only difference is the number of bits per pixel used for hiding the data. In this proposed modified method of the LSB steganography we are using one pixel (i.e. one pixel each of red, green and blue image matrix) to hide one character of the data. The rest of the process is similar in terms of encoding and decoding.

The clear advantage of using this method is the increased amount of space per image available for steganography without adversely affecting the quality of the image. Image color is defined by the first five MSBs of any pixel, so altering the last three LSBs will not have any such effect on the image quality that could be detected by the human eye. In this method of steganography, the first three MSBs of the data character are being shifted to the last three LSBs of red matrix pixel, then next three bits from data character in the last three LSBs of the green matrix and finally the last two bits of the character of the data is hidden in the last two LSBs of the blue matrix.

The order and the number of bits used per pixel is arbitrarily decided. The alteration in image quality by implementing modified LSB steganography instead of simple LSB steganography is quite higher, for its increased in mean square error and decrease in signal to noise ratio, but still this technique can be implemented in industry for various security purposes which are not quite prone to Visual Attacks like copyrighting,etc.

2.4IMAGE WITHIN IMAGE STEGANOGRAPHY

The steganography discussed so far is about hiding a text message within an image but the hiding data could also be the image itself. This is one of the steganography types added in the project.

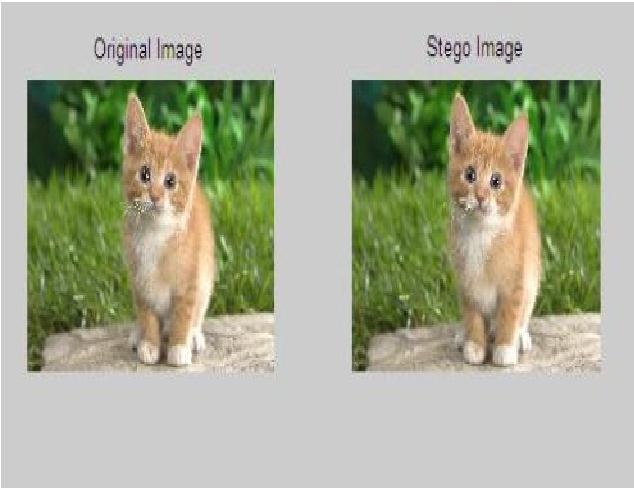


Fig.4 Modified LSB steganography results

The method however is much difficult to understand and implement than simple text in an image steganography. This is because of the fact that in the latter case the data format for both the cover image and the data is the same. Plus the image has a huge amount of complex data that needs to be successfully embedded and retrieved from the cover image.

- **Encoding**

The encoding process involves the following steps:

Step1: The cover image and the image to be hidden are both read and converted into matrix format. One of the three matrices (i.e. red, green and blue) are chosen for alteration.

Step2: In cover image the identity text is embedded (in our case identity text is "yes)"which can be used in decoding purpose to conclude whether the image has been steganalized or not.

Step3: The value of each and every pixel of the secret image is broken down into the binary format and then they are incorporated into the LSBs of the pixel values (in binary format) of the cover image

- **Decoding**

The decoding process is essentially the reversal of the encoding process. It involves the following steps:

Step1: The stego image i.e. the image that has the secret data in it is read and broken down into the matrix form and checked for identity text, whether present or not and if present further steps are followed.

Step2. The pixels of the matrix(Red, Green or Blue-whatsoever has been used in encoding) are one by one converted into binary corresponding values and then the LSBs of each pixel is extracted and stored separately in groups of 8

Step3. The data from the variable is accessed and processed to give us back the original secret image.

The method is very good in terms of hiding the secret image in the cover image with absolutely no visibility of the secret image at all. The user has to ensure that the size of the secret image should be less than or equal to 1/16th of the size of the cover image. This is because of the severe size consideration that comes into action while embedding bits of the secret image into the cover image.

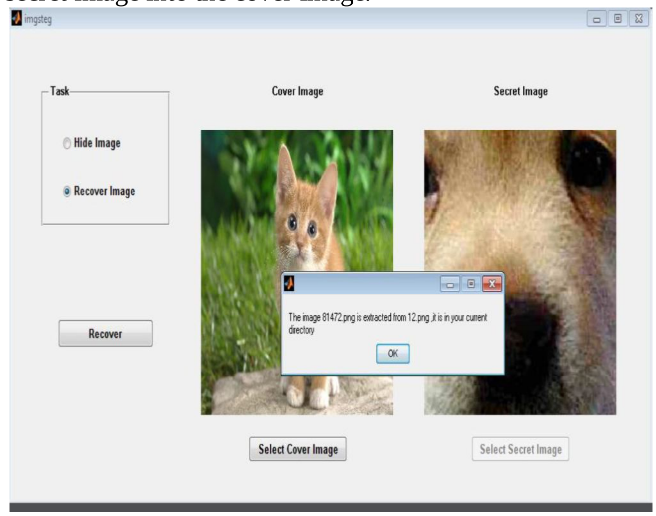


Fig.5Image within image steganography

II. MATHEMATICAL RESULTS

The resultant image in all cases of steganography has an increased value of mean square error and decreased value of signal to noise ratio. Hence as such one would conclude no form of steganography is secure which to some extent is correct as well, as developments in recent past have been such that the fate of steganography itself seems to be obscure rather than providing data obscurity. The implementation of various digital signal processing functions on the image to change it into transfer domain (as discussed in next chapter) has been providing newer options for steganography and hence engineers and researchers switching to transfer domain steganography shouldn't be viewed as a surprise. In earlier type of steganography discussed in this report, often referred to spatial steganography since there is increase in error signal with increase in data a programmer needs to make a compromise between the technique used and amount of data being embedded in an image. Few of the results that we have given below have been calculated for same data and same image using various techniques of steganography.

a)Pixel steganography
Mean Square Error: 0.0092
Signal to Noise Ratio: 68.5114
Capacity: (m×n) (where m and n specify the dimensions of the image

b) LSB Steganography
Mean square error: 0.0042
Signal to Noise Ratio: 71.8538

Capacity: (m×n)/3 (where m and n specify the dimensions of the image)

c) Modified LSB Steganography
Mean Square Error: 0.0043
Signal to Noise Ratio: 71.8167
Capacity: m×n (where m and n specify the dimensions of the image)

From above results which have been calculated in each case for same image and similar data it can be observed that there is clear increase in Signal to Noise Ratio as we employ technique of LSB and Modified LSB instead of Pixel Steganography while comparing LSB and Modified LSB Steganography there is threefold increase in capacity without any noticeable change in Signal to Noise Ratio.In pixel steganography,if we zoom the stegao image,we notice that some data should be embedded but in case of lsb and modified lsb steganography we can't notice the embedded data.

III. STEGANOGRAPHY IN OTHER MULTIMEDIA

4.1 TEXT STEGANOGRAPHY

We can build a simple application that is able to send and receive encrypted messages embedded in Rich Text Format: *.DOC, *.RTF, EMAIL /MessageBody/, etc. The user has the ability to choose the fake text he wants and the program must be able to tell whether or not this fake text will suit the real text.The user can set a different password for every message he sends. This will enable the manager to transmit to two groups two different messages with two different passwords using the same fake text. Thus, you will be able to send encrypted and hidden messages in any source code that you choose! Just take a look at the EXE to see for yourself.

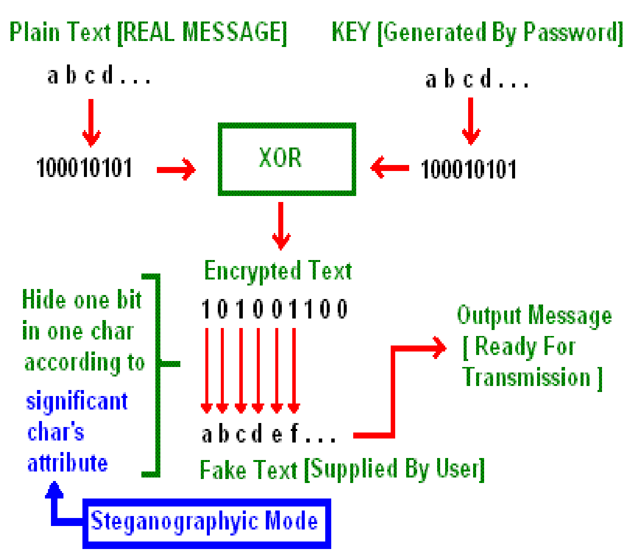


Fig.6 Text steganography

How does it work?
We will not change the text itself, but we will change the unseen attributes of the text. These attributes are many and it is impossible for web servers to track them all. There are lots of Steganographic methods and tracking them will waste huge amounts of processing for uncertain results. Be aware that Steganography is more effective than encryption when used in the right way. The deletion of all attributes is not an option, so we will choose the size and the color. This figure will under-score my point.

4.2 AUDIO STEGANOGRAPHY

We believe that the flexibility of audio steganography is what makes it so potentially powerful. The methods discussed provide users with a large amount of choice and makes the technology more accessible to everyone. A party that wishes to communicate can rank the importance of factors such as data transmission rate, bandwidth, robustness, and noise audibility and then select the method that best fits their specifications. For example, two individuals who just want to send the occasional secret message back and forth might use the LSB coding method that iseasily implemented. On the other hand, a large corporation wishing to protect its intellectual property from "digital pirates" may consider a more sophisticated method such as phase coding, SS, or echo hiding.

4.3 VIDEO STEGANOGRAPHY

As video file is a combination of both image an audio. So, video steganography is nothing but a combination of image and audio steganography. So, the combined evaluations i.e., the evaluations for image and audio steganography can be taken together for the evaluation of video steganography. While doing video steganography, the effect on video has to be kept in mind to achieve a secure communicating media.

CONCLUSION

Many different techniques exist and continue to be developed, while the ways of detecting hiddenmessages also advance quickly. Since detection can never give a guarantee of finding all hidden information, it can be used together with methods of defeating steganography, to minimize the chances of hidden communication taking place. Even then, perfect steganography, where the secret key will merely point out parts of a cover source which form the message, will pass undetected, because the cover source contains no information about the secret message at all. The future of steganography is bright. We can hide data in audio as well as video files so that we have a lot of variety available in terms of cover and it becomes almost impossible for the intruder to make the lucky guess i.e. which data has hidden information in it. Even in image steganography, we can use the DCT compression technique to embed the LSBs of the data in the high definition region of an image which the human eye can never ever detect. We can also hide image in an image. We have ourselves tried to give a method of image in an image steganography here.

REFERENCES

[1] http://www.mathworks.com/access/helpdesk/help/toolbox/images

[2] Image processing toolbox using MATLAB, User's Guide version 2

[3] Digital image processing using MATLAB (Rafael C. Gonzalez, R Woods and S Eddins)

[4] A guide to MATLAB for beginners and experienced users (Brian R. Hunt, Ronald L. Lipsman and Jonathan M. Rosenburg).

[5] A Novel information hiding technique for security by using Image Steganography(M.Sitaram Prasad, S Naganjaneyulu, Ch Gopi Krishna, C Nagaraju/ Journal of Theoretical and Applied Information Technology).

[6] http://en.wikipedia.org/wiki/Steganography.

[7] Roshidi Din, and Hanizan Shaker Hussain, and Sallehuddin Shuib, ―Hiding secret messages in images: suitability of different image file types, WSEAS TRANSACTIONS on COMPUTERS, vol. 6(1), January 2006, pp. 127 -132.

[8] http://www.worldses.org/journals/computers/computers-anuary2007.doc

[9] G. Luo, X. Sun, L. Xiang, and J. Huang, ―An evaluation scheme for steganalysis-proof ability of steganalysis algorithms, International Conference on Intelligent Information Hiding and Multimedia SiProcessing (IIHMSP), vol. 2, 26-28 Nov 2007, pp. 126 - 129.

Different Approaches to Noises and Denoising Techniques for Images

Priya Dhawan[1] and Bimal Kumar[2]
[1] M-Tech, Department of Electronics & Communication Engineering, ACET, Amritsar, India
[2]Associate Professor, Department of Electronics & Communication Engineering, ACET, Amritsar, India
Email: priya.dh20@yahoo.com, engineerbimal@rediffmail.com

Abstract — Image processing is basically the use of computer algorithms to perform image processing on digital images. Images are often degraded by noises. Visual information transmitted in the form of digital images is becoming a major method of communication in the modern age, but the image obtained after transmission is often corrupted with noise. Noise can occur during image capture, transmission, etc. Noise removal is an important task in image processing. The received image needs processing before it can be used in applications. In general the results of the noise removal have a strong influence on the quality of the image processing technique. Image denoising involves the manipulation of the image data to produce a visually high quality image. Wavelet transforms have become a very powerful tool for de-noising an image. One of the most popular methods is wiener filter. In this work four types of noise (Gaussian noise , Salt & Pepper noise, Speckle noise and Poisson noise) is used and image de-noising performed for different noise by Mean filter, Median filter and Wiener filter Selection of the denoising algorithm is application dependent. Hence, it is necessary to have knowledge about the noise present in the image so as to select the appropriate denoising algorithm. The filtering approach has been proved to be the best when the image is corrupted with salt and pepper noise. The wavelet based approach finds applications in denoising images corrupted with Gaussian noise. In the case where the noise characteristics are complex, the multifractal approach can be used. A quantitative measure of comparison is provided by the signal to noise ratio of the image.

I. INTRODUCTION

Noise is the result of errors in the image acquisition process that results in pixel values that do not reflect the true intensities of the real scene. Image de-noising is an vital image processing task i.e. as a process itself as well as a component in other processes. The image captured by the sensor undergoes filtering by different smoothing filters and the resultant images. The fundamental problem of image processing is to reduce noise from a digital color image. The two most commonly occurring types of noise are i) Impulse noise, ii) Additive noise (e.g. Gaussian noise) and iii) Multiplicative noise (e.g. Speckle noise) Noise reduction is the process of removing noise from a signal. Denoising is often a necessary and the first step to be taken before the images data is analyzed. It is necessary to apply an efficient denoising technique to compensate for such data corruption. The important property of a good image denoising model is that it should completely remove noise as far as possible as well as preserve edges. Traditionally, there are two types of models i.e. linear model and non-liner model. Generally, linear models are used. The benefits of linear noise removing models is the speed and the limitations of the linear models is, the models are not able to preserve edges of the images in a efficient manner i.e the edges, which are recognized as discontinuities in the image, are smeared out. On the other hand, Non-linear models can handle edges in a much better way than linear models. The scope of the paper is to focus on different noises occurring in the images and noise removal techniques for natural images

II. TYPES OF NOISES

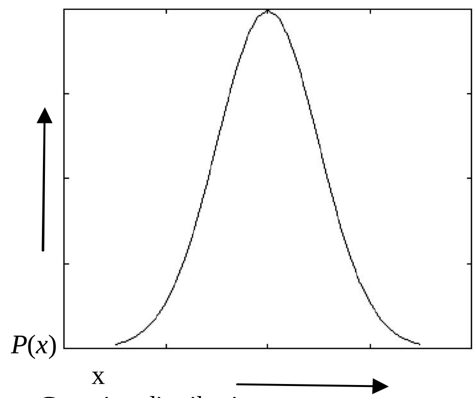
Noise is undesired information that contaminates the image.,or we can say, Noise is any degradation in the image signal caused by external disturbance. Theerrors in the image will appear on the image output in different ways depending on the type of disturbance in the signal. In the image denoising process,information about the type of noise present in the original image plays a significant role.Typical images are corrupted with noise modeled with either a Amplifier noise(Gaussian noise), Salt-and-pepper noise (Impulse noise),Shot noise,Quantization noise (uniform noise),Film grain,on-isotropic noise,Speckle noise (Multiplicative noise) and Periodic noise. Noise is present in an image either in an additive or multiplicative form. An **additive noise** follows the rule ***w*(*x*, *y*) = *s*(*x*, *y*) + *n*(*x*, *y*) ,** while the **multiplicative noise** satisfies ***w*(*x*, *y*) = *s*(*x*, *y*)× *n*(*x*, *y*) ,** where *s*(*x*,*y*) is the original signal, *n*(*x*,*y*) denotes the noise introduced into the signal to produce the corrupted image *w*(*x*,*y*), and (*x*,*y*) represents the pixel location. The above image algebra is done at pixel level. Image addition also finds applications in image morphing(Morphing is a special effect in motion pictures and animations that changes (or morphs) one image into another through a seamless transition). By image multiplication, we mean the brightness of the image is varied.

Amplifier noise (Gaussian noise)

The standard model of amplifier noise is additive, Gaussian, independent at each pixel and independent of the signal intensity, caused primarily by Johnson–Nyquist noise (thermal noise), including that which comes from the reset noise of capacitors ("kTC noise") In color cameras where more amplification is used in the blue color channel than in the green or red channel, there can be more noise in the blue channel. Amplifier noise is a major part of the "read noise" of an image sensor, that is, of the constant noise level in dark areas of the image . In Gaussian noise, each pixel in the image will be changed from its original value by a (usually) small amount. A histogram, a plot of the amount of distortion of a pixel value against the frequency with which it occurs, shows a normal distribution of noise.Gaussian noise is evenly distributed over the signal. This means that each pixel in the noisy image is the sum of the true pixel value and a random Gaussian distributed noise value. As the name indicates, this type of noise has a Gaussian distribution, which has a bell shaped probability distribution function given by,

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

x is the gray level, μ is the mean, σ is the standard deviation and σ^2 is the variance



Gaussian distribution

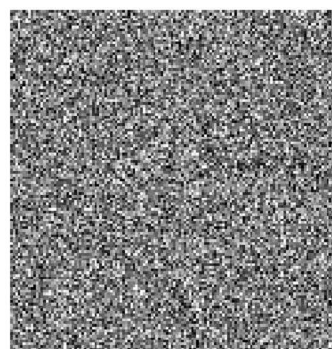


Image 1: Gaussian noise (mean=0, variance 0.05) [4]

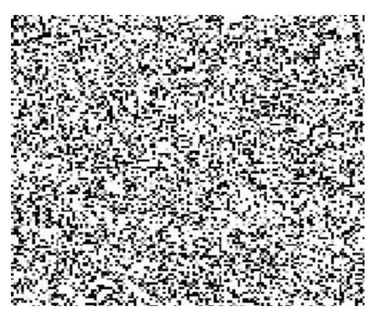


Image 2: Gaussian noise (mean=1.5, variance 10)[4]

Image 2 illustrates the Gaussian noise with mean (variance) as 1.5 (10) over a base image with a constant pixel value of 100.

Salt-and-Pepper Noise (Impulse Noise)

Salt and pepper noise is sometimes called impulse noise or spike noise or random noise or independent noise This is caused generally due to errors in data transmission. It has only two possible values, *a* and *b*. The probability of each is typically less than 0.1. It has only two possible values, *a* and *b*. The probability of each is typically less than 0.1. The corrupted pixels are set alternatively to the minimum or to the maximum value, giving the image a “salt and pepper” like appearance. Unaffected pixels remain unchanged. Salt and pepper degradation can be caused by sharp and sudden disturbance in the image signal. Generally this type of noise will only affect a small number of image pixels. When viewed, the image contains dark and white dots, hence the term salt and pepper noise. An image containing salt-and-pepper noise will have dark pixels in bright regions and bright pixels in dark regions. This type of noise can be caused by analog-to-digital converter errors, bit errors in transmission etc. For an 8-bit image, the typical value for pepper noise is 0 and for salt noise 255. The salt and pepper noise is generally caused by malfunctioning of pixel elements in the camera sensors, faulty memory locations, or timing errors in the digitization process.

Image 3: salt and pepper noise[6]

Speckle Noise

While Gaussian noise can be modeled by random values added to an image, speckle noise can be modeled by random values multiplied by pixel values hence it is also called multiplicative noise. This type of noise occurs in almost all coherent imaging systems such as laser, acoustics and SAR(Synthetic Aperture Radar) imagery. Speckle noise in SAR is generally more serious, causing difficulties for image interpretation. The source of this noise is attributed to random interference between the coherent returns. It is caused by coherent processing of backscattered signals from multiple distributed targets.

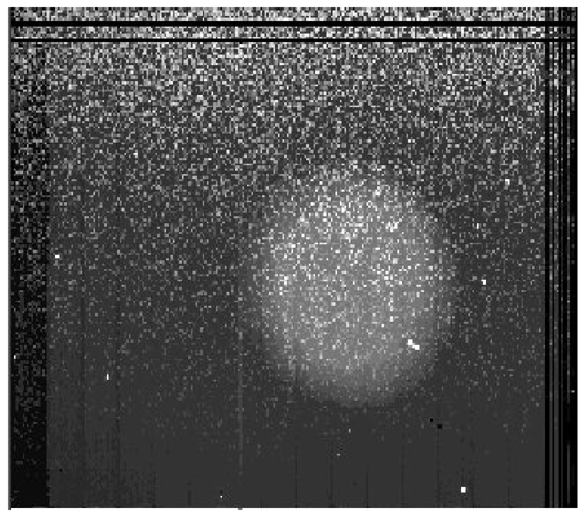


Image 4: speckle noise[6]

Shot Noise

The dominant noise in the lighter parts of an image from an image sensor is typically that caused by statistical quantum fluctuations, that is, variation in the number of photons sensed at a given exposure level; this noise is known as photon shot noise. Shot noise follows a Poisson distribution, which is usually not very different from Gaussian.

In addition to photon shot noise,there can be additional shot noise from the dark leakage current in the image sensor; this noise is otherwise known as "dark shot noise" or"dark-current shot noise".

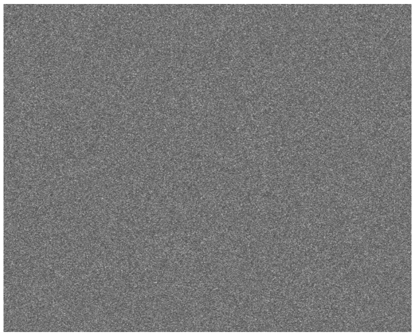


Image 4: Shot Noise[4]

Quantization noise (uniform noise)

The noise caused by quantizing the pixels of a sensed image to a number of discrete levels is known as quantization noise; it has an approximately uniform distribution, and can be signal dependent, though it will be signal independent if other noise sources are big enough to cause dithering(Dithering is a technique to simulate the display of colors that are not in the current color palette of an image), or if dithering is explicitly applied.

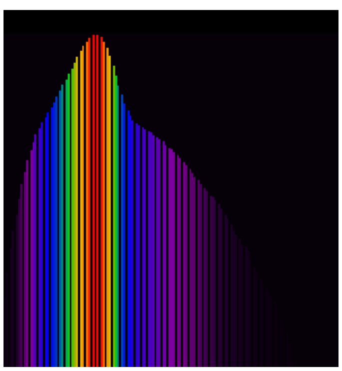


Image 5:Quantization noise[6]

Brownian Noise

Brownian noise, also known as Brown noise or red noise, is the kind of signal noise produced by Brownian motion, hence its alternative name of random walk noise. The term "Brown noise" comes not from the color, but after Robert Brown, the discoverer of Brownian motion. Brownian noise comes under the category of fractal or $1/f$ noises. The mathematical model for $1/f$ noise is fractional Brownian motion [Ma68]. Fractal Brownian motion is a non-stationary stochastic process that follows a normal distribution. Brownian noise is a special case of $1/f$ noise. It is obtained by integrating white noise.

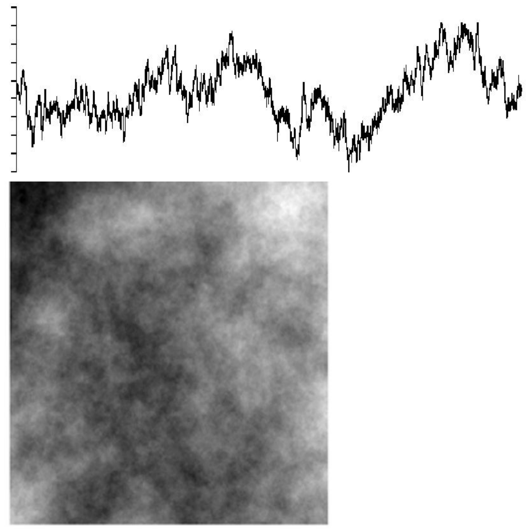


Image 6:Brownian noise[6]

III. Classification of Denoising Algorithms

Spatial Filtering

A traditional way to remove noise from image data is to employ spatial filters. Spatial filters can be further classified into non-linear and linear filters

Non-Linear Filters

With non-linear filters, the noise is removed without any attempts to explicitly identify it. Spatial filters employ a low pass filtering on groups of pixels with the assumption that the noise occupies the higher region of frequency spectrum. Generally spatial filters remove noise to a reasonable extent but at the cost of blurring images which in turn makes the edges in pictures invisible. In recent years, a variety of nonlinear median type filters such as weighted median, rank conditioned rank selection, and relaxed median have been developed to overcome this shortcoming.

Median Filter

The Median filter is a nonlinear digital filtering technique, often used to remove noise. Median filtering is very widely used in digital image processing because under certain conditions, it preserves edges whilst removing noise. The main idea of the median filter is to run through the signal entry by entry, replacing each entry with the median of neighboring entries. To run a media filter:

1. Consider each pixel in the image
2. Sort the neighboring pixels into order based upon their intensities
3. replace the original value of the pixel with the median value from the list

Median filtering is done by, first sorting all the pixel values from the surrounding neighborhood into numerical order and then replacing the pixel being considered with the middle pixel value. Note that the median value must be written to a separate array or buffer so that the results are not corrupted as the process is performed. if the window has an odd number of entries,then the median is simple to define: it is just the middle value after all the entries in the window are sorted numerically. For an even number of entries, there is more than one possible median.The median filter is a robust filter . Median filters are widely used as smoothers for image processing, as well as in signal processing and time series processing. A major advantage of the median filter over linear filters is that the median filter can eliminate the effect of input noise values with extremely large magnitudes. (In contrast, linear filters are sensitive to this type of noise - that is, the output may be degraded severely by even by a small fraction of anomalous noise values). The output y of the median filter at the moment t is calculated as the median of the input values corresponding to the moments adjacent to t:

$$y(t) = \text{median}((x(t-T/2),x(t-T1+1),\ldots,x(t),\ldots,x(t+T/2)).$$

where t is the size of the window of the median filter. Besides the one-dimensional median filter described above, there are two-dimensional filters used in image processing .Normally images are represented in discrete form as twodimensional arrays of image elements, or "pixels" - i.e. sets of non-negative values Bij ordered by two indexes –

$$i = 1,\ldots, Ny \text{ (rows) and } j = 1,\ldots,Ny \text{ (column)}$$

.where the elements Bij are scalar values, there are methods for processing color images, where each pixel is represented by several values, e.g. by its "red", "green", "blue" values determining the color of the pixel.

Image7: Input to median filter[4]

Image8: Output from median filter[4]

Linear Filters

Linear filter used to remove certain types of noise. Averaging or Gaussian filters are appropriate for this purpose. Linear filters also tend to blur sharp edges, destroy lines and other fine image details, and perform poorly in the presence of signal-dependent noise

Mean Filter

A mean filter acts on an image by smoothing it; that is, it reduces the intensity variation between adjacent pixels. The mean filter is nothing but a simple sliding window spatial filter that replaces the center value in the window with the average of all the neighboring pixel values including itself. By doing this, it replaces pixels, that are unrepresentative of their surroundings. The Mean Filter is a linear filter which uses a mask over each pixel in the signal. Each of the components of the pixels which fall under the mask are averaged together to form a single pixel. This filter is also called as average filter. The Mean Filter is poor in edge preserving.Image 9 is the one corrupted with salt and pepper noise with a variance of 0.05.

Image9: Input to mean filter[4]

The output image after Image 9 is subjected to mean filtering is shown in Image 10. It can be observed from the output that the noise dominating in Image 9 is reduced in Image 10. The white and dark pixel values of the noise are hanged to be closer to the pixel values of the surrounding ones. Also, the brightness of the input image remains unchanged because of the use of the mask, whose coefficients sum up to the value one. The mean filter is used in applications where the noise in certain regions of the image needs to be removed. In other words, the mean filter is useful when only a part of the image needs to be processed

Image10: Image after mean corrupted with salt and pepper noise filtering[4]

Adaptive Filter

An adaptive filter does a better job of denoising images compared to the Mean or averaging filter. The fundamental difference between the mean filter and the adaptive filter lies in the fact that the weight matrix varies after each iteration in the adaptive filter while it remains constant throughout the iterations in the mean filter.

Adaptive filters are capable of denoising non-stationary images, that is, images that have abrupt changes in intensity. Such filters are known for their ability in automatically tracking an unknown circumstance or when a signal is variable with little a prior knowledge about the signal to be processed . In general, an adaptive filter iteratively adjusts its parameters during scanning the image to match the image generating mechanism. This mechanism is more significant in practical images, which tend to be non-stationary.

The adaptive filter is more selective than a comparable linear filter, preserving edges and other high-frequency parts of an image.

Image11: Input to adaptive filter[4]

Image12: Image after adaptive Corrupted with salt and pepper noise filtering[4]

CONCLUSION

In this paper, we discussed different types of noises occurring in images and also different types of filtering techniques for removing noises in color image. Furthermore, we presented and compared results for these filtering or denoising techniques. The results obtained so far for salt and pepper noise, the median filter is optimal compared to mean filter and adaptive filter. It produces the maximum SNR for the output image compared to the linear filters considered. The adaptive filter proves to be better than the mean filter but has more time complexity Since selection of the right denoising procedure plays a major role, it is important to experiment and compare the methods. As future research, we would like to work further on the comparison of the denoising techniques. One area is in improving the de-noising along the edges as the method we used did not perform so well along the edges. If the features of the denoised signal are fed into a neural network pattern recognizer, then the rate of successful classification should determine the ultimate measure by which to compare various denoising procedures.

REFERENCES

[1] Patidar Pawan,et.al,"Image de-noising by various filters for different noises", International Journal of Computer Applications (0975 – 8887) Volume 9– No.4, November 2010.

[2] Motwani.C.Mukesh,et.al,"Survey of image denoising techniques"

[3] Mythili.C,Kavitha.V.Dr.,"Efficient technique for colour image noise reduction", The Research bulletian of Jordan,Vol 2

[4] Dangeti Sarita,".Denoising Techniques,A Comparison"

[5] Buades Antoni,et.al, "On image denoising methods"

[6] L. Alvarez and P-L. Lions and J-M. Morel, "Image selective smoothing and edge detection by non-linear di®usion (II)", SIAM Journal of numerical analysis 29, pp. 845-866, 1992.

[7] M. Emre Celebi, Hassan A. Kingravi, Y. Alp Aslandogan "Nonlinear vector filtering for impulsive noise removal from color Images". Journal of Electronic Imaging. 2007, 1-21.

[8] A. Buades, B. Coll, and J. M. Morel, "A review of image denoisingalgorithms, with a new one," Multiscale Modeling and Simulation (SIAM interdisciplinary journal), vol. 4, no. 2, pp. 490–530, 2005

Demand Side Management integrate With Distribution Generation in Active Network

Ajaypal Singh
Assistant Prof., Department of Electrical and Electronics Engineering, ACET, Amritsar
Email id. ajay.0001@yahoo.co.in

Abstract:- In the last decay, it has been witnessed an increasing allocation of distribution generation (DG) in the MV distribution networks, due to technology improvement, energy market liberalization and environmental issues. Every year, Distribution Network Operators (DNOs) receive several requests for installations of new generators in the networks. This situation is likely implying a revolution of the distribution networks, actually categorized by radial structure and passive operation, in order to exalt the benefits and reduce the drawbacks that DG can introduce. Therefore, it is predictable near future the implementation of an active management of the distribution networks.[1] Active network in the distribution of the future will require the large use of automatic control system to control loads and generators, with the aim of allowing the participation DERs to the energy and service market and of maintain a sufficient level of system reliability. Into the scenario, DG and Demand Side Management (DSM) action should become options that distributors have to consider in order to solve electric supply problems, as requested by the more recent EU directives. In this thesis it has been analyzed how DSM policies can be avalid opportunity to facilitate the development of DG in a given distribution system and which economical benefits the utilities can derive by the complementary employ ofboth these distributed resources.Simulation studies have been performed on a real distribution network, showing the effect of DSM action on the growth of DG in the distribution system and on the technical and economic benefits,they permit to realize.

INTRODUCTION

Distribution generation is an electric power source connected directly to the distribution network or on the customer side of the meter .The definition given does purposely lack

Information concerning

- Power rating and technology
- Environmental impacts
- Delivery area
- Mode of operation

This allow a more general discussion of various aspects, since many issues regarding the interaction between DG and the existing grid- as e.g. deep and shallow connection charges or protection aspects – are similar for the different types of DG.[5]

Power rating categories are additionally introduced as:

- Micro distribution generation < 5 KW
- Small distribution generation 5 KW – 5 MW
- Medium distribution gen. 5 MW – 50 MW
- Large distribution generation > 50 MW

Demand Side Management (DSM) is another option which is equally important and beneficial as that of distributed generation in improving the energy scenario. It is observed that, by inclusion of new energy sources we are only supplying the increasing demand, while an inherent deficit prevails.Therefore, on the demand side, more efficient ways of utilization of the available energy has to be employed. Some measures already in place include load control and/or load shifting by utility companies in order to obtain a better and profitable load curve. In times to come, the possibility of practicing DSM in real time is being studied and also implemented on pilot basis by some Electrical Utility companies. By introducing Smart Metering at load centre and intelligent devices for real time control, it is possible to monitor and control the energy demand more economically, increase profitability and also conserve energy. Connection of DG and introducing schemes like DSM in the distribution system would calls for improved control, automation and protection systems. Automation of distribution network operations can be beneficial in maintaining nominal operating conditions and offer supply security and protection, in the ever changing network configuration of the present day distribution system. The concept of Active Network Management (ANM) is being studied in this regard. ANM applies distribution automation to control network parameters for better operation with DG. For example, allow more connections of energy sources on the network by active voltage control methods. Demand side management could also be part of ANM for load control and hence better energy conservation and consequent environmental benefits.

Electric utility demand side management (DSM) programs seek to reduce electric loads from the end-user or consumer through energy efficiency (EE) and load shaping measures.[6] Successful demand side management programs, stimulated by state incentives, requirements, and financial structures, can reduce the amount of electricity a utility must provide, decreasing the need for new generation sources reduction in electricity demand generally translates to reduction in greenhouse gas emissions produced by generation. Estimations of DSM load reduction potential inform utility management strategies and climate policy. The accuracy of such estimates can affect plans for future programs, policies, generation sources, etc. The purpose of demand side management is energy conservation and the salient features of energy conservation are [3]:

- Setting up of energy conservation standards for any equipment or appliance
- consuming, generation, transmitting or supplying energy.
- Certain industries, establishments and user ofenergy to be notified as designated consumers keeping in the view of intensity and quality of energy consumed.
- Mandatory energy audit for all designated consumers, as and when required by the designated authority.
- Promotion of mass awareness at both the Central and the State levels for energy conservation, consumer education and guidance.
- Government to take step to encourage preferential use of energy efficient equipment and appliances.
- Constitution of an Energy Conservation Funds at the Centre and the State for utilizing any grant or loans made available for promoting energy conservation.

LITRATURE REVIEW

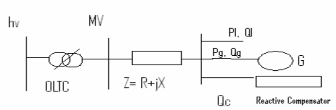
The promise of a new dawn of widespread DG sources proliferated throughout distribution systems has lead to a certain degree of hype in the electricity industry. Nonetheless, DG does offer many benefits as well as presenting numerous challenges. A large amount of research has been carried out investigating the impact of DG. A selection of the more relevant publications is given here. A number of publications have looked at optimizing the placement and sizing of DG based on various criteria. Wallace and Harrison (2003), the authors employ an optimal power flow (OPF) technique to maximize DG capacity with respect to voltage and thermal constraints. Short circuit levels, short circuit ratio, equipment ratings and losses are not considered. The effect of network sterilization is clearly demonstrated by comparison between allocating generation to buses individually rather than as a group. Vovos et al. (2005), a method is presented utilizing OPF for the allocation of generation capacity, which includes a detailed fault level constraint. Kuri et al. (2004) genetic algorithms were used to place generation such that losses, costs and network disruption were minimized and the rating of the generator maximized. The constraints considered were voltage, thermal, short circuit and generator active and reactive power capabilities. Generation is placed in single units at individual buses, while ignoring the interdependence of the buses and the network sterilization that can result from improper DG placement. El-Khattam et al. (2004) the authors use a heuristic approach to determine the optimal DG size and site from an investment point of view. Once again short circuit constraints are not considered and the focus of the objective function is on optimal investment rather than maximizing renewable energy. It uses a cost benefit analysis to evaluate various placements of DG. Celli et al. (2005), a multiobjective planning strategy is presented using a genetic algorithm to identify the best compromise DG sizing and sitting. Chowdhury et al. (2003), a probabilistic reliability model is presented to determine the impact of DG for use in distribution planning studies. In McDermott and Dugan (2003), the impact of DG on reliability and power quality is measured. Reliability and power quality indices were applied to a sample feeder to assess the impact. Work has also been done evaluating the contribution of wind generation, in particular, to reliability (Clark and Miller, 2006; Karki et al., 2006). The issues of load growth and load patterns in distribution planning are discussed in Willis (2004) and a multi stage approach to planning is described in Kuwubara and Nara (1997). Quezada et al. (2006), the amount of losses incurred with increasing penetrations of various DG sources is examined. Wang and Nehrir (2004), the authors propose a method which places DG at the optimal place along feeders and within networked systems with respect to losses. The allocation of losses to distributed generators in the network has been addressed in Costa and Matos (2004). Previous work has at tempted to quantify the net benefits of DG (Chiradeja and Ramakumar, 2004), where a number of benefits such as reduced losses and voltage profile were assessed.

PROBLEM FORMULATION

The voltage rise caused by DG is a well known effect and can be illustrated using the simple circuit shown in Fig. 1. This figure represents the basic features of a distribution system into which a distributed generator, G, is connected at the MV level.[4] The generator is connected to the distribution system through a weak network (high equivalent impedance Z at the connection point) and an on-load tap changing transformer. The scheme is completed with a load and a reactive power compensator. The voltage at bus 2 is expressed by (1), where Pg and Qg are the generated active and reactive power drawn, Pl and Ql ate the active and reactive power drawn by the load, Qc is the reactive power required by the reactive power compensator. Pdsm is the portion of total load that can be moved from peak to off-peak hours.

$V^2=V^1+R(Pg-Pl\pm Pdsm)\pm X(\pm Qg-Ql\pm Qc)$ ---------(1) Eq.

This simple equation can be used to qualitatively analyse the relationship between the voltage at bus 2, and the amount of generation that can be connected to the distribution network, as well as the impact of alternative control action.



As indicated earlier, the general practice in distribution networks is to limit the capacity of the connected DG based on the extreme conditions of minimum load (Pl=0, Ql=0 and Pdsm=0) and maximum generation (Pg=Pgmax). Under such hypothesis the maximum level of generation that can be connected to the network (with no reactive power compensation and generator at unitary power factor) is expressed by Eq.(2)

$Pgmax=(V^2max-V1)/R+Pdsm/R$--------------(2) Eq.

Where V2max represents the maximum voltage rise admitted at busbar 2. It is trivial to observe that, by applying the rule, only few generators can be connected without reinforcing the network. Of course, the reactive power regulation can help to increase the maximum allowable penetration of DG but for sake of simplicity, in this paper only active power has been considered. The probability of occurrence of the extreme situation in (2) is generally quite low and for this reason in it has been supposed to actively managing, the network so the generation is curtailed during inadmissible over-voltages at busbar 2.

The probability of the coincidence of minimum load demand and maximum generation will determine the total annual energy curtailed. As the price of electricity is primatily driven by load demand, and generation curtailment occurs typically during period of loaw load, the value of this energy curtailed is likely to be telatively low.

In this thesis a similar approach has been followed but, in order to make a comparison, the effect of a DSM action able reduce the peak load by increasing the energy demand during the low load condition has been investigated. The Eq.(2) shows a linear relationship between the portion of loads involved in the DSM program and the increasing of maximum allowable level of DG.

RESULT AND DISCUSSION

In order to study the performance of demand side management action, some test has been performed on the small test system depicted in fig 1.[3] In order to be able to make a comparison with the result presented, the generation curtailment has been also applied to the same case study. To simplify the model only two energy prices have been considering, one for the off-peak hour. Here load have been grouped into some lumped load. During off-peak hour the generator connected with load bus has been assumed to produce constant power this is because load is constant, under such conditions if load power we assumed it is less than the power coming from bus 1(assume line losses negligible) than result has been reactive power absorbed by distribution generator unit and reactive power compensation unit. By varying the power of output load up-to such instant that it is not cross the power coming from bus 1, we take some results which show what is being the maximum value of distribution generation unit under such conditions. In second case during peak load conditions (when load has been cross the limit of power coming from bus 1) we have assume some peak values of load power. Under such condition the generator and reactive power compensation device will give reactive power to the load, which may increases line losses. These losses is being reduce by using shunt capacitor and obtain the result which shows what has been the distribution generation under such conditions corresponding with the variation with load. The result obtained in peak and off-peak hours is by using mathematical relation and by using computer programming technique (C++).

After calculating the value of distribution generation produce by distributor generator at different instant of load. Our next aim in this thesis has to determine the cost of distributor generator at different instant of load. This has been done by using the computer programming technique. The result obtain here has clearly showed that greater has been the load more be the cost of distributor generator unit and vice versa. Clearly, the benefit of demand side management depends on the fee paid by DG. Whether the DSM fee is established equal to energy price during the off-peak hours, the generation curtailment is slightly more advantageous, especially for the highest value of DG installed. If the cost of DSM was set equal to the difference, which has to be recognized to those customers that change their load curve to increase the energy demand during the off-peak hours, the DG owner revenue is greater than in the curtailment case. From this simple example, it is evident that the generation curtailment is more convenient than the load control. Furthermore, whether capital costs are taken into account, the comparison is much less favorable to the control, because it is more expensive than the other "dynamic" method such as generation curtailment or constraining and power factor control

REFERENCES

[1] ENIRDGnet, D2, Concepts and Opportunities of DG, WP1 – concepts and opportunities of DG: the driving european forces and trends

[2] ENIRDGnet, D3, The Driving European Forces and Trends, WP1 - concepts and opportunities of DG: the driving european forces and trends.

[3] http://www.distributed generation.com/dpca/ what.html.

[4] Impact of increasing contribution of dispersed generation on the Power System, CIGRE SC #37,1998.

[5] Distributed generation in liberalized electricitymarket, IEA Publications, 2002.

[6] Bailey, Bruce H., et. al., (1993). "Evaluation of a Demand Side ManagementPhotovoltaic System."IEEE Transactions on Energy Conversion.

[7] Davidson, Keith G. and Gerald Braun (1993). "Thinking Small: Onsite Power Generation Many Soon Be Big." Public Utilities.

[8] Hoff, Thomas E. and Christy Herig (1996). "Managing Risk Using Renewable Energy Technologies," in The Virtual Utility: Accounting, Technology & Competitive Aspects of the Emerging Industry. Shimon Awerbuch and Alistair Preston, eds., KluwerAcademic Publishers, Norwell: MA, forthcoming.

Wireless Networks Architecture Using Cognitive Radio

Gourav Soni[1], Narinder Sharma[2] and Sandeep kaushal[3]
[1, 2, 3]Associate Professor, Department of Electronics & Communication Engineering, ACET, Amritsar, India

***Abstract*:- Cognitive radio (CR) is seen as a solution to the current low-usage of the radio spectrum and the problem of the fixed spectrum allocation. Quality of service (QoS) provisioning is an important issue in the deployment of broadband wireless access networks with realtime and non-real-time traffic integration in wireless spectrum. The connection-level and packet-level scheduling scheme is essential to guarantee the QoS requirements of different service classes. solution can expand the capacity of the WiMAX system while providing QoS to real-time and non-real-time traffic.. In this paper Cognitive Radio architecture for intelligent wireless networks, its requirements, spectrum sensing and future scopes are discussed.**

Keywords:-Cognitive Radio, Spectrum Sensing, Intelligent wireless networks.

I. INTRODUCTION

Recently, broadband wireless access becomes ubiquitous. This makes the already heavily crowded radio spectrum much scarcer [1]. Cognitive radio (CR) is a promising technology to alleviate the increasing stress on the fixed radio spectrum. In cognitive radio based opportunistic spectrum access (OSA) networks, the secondary (unlicensed) users can periodically sense and identify available channels which are referred to as white spaces (WS) as the parts of the spectrum not in active use by the primary users. Based on the results of spectrum sensing, the secondary users dynamically tune its transceivers to the identified WSs to access the wireless channel without disturbing the communications of the primary users.Cognitive radio is a particular extension of software radio that employs model-based reasoning about users, multimedia content, and communications context. Cognitive radio (CR) is one of the new long term developments taking place and radio receiver and radio communications technology. A cognitive radio may be defined as a radio that is aware of its environment, and the internal state with knowledge of the elements and any stored pre-defined objectives can make and implement decisions about its behavior. In general, the cognitive radio may be expected to look at parameters such as channel occupancy, free channels, the type of data to be transmitted and the modulation types that may be used. It must also look at the regulatory requirements. A cognitive radio is a radio which can sense its environment and has the capability to adapt some of its features, such as carrier frequency, modulation, transmission bandwidth and transmission power allowing dynamic reuse of the available spectrum. Radio spectrum resource is of fundamental importance in wireless communications. The access to radio spectrum in the present days is largely based on fixed spectrum allocation principle. With the deployment of more wireless applications/services, most of the available spectrum has been well allocated, thus many countries are facing the problem of spectrum scarcity. Cognitive radio is the enabling technology for dynamic spectrum access. The essential components in a cognitive radio system include spectrum sensing, Cognitive medium access control and Cognitive networking. With the rapid deployment of various wireless systems, the limited radio spectrum is becoming crowded increasingly. On the other hand, it is evident that most of the allocated spectrum experience low utilization [1]. Cognitive radio (CR) [2], which allows secondary users /networks to coexist with the primary users /networks through spectrum sharing, can dramatically improve the spectrum utilization, thus supporting more new services.[3] In challenging scenarios, such as a battlefield ad hoc network, a communication unit may not be able to operate in a fixed assigned band due to environmental constraints or application constraints, but rather have to search for an appropriate band in which to operate from time to time. On the other hand, the spectrum utilization information is not directly available for users. These can impose major difficulties for system design. However, cognitive radio, a technique that allows users to dynamically sense the frequency spectrum, find the available spectrum bands in a target spectral range, and then transmit without introducing excessive interference to the existing users in this spectral range, provides a technique to solve this problem. There are numerous advantages of Cognitive Radios like

Improved spectrum sensing: By using cognitive radio networks, it is possible to gain significant advantages in terms of spectrum sensing. Intelligent spectrum management Recently, the traditional approaches for spectrum management have been reconsidered to the actual use of spectrum. FCC's (Federal Communications Commission) Spectrum Policy Task Force has reported plentiful temporal and geographic variations in the usage of allocated spectrum [1]. One way of increasing spectrum utilization is to reuse the spectrum when their primary users are non-active. With the concept of the opportunistic spectrum sharing, secondary users are allowed to access to unlicensed bands without agreement from the primary users. The unlicensed bands play a key role in this wireless system since the deployment of applications in these bands is unencumbered by regulatory delays and which resulted in a rich of new applications. CR is an intelligent wireless communication system [8]that is aware of its surrounding environment, and uses the method of understanding-by-building to learn from the environment and adapts its internal states to statistical variations in the incoming radio frequency stimulate by making corresponding changes of its parameters in time [9]. The radio dynamically identifies portions of the spectrum that are not in use by primary users, and configure the radio to operate in the appropriate White Spaces. Fig. 1a [10] shows the intelligent spectrum access white spaces in tempo al-spectrum domain. The CR-based wireless system is able to utilize the large amount of unused spectrum in an intelligent way while not interfering with other incumbent devices in the frequency bands already licensed for specific uses.

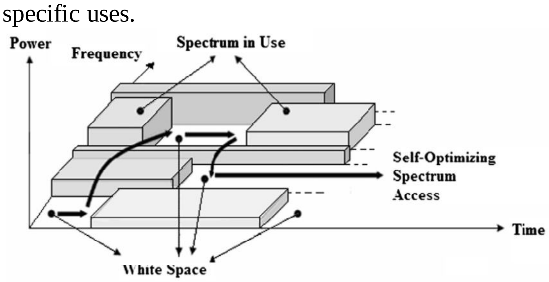


Fig. 1a. Intelligent spectrum access white spaces in temporal-spectrum domain [10].

Autonomic computing

An AC paradigm [5,12,13] has a mechanism whereby changes in its essential variables can trigger changes in the behavior of the computing system such that the system can be brought back into equilibrium with respect to the environment. The vision of AC [14] is to improve the management of complex IT systems by introducing selfmanagement schemes. Multiple interacting autonomic elements (AEs) form the building blocks of the AC system./The AC system monitors data sensed by sensors, analyzes them, and adjusts its operations according to policies, thus reducing complexity [12]. IBM has defined an AC architecture [15] that focuses on enabling self-management functionality. This has shifted the burden of managing systems from people to technologies based on an enterprise view of appropriate policies. The architecture is built based on an intelligent control loop that monitors, analyzes, plans and executes based on a perception of the current environment [16]. Autonomic applications and systems are composed from autonomic elements, and are capable of managing their behaviors and their relationships with other systems/ applications in accordance with high-level policies.

Improved Coverage: By setting up cognitive radio network, it is possible to relay data from one node to the next. In this way power levels can be reduced and performance maintained. Conventionally, to improve spectral efficiency and reliability of multiple access channels (MAC), multiple antennas are deployed at the base stations [4, 5]. Since, single-antenna mobile users are quite common due to the size and cost limitations of mobile terminals, this is termed as single input multiple output multiple access channels (SIMO-MAC).Consider the example of cellular network bands which are overloaded in most parts of the world, but amateur radio and paging frequencies are not. Independent studies performed in some countries confirmed that observation, [7, 8] concluded that spectrum utilization depends strongly on time and place. Since, fixed spectrum allocation prevents rarely used frequencies (assigned to specific services) from being used by unlicensed users, even when their transmissions would not interfere at all with the assigned service. This was the reason for allowing unlicensed users to utilize licensed bands whenever it would not cause any interference. This paradigm for wireless communication is known as cognitive radio .In this paper architecture; functions of cognitive radio are discussed for intelligent networks.

II. ARCHITECTURE OF COGNITIVE RADIO:

The Digital Radio system mainly consists of three functional blocks [3]: RF section, IF section and Baseband section

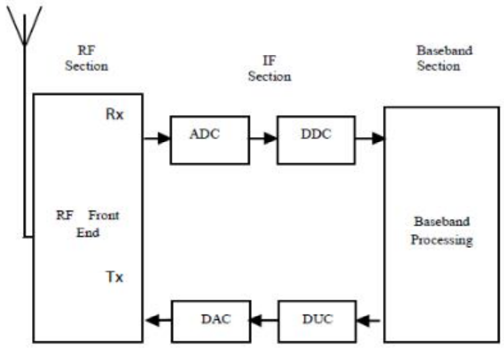


Figure 1: Block Diagram of a generic Digital Transceiver

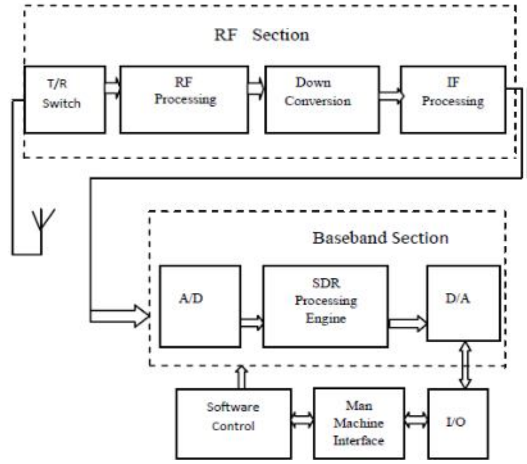


Fig1.b:Conceptual Block Diagram of Software Defined Radio

The architecture of Cognitive radio is more or less related to SDR. The RF section consists of essentially analog hardware modules while IF and baseband section consists of digital hardware modules. The RF section is responsible for transmitting/receiving the radio frequency signal from the antenna and converting the RF signal to an intermediate frequency (IF) signal. On the transmit path, it performs analog up conversion and RF power amplification. The ADC/DAC blocks perform Analog to Digital conversion on the receive path and Digital to Analog conversion on the transmit path, respectively. DUC/DDC blocks perform Digital Down Conversion (receive path) and Digital Up Conversion (transmit path) respectively. Cognitive Radio networks have different requirements compared to ordinary networks. Some physical layer requirements are as under:

1) The multiple services should be supported including real-time voice, data message, still pictures and video. To support multiple services, different constraints on QoS have to be met. 2) The radio needs to be robust to combat bad physical channel conditions. 3) Energy-efficiency is a major concern because the battery life of radio devices can be a limitation for successful operations. 4) The radio should be operational in presence of intentional jamming. All these requirements have to be supported by flexible and reconfigurable radio architecture. In addition to the level of processing required for cognitive radio, the RF sections will need to be particularly flexible. In order to achieve the required level of performance they need a very flexible front end. Traditional front end technology cannot handle all these requirements since they are band limited, for both the form of modulation used and the frequency band in which they operate. But wide band receivers have limitations and generally operate by switching front ends accordingly, the required level of performance can only be achieved by converting to and from the signal as close to the antenna as possible. Hence in this way no analogue signal processing will be needed, and all the processing will be handled by the digital signal processing. The conversion to and from the digital format is handled by digital to analogue converters (DACs) and analogue to digital converters (ADCs). To achieve the performance required for a cognitive radio, not only must the DACs and ADCs have an enormous dynamic range, and be able to operate over a very wide range, extending up to many GHz, but in the case of the transmitter they must be able to handle significant levels of power.

III. FUNCTIONS OF COGNITIVE RADIO

(1) Spectrum Sensing in Cognitive Radio

Cognitive radio spectrum sensing is one of the key algorithms associated with the whole field of cognitive radio.In order to identify the licensed user and locate unused spectrum, the system has to sense the spectrum. For spectrum sensing, there are three signal processing techniques: matched filtering, energy detection and cyclostationary feature detection [8]. Matched filtering is an optimal way used for signal detection in communication systems. However, it requires prior knowledge on the licensed user signal which may not be available. Energy detection is mostly used to determine the presence of signals without prior knowledge. But, there are limitations for the energy detection: [1] the decision threshold is subjected to changing signal to noise ratios [2] It cannot distinguish interference from a user signal [3] it is not effective for signals whose signal power has been spread over wideband. Therefore, the power detection is not adequate for spectrum sensing. Cyclostationary feature detection is therefore used to extract signal features in the background of noise [9].When sensing the spectrum occupancy; the cognitive radio system must accommodate a variety of considerations: Continuous spectrum sensing, Monitor for alternative empty spectrum, Monitor type of transmission.

The ways in which cognitive radio spectrum sensing can be performed falls into one of two categories: Non-cooperative spectrum sensing[7] Cooperative spectrum sensing. There are a number of attributes like Spectrum sensing bandwidth, Transmission type sensing, Spectrum sensing accuracy, Spectrum sensing timing windows etc. which must be incorporated into any cognitive radio spectrum sensing scheme. This ensures that the spectrum sensing is undertaken to meet the requirements of particular applications. The methodology and attributes assigned to the spectrum sensing ensures that the cognitive radio system is able to avoid interference to other users while maintaining its own performance. I. Backoff Interval for Channel Sensing The IEEE 802.11 standard at the physical layer provides a backoff timer that chooses a random number „r', in [0, CW], where CW is contention window. The node sets the timer to the value 20 μs×r and then counts down to zero as shown in Figure 2 after the completion of the time which is given by the Distributed Inter Frame Spacing (DIFS) [10]. Now, during this countdown interval, the channel is still sensed by the radio. If it is found busy at any particular slot, then the backoff counter is frozen for the duration that the channel is found to be occupied. But, the classical approach of tuning individually to all available channels and gathering channel statistics has the following limitations: The short time duration for which the timer is frozen poses an upper limit on the number of channels that may be sensed in succession. Also estimating the channel occupancy by received signal power alone may not provide accurate results.

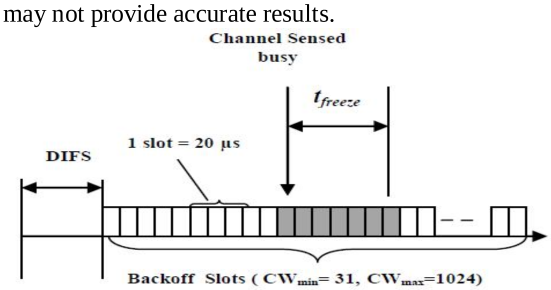


Figure 2: Packet transfer for the radio during the freeze duration

(2) Spectrum Management [10]

The best available spectrum is captured in order to meet user communication requirements. Cognitive radios therefore should decide on the best spectrum band to meet the Quality of service (QOS) requirements over all available spectrum bands, hence spectrum management functions are required for Cognitive radios which are classified as: Spectrum analysis and Spectrum decision.

(3) Spectrum Mobility [11]

It is the process when a cognitive radio user exchanges its frequency of operation and Cognitive radio networks target to use the spectrum in a dynamic manner by allowing the radio terminals to operate in the best available frequency band, thus maintaining seamless communication requirements during the transition to achieve the best spectrum.

(4) Spectrum Sharing

The major challenge in open spectrum usage is the spectrum sharing. It is similar to generic media access control MAC problems in the existing systems.CR can improve spectrum Utilization by (i)Allocating the frequency usage in a network (ii) Assist secondary markets with frequency use, implemented by mutual agreements (iii)Negotiate frequency use between users(iv) Provide automated frequency coordination (v) Enable unlicensed users when spectrum not in use. CR can also improve network management efficiency by (i) time-space characterization of demand is possible, (ii)Cognitive Radio Learns plans of the user to move and use wireless resources, Expresses its plans to the network reducing uncertainty about future demand,(iii)The network can use its resources more efficiently.

IV. CONCLUSION

With Wireless and Radio communications becoming far more widely used, and the current levels of growth looking to increase, ideas such as cognitive radio will become more important. Since, the drawbacks of SDR include Power consumption, Security, Cost, keeping up with higher data rates, also both subscriber and base units should be SDR for maximum benefit. Hence, Cognitive Radio is introduced for Intelligent wireless networks. Applications of Cognitive Radio (CR) allow user terminals to sense whether a portion of the spectrum is being used or not, in order to share the spectrum among neighbor users. It determines the occupancy of the available spectrum, and then decide the best power level, mode of transmission and other necessary characteristics. Additionally the radio also judges the level of interference it may cause to other users. which is important requirement for the radio communications system to operate effectively. Planning applications are one of the new attractive technologies that CR will enable. Navigation using databases and GPS signals will increase. The CR can remember prior paths and can learn better routes. Some of the planning applications envisioned include: Route planning, Battery Management, Noise Discipline, Management of Information Flow ,Role Assignment as a function of the operator□s skills, Smart Calibration Smart Bridge. Thus, Cognitive radio will be an enabling technology in future generations of LTE.

REFERENCES

[1] FCC Spectrum Policy Task Force, Report of the Spectrum Efficiency Working Group, Technical Report 02-135, November 2002. Available from: http://www.fcc.gov/sptf/files/SEWGFinalReport1.pdf>.

[2] Qiwei Zhang, Andre B.J. Kokkeler and Gerard J.M. Smit," A Reconfigurable Radio Architecture for Cognitive Radio in Emergency Networks" European Conference on Wireless Technology, Manchester, September 2006.

[3] Xin Liu, Optimal opportunistic scheduling in wireless networks, in: Proceedings of IEEE 58th Vehicular Technology Conference, vol. 3, 2003, pp. 1417–1421.

[4] M.-R. Garey, D.-S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W.H. Freeman, p. 340.

[5] S. Hariri, M. Parashar, Autonomic Computing: An overview, in: Lecture Notes in Computer Science, Springer, 2005, pp. 247–259.

[6] Won-Yeol Lee, I.F. Akyildiz, Optimal spectrum sensing framework for cognitive radio networks, IEEE Transactions on Wireless Communications 7 (10) (2008) 3845–3857.

[7] ERC/DEC/(99)23, ERC Decision on the harmonized frequency bands to be designed for the introduction of HIPERLANS, 29-11-1999. Available from: <http://www.ero.dk/doc98/official/pdf/ DEC9923E.PDF>

[8] . Y. Yuan, P. Bahl, R. Chandra, T. Moscibroda, Y. Wu, Allocating dynamic time-spectrum blocks in cognitive radio networks, in Proceedings of the 8th ACM International Symposium on Mobile a hoc Networking and Computing, 2007, pp. 130–139.

[9] S. Haykin, Cognitive radio: brain-empowered wireless communications, IEEE Journal on Selected Areas in Communications 23 (2) (2005) 201–220.

[10] I.F. Akyildiz, W.Y. Lee, M.C. Vuran, S. Mohanty, Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey, in: Computer Networks Journal, Elsevier, 2006.

[11] Q. Zhao, B.M. Sadler, A survey of dynamic spectrum access: signal processing, networking, and regulatory policy, IEEE Signal Processing Magazine (2007).

[12] M.R. Nami, M. Sharifi, Autonomic computing: a new approach, in: First Asia International Conference on Modelling and Simulation, AMS'07, 2007, pp. 352–357.

[13] M.R. Nami, K. Bertels, A survey of autonomic computing systems, in: Third International Conference on Autonomic and Autonomous Systems (ICAS07), 2007, pp. 26–26.

[14] J.O. Kephart, D.M. Chess, The vision of autonomic computing, IEEE Computer 36 (2003) 41–50.

[15] IBM, An Architectural Blueprint for Autonomic Computing, April 2003. Available from: <http://www03.ibm.com/autonomic/pdfs/AC%20Blueprint %20White%20Paper%20V7.pdf>

[16] S. Davy, K. Barrett, S. Balasubramaniam, S. Meer, B. Jennings, J. Strassner, Policy-based architecture to enable autonomic communications – a position paper, in: 3rd IEEE Consumer Communications and Networking Conference, CCNC, vol. 1, 2006,pp. 590–594.

Dispersion Compensation Techniques in Optical Fibres

Gourav Soni[1], Sandeep kaushal [2] and Narinder Sharma [3]
[1, 2, 3]Associate Professor, Department of Electronics & Communication Engineering, ACET, Amritsar, India

Abstract: **Dispersion in a single mode fiber is the bottleneck of long haul optical communication systems, which limits the bit rate and repeater-less distance. Chromatic dispersion (CD) of a single mode fiber (SMF) is an important aspect in a long-haul optical communication system. Chromatic dispersion leads to inter-symbol interference and degrades the quality of the signal. In this paper we review different dispersion compensation techniques.**

Keywords: Single mode fiber, Chromatic dispersion, all pass filter, Fiber bragg grating

I. INTRODUCTION

With the rapid development of wireless communications [1,2], the bandwidth requirement of the radio signal increases rapidly in order to realize multigigabit/s broadband wireless access, and RoF techniques allowing for the integration of optical and wireless systems, have become attractive solutions for increasing the capacity, bandwidth and mobility to serve both fixed and mobile users [3–5]. As a consequence, many research works related to RoF system have been intensively conducted, such as mm-wave generation and distribution, RoF system integrated with DWDM, and suppression of nonlinear distortion [6–12]. In order to simplify the implementation and reduce the cost of the RoF system, mm-wave signals are generated in a central location directly in the optical domain, and distributed over optical fiber to remote antenna locations. Many different approaches have been reported to generate optical mm-wave, such as direct modulation [9], optical heterodyne techniques [10,11], external opticalmodulation [12,13.14], and methods based on optical nonlinear effects [16-19]. Methods based on external optical modulation have shown great potential for high frequency mm-wave generation because of its wide bandwidth and simple configuration.In RoF systems, data signal is modulated on the optical mmwave signal[13,15] for transmission. In order to avoid the time shift of code edges caused by fiber chromatic dispersion, data signals have to be modulated on only one sideband of the generated optical mm-wave signal. If the methods mentioned above are used, other optical components such as optical interleaver, fiber Brag grating (FBG), optical circulator, optical intensity modulator and optical coupler (OC) have to be used to modulate the data signal onto one of the optical sidebands, which will make the system very complicated and lead to big insertion loss. Optical fiber communication is a way of transmitting the information from one place to another by modulating the light signal with the information signal. Optical fiber communication systems primarily operate at wavelengths near 1.55 µm in order to coincide with the minimum loss point of optical fiber and thereby maximizing the transmission distance.[1] But, at this wavelength, there is a significant amount of group velocity dispersion (GVD) that limits the achievable propagation distance. At the operating wavelength window of 1.55µm, the dispersion in a single mode fiber(SMF) is known as chromatic dispersion(CD) which results in pulse spreading and causes intersymbol interference(ISI)[2].CD is made of material dispersion and waveguide dispersion.The phenomenon of different wavelengths travelling at different speed due to variation of refractive index of the SMF is known as material dispersion.A proportion of the light will also travel in the cladding of the SMF ,which has a different refractive index compared to the core, hence introduces an effect known as waveguide dispersion.[3]

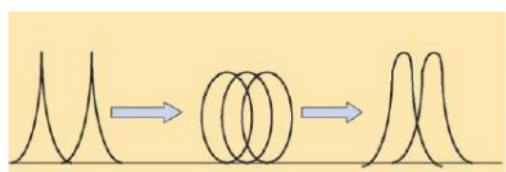


Fig1.Chromatic dispersion effect on optical pulse

The length of the SMF in a high bit rate long-haul optical communication system is limited by the bit rate and the CD [2,4], which can be described by the bit rate length product

$$B^2L=c/(4D\lambda_0^2)$$

where B is the bit rate, L is the SMF length,c the velocity of light, D is the dispersion coefficient and $\lambda 0$ the operating wavelength. Therefore doubling the bit rate should reduce the fiber length by a factor of four for a fixed dispersion factor. Therefore in order to realize high bit rate transmission over long distances using the SMF, CD[22] compensation techniques must be employed to overcome signal distortion resulting from dispersion.[5,6]

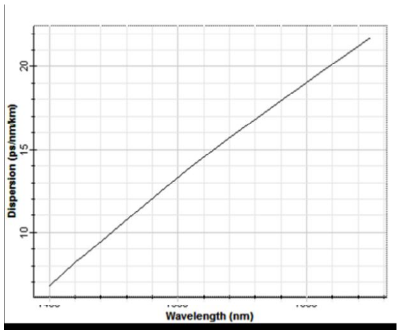


Fig.2 Variation of chromatic dispersion with wavelength

II.CDCOMPENSATION: TECHNIQUES [22-23]

A number of compensating techniques have been reported in the literature including dispersion compensating fibers (DCFs), Fiber Bragg gratings (FBGs), Electronic Dispersion compensation (EDC) each having its own advantages and disadvantages.

1. Dispersion compensating fibers (DCF): DCF is a specially designed fiber with negative dispersion, employed to compensate for positive dispersion over large lengths of SMF. DCF typically has a much narrower core than a SMF, causing the optical signal to be more tightly confined and accentuating the problem caused by nonlinear high power effects which results in higher attenuation compared to the SMF .DCF is a loop of fiber which can be inserted at either beginning or end between two optical amplifiers accordingly it is referred to as pre compensation or post-compensation techniques. A third option is to have a DCF at both ends. It is a passive compensation in which loop of fiber having dispersion characteristics that negates the accumulated dispersion of the transmission fiber is inserted at or end of the transmission fiber.

2. Fiber bragg grating: Chirp FBG can compensate CD of a SMF by using the varying distance of grating to delay the faster wavelengths in relation to the slower wavelengths of an optical pulse. By recombining all the wavelengths of an optical pulse at the receiving end, the original optical pulse can be restored. The chirp FBG is limited by its narrow bandwidth and ripple in the opposite GVD. Tunable dispersion is typically achieved by adjusting the chirp on the FBG but causes a large shift in the centre wavelength which is a critical problem.

3. Electronic dispersion compensation (EDC): Since there is direct detection at the receiver, linear distortions in the optical domain, e.g. chromatic dispersion, are translated into no distortions after optical-to-electrical conversion .Due to this reason that the concept of nonlinear cancellation and nonlinear channel modeling attracted increasing attention Electronic compensation circuits can improve system performance at a small cost premium once they have entered large volumes. The simplified equalization circuits feed forward equalizer (FFE) and decision feedback equalizers (DFE). All pass filters are linear systems having variable phase response and constant amplitude response. The variable phase response of the APFs makes them to be used as the phase equalizers to compensate the chromatic dispersion.

III. COMPARISON OF COMPENSATION TECHNIQUES

Dispersion compensating fiber: DCF is the predominant technology for dispersion compensation. It consists of an optical fiber that has a special design such as providing a large negative dispersion coefficient while the dispersion of the transport fiber is positive. A proper length of DCF allows the compensation of the chromatic dispersion accumulated over a given length of the transport fiber, although standard modules with predetermined dispersion values are commercially available. Fiber Bragg grating technique: One of the most advanced technologies being incorporated into dispersion compensation methods is FBGs, short lengths of optical fiber that reflect a particular wavelength. FBG feature periodically spaced zones in the fiber core that have been altered to have different refractive indexes slightly higher than the core. This structure selectively reflects a very narrow range of wavelengths while transmitting others. Optical phase conjugation: In this technique we use a device called optical phase conjugator (OPC) in the middle of the link to invert the spectrum. This process changes the short wavelengths to long ones and the long wavelengths to short ones. If we invert the spectrum in the middle of a link (using standard fiber) the second half of the link acts in the opposite direction (really the same direction but the input has been exactly pre-emphasized). *Negative dispersion fiber:* NDF has been used as the transmission medium in digital lightwave transport systems. Some work was done to compensate dispersion using negative dispersion fiber for regional metro network and at bit rates of 10 Gbps . NDF, which has negative dispersion, was found to compensate positive laser chirp and improve the dispersion tolerance in a directly modulated CATV transport system Reverse dispersion fiber: RDF has several advantages compared with the conventional DCF, including lower loss, lower non-linearity and lower polarization mode dispersion (PMD). RDF is designed for managing total dispersion with combination of SMF. It shows negative dispersion and slope, reverse to those of SMF. It has been reported that a desired zero dispersion can be constructed by using SMF and RDF combination in digital lightwave transport systems.

CONCLUSION

Chromatic dispersion (CD) is the most important impairment which can limit the transmission reach to less than 100km for 10Gbit/s OOK signal and 6km at 40Gbit/s. The commercial optical networks employ inline dispersion compensation fibre (DCF), which is bulky and expensive with significant power attenuation. It's length has to be manually adjusted to achieve proper CD compensation with the result that link provisioning is expensive and time consumption. Electronic dispersion compensation (EDC) has attracted much interest recently for extending reach in legacy multimode optical fiber as well as in metro and long-haul optical transmission systems. Its advantages compared to the optical compensation method include: Reduced costs by eliminating the need for DCF modules including the cost of DCFs and the associated cost for compensating the loss from the DCFs. Simplification of the deployment and configuration Flexible and adaptive compensation required in dynamic optical networks. Easy for integration in transmitter and receiver.

REFERENCES

[1] D. Wake, A. Nkansah, N.J. Gomes, Radio over fiber link design for next generation wireless systems, J. Lightwave Technol. 28 (2010) 456–2464.

[2] J. Ma, X. Xin, J. Yu, C. Yu, K. Wang, H. Huang, L. Rao, Optical millimeter wave generated by octupling the frequency of the local oscillator, J. Opt. Network. 7 (2008) 837–845.

[3] J. Zhao, Z. Liu, X. Liu, T. Shang, P. Yue, Optimisation of carrier-to-sideband ratio by triple-arm Mach–Zehnder modulators in radio-over-fibre links, IET Optoelectron. 4 (2010) 183–188.

[4] S. Pan, J. Yao, A frequency-doubling optoelectronic oscillator using a polarization modulator, IEEE Photon. Technol. Lett. 21 (2009) 929–931.

[5] J. Zhang, H. Chen, M. Chen, T. Wang, S. Xie, A photonic microwave frequency quadrupler using two cascaded intensity modulators with repetitious optical carrier suppression, IEEE Photon. Technol. Lett. 19 (2007) 1057–1059.

[6] J. Zhang, H. Chen, M. Chen, T. Wang, S. Xie, Photonic generation of a millimeter-wave signal based on sextuple-frequency multiplication, Opt. Lett. 32 (2007) 1020–1022.

[7] L. Chen, Y. Shao, X. Lei, H. Wen, S. Wen, A novel radio-over-fiber system with wavelength reuse for upstream data connection, IEEE Photon. Technol. Lett. 19 (2007) 387–389.

[8] N.J. Gomes, M. Morant, A. Alphones, B. Cabon, J.E. Mitchell, C. Lethien, M. Csörnyei, A. Stöhr, S. Iezekiel, Radio-over-fiber transport for the support of wireless broadband services [Invited], J. Opt. Network. 8 (2009) 156–178.

[9] J. Yu, M. Huang, Z. Jia, T. Wang, G. Chang, A novel scheme to generate singlesideband millimeter-wave signals by using low-frequency local oscillator signal, IEEE Photon. Technol. Lett. 20 (2008) 478–480.

[10] **J**. Ma, J. Yu, C. Yu, Z. Jia, G. Chang, X. Sang, Optical mm-wave generation by using external modulator based on optical carrier suppression, Opt. Commun. 268 (2006) 51–57.

[11] T. Schneider, D. Hannover, M. Junker, Investigation of Brillouin scattering in optical fibers for the generation of millimeter waves, J. Lightwave Technol. 24 (2006) 295–304.

[12] M. Sauer, A. Kobuakov, J. George, Radio over fiber for picocellular network architectures, J. Lightwave Technol. 25 (2006) 3301–3320.

[13] J. Yu, Z. Jia, L. Yi, Y. Su, G. Chang, Optical millimeter-wave generation or upconversion using external modulators, IEEE Photon. Technol. Lett. 18 (2006) 265–267.

[14] J. Yu, Z. Jia, L. Xu, L. Chen, T. Wang, G. Chang, DWDM optical millimeter-wave generation for radio-over-fiber using an optical phase modulator and an optical interleaver, IEEE Photon. Technol. Lett. 18 (2006) 1418–1420.

[15] T. Cho, K. Kim, Effect of third-order intermodulation on radio-over-fiber systems by a dual-electrode Mach–Zehnder modulator with ODSB and OSSB signals, J. Lightwave Technol. 24 (2006) 2052–2058 .

[16] Z. Jia, J. Yu, G. Chang, A full-duplex radio-over-fiber system based on optical carrier suppression and reuse, IEEE Photon. Technol. Lett. 18 (2006) 1726– 1728.

Edge Detection: An Essential Tool for Image Processing

Jaspreet Kaur
Department of Electronics and Communication Engineering,
Amritsar College Of Engineering and Technology, Amritsar (Punjab)
Email: rosy.jaspreet@gmail.com

***Abstract* — Edge detection is one of the most frequently used techniques in digital image processing. Its application area reaches from astronomy to medicine where isolation of objects focused on from the unwanted background is of great interest. It constitutes a significant portion of the information contained in images. Edge detecting an image significantly reduces the amount of data and filters out useless information, while preserving the important structural properties in an image. Edge detection is one of the subjects of basic importance in image processing. In this paper, the main focus is to highlight the basic importance of edge detection in image processing**

Index Terms - Edge detection, images, information

I. INTRODUCTION

Edges in a digital image provide important information about the objects contained within the image since they constitute the boundaries between the objects in the image. Edge detection is a frequently performed operation in many image processing applications because it is usually the first operation that is performed before other image processing tasks such as image segmentation, boundary detection, object recognition and classification, image registration, and so on. Consequently, the success of these subsequent image processing tasks is strictly dependent on the performance of the edge detection operation. Color images provide more information than gray scale images. There are numerous techniques available to detect meaningful edges based on the approaches such as genetic algorithms, fuzzy logic, and neural network, Bayesian approach, wavelet domain, and mathematical morphology and so on. Edges are often associated with the boundaries of objects in a scene. Traditional approaches using classical edge detectors fail when the images are noisy. In the case of noisy images, it may produce false edges caused by discontinuities in gray level due to noise. At present, ultrasound image system is a tremendously valuable technique due to its safety, cost-effectiveness and real-time imaging capability. It poses unique challenges for standard edge detection algorithms, because the boundaries between regions of interest (ROI) in ultrasound images are generally bright steaks between similar intensity regions, rather than differing contrast. Furthermore, the speckle noise that presented in the ultrasound images is mainly reason to make ultrasonic images degenerate and increase the difficulty in discriminating fine detail in images during diagnosis examination. Therefore, edge detection in ultrasound images is very difficult task because edge detection algorithms may be sensitive to noise. This problem had led researchers to develop different optimal methods for edge detection in ultrasound images In this case it must attempt to find edges in smoothed images instead of the original ones to reduce the effect of noise. First, the algorithm is firstly applied to the ultrasonic images in order to convert the multiplicative noise into the additive noise case. The wavelet based denoising is used to remove the speckle noise in ultrasonic images. In this case, soft threshold is adopted. Then, the inverse wavelet transform is computed and applied the inverse logarithm. Second, when an image is examined for intensity variations, several scales generally are of interest. The detection of certain feature in an image is optimal at a certain scale. Therefore, it is necessary to perform and combine information of edge detection at multiple scales. Since the wavelet transform (WT) is obviously related to multi-scale analysis, the WT could play an important role in multi-scale edge detection. In general, edge detection is one of the most commonly used operations in image analysis. The reason for this is that edges form the outline of an object. An edge is the boundary between an object and the background, and indicates the boundary between overlapping objects. This means that if the edges in an image can be identified accurately, all of the objects can be located and basic properties such as area, perimeter, and shape can be measured. Since computer vision involves the identification and classification of objects in an image, edge detections is an essential tool of it.

II. MEANING OF EDGE DETECTION

Edge detection is one of the most critical and hot topic for digital images for segmenting images and to improve the quality of the images. As we know, about data abstraction, i.e. it focuses on some of its data but eliminates unwanted data. In the same way, Edge Detection is used to trim down and strain some amount of data and inadequate information, at the same time preserving the important structural (edges) properties in an image. Edge detection is used to capture the discontinuities in the image brightness where discontinuities can be either at discontinuities in depth, discontinuities in surface orientation or changes in material properties etc.

III. IMPORTANCE OF EDGE DETECTION IN IMAGE PROCESSING

Edge detection is a very important area in the field of image processing. Edges define the boundaries between regions in an image, which helps with segmentation and object recognition. They can show where shadows fall in an image or any other distinct change in the intensity of an image. Edge detection is a fundamental of low-level image processing and good edges are necessary for higher level processing. Edge detection is one of the most commonly used operations in image analysis. An edge is defined by a discontinuity in gray level values. In other words, an edge is the boundary between an object and the background. The shape of edges in images depends on many parameters: The geometrical and optical properties of the object, the illumination conditions, and the noise level in the images. The main goal of edge detection is to mark the points in an image at which the intensity changes sharply. Sharp changes in image properties usually reflect important events and changes in world properties.

IV. CHALLENGES FOR EDGE DETECTION

Many of the basic image processing tasks, such as edge detection, color conversion, histogram operations, threshold, correlation, interpolation for data extraction, etc., act on the image at a pixel level to extract useful high level information. These basic tasks are referred to as pixel vision in this report. One of the challenges for performing these tasks is the time taken by the processor to operate over the whole image. Another challenge is to handle the complexity of the input image. In order to overcome these two limiting factors and to strike a balance between them, custom made platforms are being developed. Even these customs made platforms suffer from some limitations, when dealing with real time deployment. The issue with real time deployment is that the system should be compatible with different algorithms depending on different conditions, like time of the day, weather conditions, etc. This calls for a flexible platform, which is not the case with these custom made platforms.

V. FINDINGS OF THE STUDY

- Edge detection algorithms for an image may reduce the quantity of data to be processed and filters out the information that may be less relevant but at the same time preserving the structural properties of the image.

- The quality of edge detection is highly dependent on lighting conditions, the presence of objects of similar intensities, density of edges in the scene, and noise.

- In the ideal case, the result of applying an edge detector to an image may lead to a set of connected curves that indicate the boundaries of objects, the boundaries of surface markings as well as curves that correspond to discontinuities in surface orientation.

VI. SUGGESTIONS

Since different edge detectors work better under different conditions, it would be ideal to have an algorithm that makes use of multiple edge detectors, applying each one when the scene conditions are most ideal for its method of detection. In order to create this system, we must first know which edge detectors perform better under which conditions that use different methods for detecting edges and then compared their results under a variety of situations to determine which detector was preferable under different sets of conditions. This data could then be used to create a multi-edge-detector system, which analyses the scene and runs the edge detector best suited for the current set of data. For one of the edge detectors we have to consider different ways of implementation. Rather than trying to find the ideal edge detector to apply to traditional photographs, it would be more efficient to merely change the method of photography to one which is more conducive to edge detection.

It makes use of a camera that takes multiple images in rapid succession under different lighting conditions. Since the hardware for this sort of edge detection is different than that used with the other edge detectors, it would not be included in the multiple edge detector system but can be considered as a viable alternative to this. As edge detection is a fundamental step in computer vision, it is necessary to point out the true edges to get the best results from the matching process. That is why it is important to choose edge detectors that fit best to the application.

VII. CONCLUSION

Edge detection is a fundamental tool in image processing, machine vision and computer vision, particularly in the areas of feature detection and feature extraction, which aim at identifying points in a digital image at which the image brightness changes sharply or, more formally, has discontinuities. Now days, edge detection is one of the most important technique for digital images for segmenting and to improve the quality of the images.

ACKNOWLEDGMENT

I hereby acknowledge the contribution of information providers to give us proper information and enable us to write this paper. I would also like to thank the eminent personalities who spared their precious time to respond to our questionnaires, personal interviews and telephonic interviews. Without the contribution of these people it was not only difficult but even impossible to write an empirical research paper. I once again thank all and one who contributed in our research work by providing us worthy information. Moreover I am highly thankful to an eminent learned personality "Dr. Vijay Kumar Banga" –Head, Department of Electronics and Communication Engineering, Amritsar College of Engineering and Technology, Amritsar, Punjab, India, for his cordial support and motivating us for doing this creative research work.

REFERENCES

[1] D. Ziou and S. Tabbone (1998) "Edge detection techniques: An overview", International Journal of

[2] Pattern Recognition and Image Analysis, 8(4):537–559, 1998

[3] J. M. Park and Y. Lu (2008) "Edge detection in grayscale, color, and range images", in B. W. Wah (editor) Encyclopedia of Computer Science and Engineering

[4] J. Canny (1986) "A computational approach to edge detection", IEEE Trans. Pattern Analysis and Machine Intelligence, vol 8, pages 679-714.

[5] Shapiro L.G. & Stockman G.C. (2001) Computer Vision. London etc.: Prentice Hall, Page 326.

[6] Evelyn Brannock, Michael Wees,"A Synopsis of Recent Work in Edge Detection using the DWT", IEEE, 2008, pages 515 -520.

[7] R, Raskar; Tan, K-H; Feris, R.; Yu, J.; Turk, M., "Non-photorealistic Camera: Depth Edge Detection and Stylized Rendering Using Multi-Flash Imaging", ACM SIGGRAPH, August 2004

[8] Jean Ponce, "Lecture 26: Edge Detection II", 12/2/2004. http://www- cvr.ai.uiuc.edu/~ponce/fall04/lect26.ppt
Received: December 31, 2007

[9] Marimont, D., Rubner Y., "A Probabilistic Frameword for Edge Detection and Scale Selection",Proceedings of the IEEE International Conference on Computer Vision, 207-214

[10]. Digital image processing by (R.C.Gonzalez and R.E.Woods)

Image Processing in Bionic Eye

Rajbir Kaur
M.Tech ECE student, Amritsar Collage Of Engineering And Technology
rajbirkaurr@yahoo.com

Abstract: **Technology has created many path ways for the mankind. The entire human body can be controlled using a single electronic chip. Now it is the turn of artificial vision through bionic eyes. Chips designed specially to imitate the characteristics of the damaged retina and the cones and rods of the organ of sight are implanted with a microsurgery.**

Keywords-Bionic, AMD, RP ,Visual Prosthesis

I. Introduction

Bionic Eye also called as BIO ELECTRONIC EYE is a electronic device that replaces the functionality of part or whole of eye.A bionic eye work by stimulating nerves which are activated by electrical pulses.In this a small device is implanted with patient which receive signals and send them to the nerves.

An external camera is worn on a pair of dark glasses which sends the images in digital form to the radio receiver placed in the eye. The radio receiver is attached to the implant chip on the retina. The implantation is of two types, epiretinal implant and sub retinal implant, based on whether the implant is placed on or behind the retina.

II. Basic Eye Disorders

The Eye disorders dealt here are listed below:

A • Retinitis Pigmentosa

B • Macular Degeneration Maintaining the Integrity of the Specifications

A. Retinitis Pigmentosa

Retinitis Pigmentosa is a name given to to a group of hereditary diseases of the retina of the eye. RP is a progressive blinding disorder of the outer retina which involves degeneration of neurons [1]. RP may be caused by a breakdown in the function of the rods or the cones in some part of the retina. The retina is so complex that, breakdowns may occur in a variety of ways and so RP is not a single disorder but a great number of disorders. The breakdown of cone function may be called Macular Degeneration

B Macular Degeneration

Macular Degeneration is a medical condition which usually affects older adults. Macular Degeneration is mainly due to the breakdown of the cones in the retina. The cone cells are responsible for distinguishing the colours of images formed on the retina. In macular degeneration, a layer beneath the retina, called the retinal pigment epithelium (RPE), gradually wears out from its lifelong duties of disposing of retinal waste products. A large proportion of macular degeneration cases are age- related and it can make it difficult to read or recognize faces, although enough peripheral vision remains to allow other activities of daily life. Age related Macular Degeneration (AMD) usually affects people over the age of 50 and there are two distinct types - wet AMD and dry AMD. Wet AMD results from the growth of new blood vessels in the choroids, causing an accumulation of fluid in the macula which leads to retinal damage. Dry AMD represents at least 80% of all AMD cases and results in atrophy of the Retina. Usually yellowish-white round spots called drusen first appear in a scattered pattern.

III. Bionic Eye:

A visual prosthesis or bionic eye is a form of neural prosthesis intended to partially restore or amplify existing vision. It usually takes the form of an externallyworn camera that is attached to a stimulator on the retina, optic nerve, or in the visual cortex, in order to produce perceptions in virtual cortex.

The external camera is on a pair of dark glasses which sends images in digital form to reciever placed in eye.

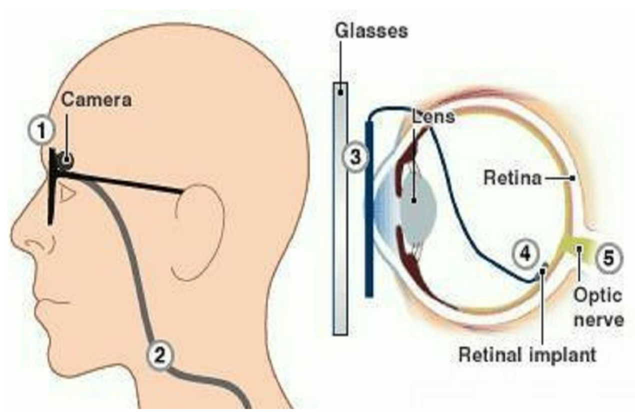


FIG 1: BIONIC EYE

IV Image processing In Bionic Eye

In this some image processing techniques are discussed related to artificial vision .research is described in 4 areas :

- Vision chip development

- CCD based systems
- Receptive field modeling
- Multiple resolution work

Vision Chip Development

In this technique depth information in bionic eye ies recognized .In this the vision chip has depth perception [2].The chip has 100 analogue sensors connected laterally by resistors giving 1 dimensional 100 pixel sensor which allows parallel processing in real time.. The output of the system is serial signal representing depth.. As real time processing has sepration for image sensing (camera) and image processing (computer) so system performance is limited by slow camera rate and low transmission rate between camera and computer.

`Artificial Retina ` chips are used that can simultaneously sense and process images. These artificial retinas consist of two dimensional variable sensitivity variable photodetection cell array, with sensitivity similar to commercially available CCDs. A variety of on chip image processing can be achievied by changing the control voltage pattern on chip. These processing functions include image sensing, edge extraction, image smoothing, pattern recognition.

CCD Based Systems

It is a image resolution reduction algorithm based on image segmentation. The processing hardware for prosthesis consists of FPGA/EPLD(Field Programmable Gate Array)/(Electrically Programmable Logic Device)[3].It consist of three SRAM memories that act as registers to store images delivered by camera. Two SRAMS support dual buffer video ,where current image is stored ,while previous image is processed simultaneously .A third frame buffer is used for intermediate computations that may occur in algorithms such as spstial convolution. A 8-bit pipeline A/D convertor supports camera which provide only analogue video. This whole board can be worn in shirt pocket or clipped to belt.

Optic nerve simulation researchers developed a resolution reduction algorithm based on image segmentation by growth of zone and implementation in low power VLSI device[4].the developed a algorithm based on extraction on main features of image ,with transmitted information being only the position and form of the relevant entities of the scene. However ,the current implementation appear only to be based on intensity. They propose to give a blind person ability to control the segmentation level producing areas of uniform illuminance matching corresponding objects or surfaces. Due to nature of this segmentation algorithms ,there is a undesirable fast transition when segmenting with successive images.

Receptive Field Modeling

It is used for eradication of sensory noises from images. It proposed a system that approximates receptive fields properties of primate retinal ganglion cells [4].Input data is fed into two distinct path ways one for center computations and one for surround .Each path performs a spatial scalar product of the pixel data ,and two dimensional gaussian, whose width determines the spatial extent of receptive field. The resulting signal is then processed by temporal low pass filter . The surround pathway signal can be optionally delayed ,and then signals from both pathways converge at mixer point. Finally the gain factor enables range adaptation and switching between ON OFF and OFF ON behaviour .The resulting signal is then used to stimulate nerve cells[5].

The hardware of receptive field modeling consists of CMOS image sensor chip with high dynamic range with respect to illumination intensity (140 db).This full dynamic range can be used within a single image frame without any distortions like blooming ,smearing or time lag .Signal processing is carried out on chip ,so an additional frame buffer is not required unlike CCD devices .The spatial filter used for ON OFF is inserted between the sensor chip and signal processor. This sensor chip in a package can be mounted on spectacle frame .

Multiple Resolution Work

In this method two CCD cameras are used and its implementation is done on VLSI chip [6].

One tele-lens camera produces high resolution in central area of the image ,while the second wide angle camera captures peripheral image area. This system processes these two images in real time to obtain a resulting image with high resolution t the center ,similar to the central part of the retina.

The image processing in this system is based on identification of main features of primary visual system lateral inhibition and graded resolution .Lateral inhibition is implemented by an edge detection filter and graded resolution is modelled using a multi resolution artifical retina based on the filtered image .

This method is also used in head mounted display unit coupled with an eye tracking system[8].In this the author claims that conventional HMDs suffers from a narrow field of view and low resolution and consequently cannot be used for applications such as tele-microsurgery. Their HMDs display high resolution at subjects view point(obtained by an eye tracker) and low resolution at the periphery ,therefore displaying images at higher precieved resolution in a wider view angle. The multi resolution approach is used in applications where bandwidth is limited. .It is most likely that the pixel density in a device would be fixed at maximum possible under manufacturing and size constraints. Improved scene understanding is expected when the entire electrode layout is used rather than applying low resolution image section to some parts of the plant.

This method is also used in head mounted display units than applying low resolution image section to some parts of the plant.

CONCLUSION

This paper discusses about bionic eyes are like to provide visual sensation to blind patients .Various image processing techniques described above aim at improving perception of users of bionic eye. Some strategies for image depth perception, resolution reduction, removal of sensory noises have been presented. The image processing methods presented in this paper are results of various research work carried. The algorithm described in this paper were employed on static images ,and in future this image processing techniques could be applied to image sequences/videos.

References

[1] Asher A , Segal,W.A, Baccus,S.A.;Yaroslavsky,L P ,P alankar,,D.V ,``Image processing for a high resolution optoelectronic retinal prothesis ````,IEEE transection on biomedical engineering ,vol 54,no.6,pp.993-1004,june 2007.

[2] A,``Computational image quality matrices :A review: ``SID Digest of technical papers , 24,pp.305-308,1993.

[3] DE Marco S, Clements M ,Vichienchom K, Liu W,Humayan M ,Weiland J ,An epi-retinal visual prothesis implementation,proceedings of First Joint BMES/EMBS Conference 1999:IEEE Engineering in Medicine and Biology 21st Annual conference and the 1999Annual Falll Meeting Of The Biomedical Engineering Society ,IEEE ,Piscataway, USA, Vol 1,pp475,1999.

[4] Gilmont T, Verians X, Legat J, Veraart C, Resolution reduction by growthof zones for visual prothesis, proceedings :International Conference On Image Processing ,IEEE, New York,USA,pp.299-9301,Vol 1,1996.

[5] Eckmiller R,Becker M, Hunnermann R, ``Dialog Concepts For Learning Retina Encoders" Proceedings: ,International Conference On Neural Networks, IEEE, NEW YORK , USA,pp.2315-2320,1997.

[6] Beckner M ,Braun M, Eckmiller R,`` Retina Implant adjustment reinenforcement learning" ,Proceedings of the 1997 IEEE International Conference On Acoustics, ``SpeechandsignalProcessing",ICASSP'98,IEEE,NewYork,USA,Vol 2,pp.1181-4, 1998.

[7] Barten P,Contrast Sensitivity of the Human Eye and its effects on image quality ,SPIE Press,Washington,1999.

[8] Capelle C ,Faik C, Trullemans C, Veraart C, ``Real time experimental Visual Prothesisusing sensory substitution of vision bu audition", Proceedings of the 16th Annual International conference of the IEEE engineering in Medicine and Biology Society,IEEE, New York, USA; Vol 1,pp.255-256,1994.

[9] ITU/R Recommendations BT.500-7,10/1995, http://www.itu.ch/.

[10] Praveen Kumar ,Guhan Senthil,``International Conference on Life Sciences and Technology" ,Vol 3,pp 91 -93,2011.

[11] N Sharmili, P.S Ramaiah, G Swamyngdhan ,``International Journal Of Computer Science & Engineering Survey" ,Vol 2.May 2011.

[12] Justine Robert Boyle,``Image and Video research Laboratory ,School Of Engineering Systems, ``Queensland University Of Technology " ,Feb 2005.

Implementation of Pattern Recognition System using NI Vision

Parvinder Singh Sandhu[1] and Gurbhej Singh[2]

[1]Assistant Professor, Department of Electronics & Communication Engineering, RIE&IT, Railmajra, India
[2]Assistant Professor, Department of Electronics & Communication Engineering, BKSJCET, Amritsar, India
Email: Er.Parvinder@gmail.com, Gurbhej86@gmail.comr

Abstract:-The widespread interest in personal identification has increased the need for an accurate and efficient ways of identification, verification and authentication. Pattern recognition is widely used now-a-days for the biometric personal identification systems [1] [2]. Pattern matching compares the user template with templates from the database using a matching metric. This paper presents a new and simple approach for pattern recognition based on template matching. Pattern Matching is implemented in NI Vision Assistant and NI LabVIEW.

Keywords:- Pattern Recognition, Convolution kernel, Color extraction, Filtering, Normalized Cross Correlation, Scale and Rotation-Invariant Matching, Pyramidal Matching, NI Vision Assistant, NI LabVIEW.

1. INTRODUCTION

In recent years, there has been a growing interest in security issues at large and biometric-based person identification systems [3] in particular. Pattern matching [4] quickly locates regions of a grayscale image that match a known reference pattern, also referred to as a model or template. A template is an idealized representation of a feature in the image. Pattern matching [5] is the fundamental tool for implementing a biometric based person identification system. Pattern matching compares the user template with templates from the database using a matching metric. The matching metric will give a measure of similarity between the two templates. It gives a range of values when comparing templates from the same image, and another range of values when comparing templates from different images. Finally, a decision with high confidence level is made to identify whether the user is authentic or imposter.

2. METHODOLOGY

In our proposed methodology, an image of eye is captured which is then pre-processed using a color plane extraction (HSL- Luminance Plane). The subsequent image is then further processed using a convolution filter of kernel size 7x7. The resultant image is matched with the template stored in the database. After that authentication process gives the exact match. The three main stages of pattern recognition system are image preprocessing, pattern matching and authentication.

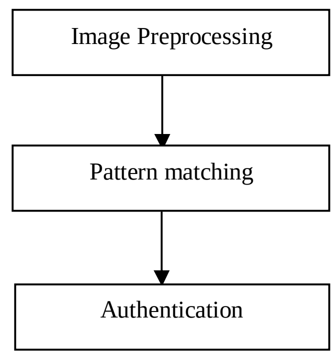


Figure 1. Flowchart illustrating overview of the process

2.1 Image Preprocessing

Image preprocessing is required for feature extraction phase. Image preprocessing consists of two phases:-

2.1.1 Color extraction phase

In this phase of image preprocessing, RGB or HSL (colored) image is converted to gray scale image [6]. This conversion is required for pattern matching as NI Vision Assistant matches gray scale templates. NI Vision converts the images by mathematical equations given below:

2.1.1.1 RGB to Grayscale

The following equations convert an RGB image into a grayscale image on a pixel-by-pixel basis.

$$\textit{Grayscale value} = 0.299R + 0.587G + 0.114B$$

This equation is part of the NTSC standard for luminance. An alternative conversion from RGB to grayscale is a simple average:

$$\text{Grayscale value} = (R + G + B) / 3$$

2.1.1.2 RGB and HSL-

There is no matrix operation that allows us to convert from the RGB color space to the HSL color space. The following equations describe the nonlinear transformation that maps the RGB color space to the HSL color space.

V2 = √3 (G – B)

V1 = 2R – G – B

L = 0.299R + 0.587G + 0.114B

H = 256tan–1 (V2 / V1) / (2π)

S = 255(1 – 3min(R, G, B) / (R + G + B))

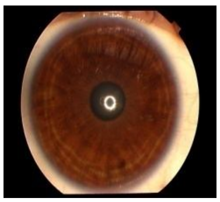


Figure 2 (a) RGB Image

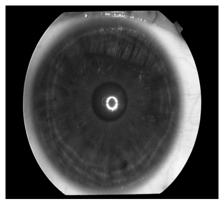


Figure 2 (b) Grayscale Image

2.1.2 *Filtering*

Then the gray scale image is further processed by using a convolution filter. A filtering operation on an image involves moving the kernel from the leftmost and topmost pixel in the image to the rightmost and bottommost point in the image. At each pixel in the image, the new value is computed using the values that lie under the kernel.

Filters are divided into two types: linear (also called convolution) and nonlinear. Convolution kernel [4] is used to filter a grayscale image. Filtering a grayscale image enhances the quality of the image to meet the requirements of our application. We used convolution filter to smooth an image, remove noise from an image, and enhance the edge information in an image.

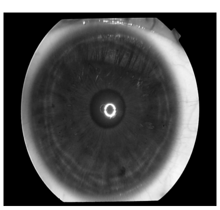


Figure 3 (a) Grayscale Image

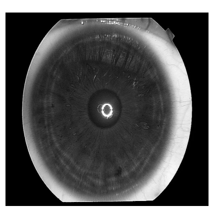


Figure 3(b) Filtered image

A convolution [7] is an algorithm that consists of recalculating the value of a pixel based on its own pixel value and the pixel values of its neighbors weighted by the coefficients of a convolution kernel. The sum of this calculation is divided by the sum of the elements in the kernel to obtain a new pixel value. The size of the convolution kernel does not have a theoretical limit and can be either square or rectangular (3 × 3, 5 × 5, 5 × 7, 9 × 3, 127 × 127, and so on).

NI Vision features a set of standard convolution kernels for each family and for the usual sizes (3 × 3, 5 × 5, and 7 × 7).

<u>Mathematical discussion of filtering</u>:

If $P_{(i,j)}$ represents the intensity of the pixel *P* with the coordinates (*i*, *j*), the pixels surrounding $P_{(i,j)}$ can be indexed as follows (in the case of a 3 × 3 matrix):

$P_{(i-1,j-1)}$	$P_{(i,j-1)}$	$P_{(i+1,j-1)}$
$P_{(i-1,j)}$	$P_{(i,j)}$	$P_{(i+1,j)}$
$P_{(i-1,j+1)}$	$P_{(i,j+1)}$	$P_{(i+1,j+1)}$

A linear filter assigns to $P_{(i,j)}$ a value that is a linear combination of its surrounding values. For example,

$$P_{(i,j)} = P_{(i,j-1)} + P_{(i-1,j)} + 2P_{(i,j)} + P_{(i+1,j)} + P_{(i,j+1)} \quad [4]$$

A nonlinear filter assigns to *P* (*i*, *j*) a value that is not a linear combination of the surrounding values. For example,

$$P_{(i,j)} = \max\left(P_{(i-1,j-1)}, P_{(i+1,j-1)}, P_{(i-1,j+1)}, P_{(i+1,j+1)}\right) \quad [4]$$

In the case of a 5 × 5 neighborhood, the *i* and *j* indexes vary from –2 to 2. The series of pixels that includes $P_{(i,j)}$ and its surrounding pixels is annotated as $P_{(n,\ m)}$.

For each pixel $P_{(i,j)}$ in an image where *i* and *j* represent the coordinates of the pixel, the convolution kernel is centered on $P_{(i,j)}$. Each pixel masked by the kernel is multiplied by the coefficient placed on top of it. $P_{(i,j)}$ becomes either the sum of these products divided by the sum of the coefficient or 1, depending on which is greater.

In the case of a 3 × 3 neighborhood, the pixels surrounding $P_{(i,j)}$ and the coefficients of the kernel, *K*, can be indexed as follows:

$P_{(i-1,j-1)}$	$P_{(i,j-1)}$	$P_{(i+1,j-1)}$
$P_{(i-1,j)}$	$P_{(i,j)}$	$P_{(i+1,j)}$
$P_{(i-1,j+1)}$	$P_{(i,j+1)}$	$P_{(i+1,j+1)}$

$K_{(i-1,j-1)}$	$K_{(i,j-1)}$	$K_{(i+1,j-1)}$
$K_{(i-1,j)}$	$K_{(i,j)}$	$K_{(i+1,j)}$
$K_{(i-1,j+1)}$	$K_{(i,j+1)}$	$K_{(i+1,j+1)}$

The pixel $P_{(i,j)}$ is given the value $(1/N) \sum K_{(a,b)} P_{(a,b)}$ with *a* ranging from (*i* – 1) to (*i* + 1), and *b*

ranging from (j – 1) to (j + 1). N is the normalization factor, equal to $\sum K_{a,b}$ or 1, whichever is greater.
If the new value $P_{(i,j)}$ is negative, it is set to 0. If the new value $P_{(i,j)}$ is greater than 255, it is set to 255 (in the case of 8-bit resolution).

The greater the absolute value of a coefficient $K_{(a,b)}$, the more the pixel $P_{(a,b)}$ contributes to the new value of $P_{(i,j)}$. If a coefficient $K_{(a,b)}$ is 0, the neighbor $P_{(a,b)}$ does not contribute to the new value of $P_{(i,j)}$ (notice that $P_{(a,b)}$ might be $P_{(i,j)}$ itself).

If the convolution kernel is

$$\begin{matrix} 0 & 0 & 0 \\ -2 & \mathbf{1} & 2 \\ 0 & 0 & 0 \end{matrix}$$

then $P_{(i,j)} = (-2P_{(i-1,j)} + P_{(i,j)} + 2P_{(i+1,j)})$

If the convolution kernel [8] is

$$\begin{matrix} 0 & 1 & 0 \\ 1 & \mathbf{0} & 1 \\ 0 & 1 & 0 \end{matrix}$$

then $P_{(i,j)} = (P_{(i,j-1)} + P_{(i-1,j)} + P_{(i+1,j)} + P_{(i,j+1)})$

2.2 Pattern Matching

After image preprocessing, pattern matching is to be done. As pattern matching is the first step in many machine vision applications, it must work reliably under various conditions. In automated machine vision applications, the visual appearance of materials or components under inspection can change because of varying factors such as part orientation, scale changes, and lighting changes. The pattern matching tool must maintain its ability to locate the reference patterns despite these changes. There are different Pattern Matching techniques used now-a-days. Pattern matching techniques include normalized cross-correlation, pyramidal matching, scale- and rotation-invariant matching, and image understanding.

2.2.1 Normalized Cross-Correlation

Normalized cross-correlation [10] is the most common method for finding a template in an image. Because the underlying mechanism for correlation is based on a series of multiplication operations, the correlation process is time consuming. The following is the basic concept of correlation: Consider a subimage $w(x, y)$ of size $K \times L$ within an image $f(x, y)$ of size $M \times N$, where $K \leq M$ and $L \leq N$. The correlation between $w(x, y)$ and $f(x, y)$ at a point (i, j) is given by

$$C(i,j) = \sum_{x=0}^{L-1}\sum_{y=0}^{k-1} w(x,y) f(x+i, y+j)$$

where $i = 0, 1 \ldots M-1$,
$j = 0, 1 \ldots N-1$, and the summation is taken over the region in the image where w and f overlap.

Basic correlation is very sensitive to amplitude changes in the image, such as intensity, and in the template. For example, if the intensity of the image f *is* doubled, so are the values of c. We can overcome sensitivity by computing the normalized correlation coefficient, which is defined as:

$$R(i,j) = \frac{\sum_{x=0}^{L-1}\sum_{y=0}^{K-1}(w(x,y)-\bar{w})\left(f(x+i,y+j)-\bar{f}(i,j)\right)}{\left[\sum_{x=0}^{L-1}\sum_{y=0}^{K-1}(w(x,y)-\bar{w})^2\right]^{1/2}\left[\sum_{x=0}^{L-1}\sum_{y=0}^{K-1}\left(f(x+i,y+j)-\bar{f}(i,j)\right)^2\right]^{1/2}}$$

where $\bar{w}$ (calculated only once) is the average intensity value of the pixels in the template w. The variable $\bar{f}$ is the average value of f in the region coincident with the current location of w. The value of R lies in the range –1 to 1 and is independent of scale changes in the intensity values of f and w [11].

Technologies such as MMX allow for parallel multiplications and reduce overall computation time. To increase the speed of the matching process, reduce the size of the image and restrict the region of the image in which the matching occurs. Pyramidal matching and image understanding are two ways to increase the speed of the matching process.

2.2.2 Scale- and Rotation-Invariant Matching

Normalized cross-correlation is a good technique for finding patterns in an image when the patterns in the image are not scaled or rotated.

Typically, cross-correlation can detect patterns of the same size up to a rotation of 5 degree. to 10 degree. Extending correlation to detect patterns that are invariant to scale changes and rotation is difficult. For scale-invariant matching [11] [12], we must repeat the process of scaling or resizing the template and then perform the correlation operation. This adds a significant amount of computation to our matching process. Normalizing for rotation is even more difficult. If a clue regarding rotation can be extracted from the image, we can simply rotate the template and perform the correlation. However, if the nature of rotation is unknown, looking for the best match requires exhaustive rotations of the template. We can also carry out correlation in the frequency domain using the Fast Fourier Transform (FFT). If the image and the template are of the same size, then this approach is more efficient than performing correlation in the spatial domain. In the frequency domain, correlation is obtained by multiplying the FFT of the image by the complex conjugate of the FFT of the template. Normalized cross-correlation is considerably more difficult to implement in the frequency domain.

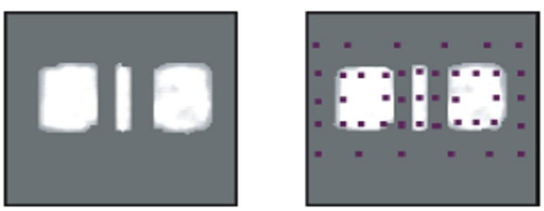
2.2.3 Pyramidal Matching

We can improve the computation time of pattern matching by reducing the size of the image and the template. In pyramidal matching, both the image and the template are sampled to smaller spatial resolutions. For instance, by sampling every other pixel [4], the image and the template can be reduced to one-fourth of their original sizes. Matching is performed on the images of reduced size. As the images are smaller, matching is faster. When matching is complete, only the areas with high match scores need to be considered as matching areas in the original image.

2.2.4 Image Understanding

A pattern matching feature is a salient pattern of pixels that describe a template. Because most images contain redundant information, using all the information in the image to match patterns is time-insensitive and inaccurate.

NI Vision uses a non-uniform sampling technique that incorporates image understanding for thorough and efficient description of the template. This intelligent sampling technique specifically includes a combination of edge pixels and region pixels as shown in Figure 4.b



a. b.

Figure 4. Good Pattern Matching Sampling Technique

NI Vision uses a similar technique when the user indicates that the pattern might be rotated in the image. This technique uses specially chosen template pixels whose values—or relative changes in values—reflect the rotation of the pattern. Intelligent sampling [13] of the template reduces the redundant information and emphasizes the feature to allow for an efficient, yet robust, cross-correlation implementation. NI Vision pattern matching is able to accurately locate objects that vary in size (±5%) and orientation (between 0 degree and 360 degree) and that have a degraded appearance.

2.3 Authentication

This is the final stage in which the result of the Pattern Matching is analyzed. If the template matches with the image, then the pattern is said to be recognized. Authentication stage is designed using NI LabVIEW. The LabVIEW code for the Pattern Matching is imported from NI Vision Assistant by using Create LabVIEW VI feature. This LabVIEW VI is modified for the authentication phase.

3. EXPERIMENTAL RESULTS

Pattern matching was performed using NI VISION ASSISTANT and NI LabVIEW. The NI Vision script for Pattern Recognition using template matching is shown in Figure 5.

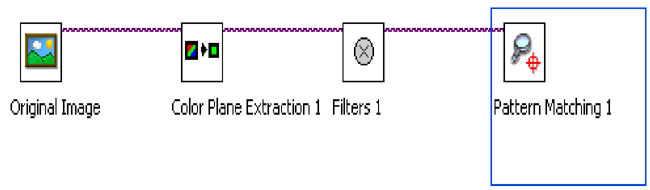


Figure 5. NI Vision Script for Pattern Matching

The results of five iris images with their own templates are shown in Fig. 6 to Fig. 10.

Results ...	1
X Position	388.00000
Y Position	299.50000
Angle	0.000000
Score	1000.00000

Figure 6.

Results ...	1
X Position	354.00000
Y Position	222.50000
Angle	0.000000
Score	1000.00000

Figure 7.

Results ...	1
X Position	463.50000
Y Position	209.50000
Angle	0.000000
Score	1000.00000

Figure 8.

Results ...	1
X Position	399.50000
Y Position	227.00000
Angle	0.000000
Score	1000.00000

Figure 9.

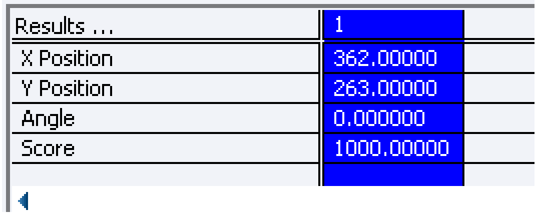


Figure 10.

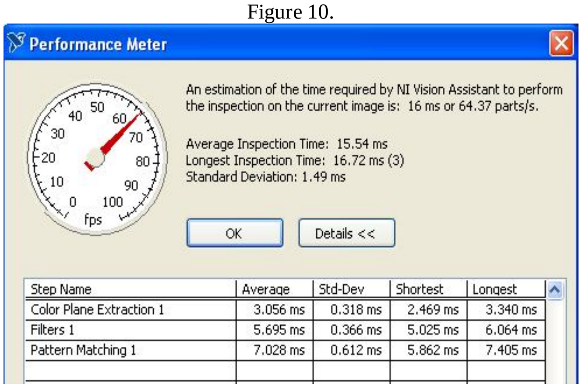


Figure 11. Performance Meter for Pattern Recognition

In the above figures, the score is 1000. This means that the pattern is matched (recognized). This experiment was performed on 21 iris images and their templates and the results were very accurate. Results for 5 images are shown in the above figures. Pattern Matching Script required 15.54 ms as shown in Fig. 11. Fig.12, Fig .13(a) and Fig. 13(b) shows the front panel and block diagram of Pattern Recognition VI.

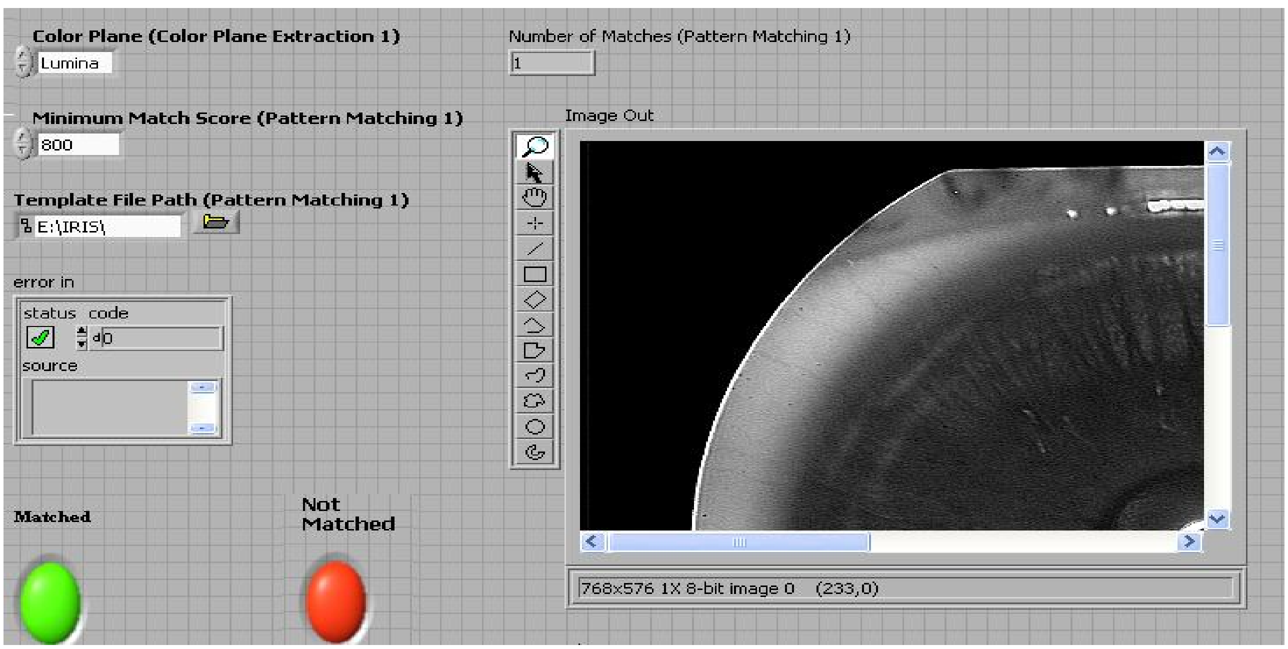


Figure 12. Front Panel of Pattern Recognition VI

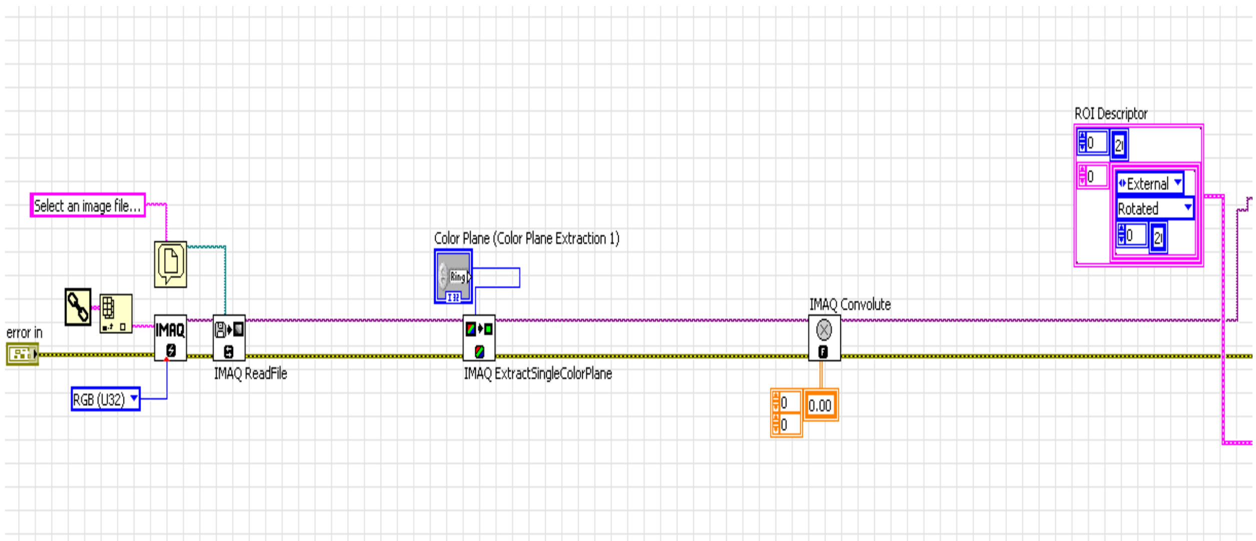


Figure 13(a) Block Diagram of Pattern Recognition VI

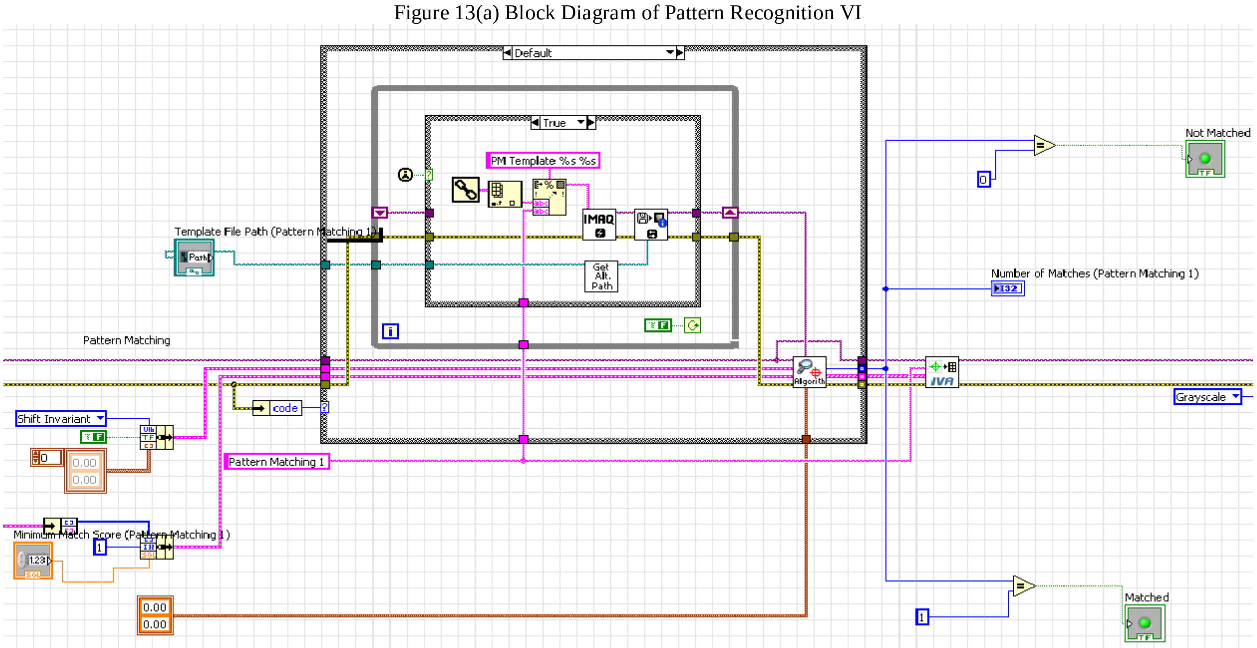


Figure 13(b) Block Diagram of Pattern Recognition VI

4. CONCLUSION AND FUTURE WORK

Pattern recognition quickly locates regions of a grayscale image that match a known reference pattern, also referred to as a model or template. Pattern recognition is widely used now-a-days for the biometric personal identification systems. This paper presented a novel approach for pattern recognition based on template matching. Template matching was performed using NI Vision Assistant. NI Vision script was generated for template matching. This script was tested on 21 images from database [14]. Results for five images were shown. NI LabVIEW was used for developing graphical user interface. Our future work will aim at developing biometric identification [15] system based on this pattern recognition.

REFERENCES

[1] Shimaa M. Elsherief, Mahmoud E. Allam, Mohamed W. Fakhr, *Biometric Personal Identification Based on Iris Recognition*. The 2006 International Conference on Computer Engineering and Systems. 5-7 Nov. 2006, 208-213.

[2] D. Lauber,” Biometrics: A Brief Overview”, SANS Institute 2003.

[3] Z. Riha and V. Matyas, "Biometric Authentication Systems", FI-MU Report Series, November 2000.

[4] NI Vision Concepts Manual, June 2008.

[5] R.O. Duda, P.E. Hart and D.G. Stork,”Pattern Classification”, Second Edition, Wiley Interscience, 2001.

[6] R. C. Gonzalez and P. Wintz, "Digital Image Processing", Second Edition, Addison-Wesley Pub., 1987.

[7] Shan Duan, Qian-Qing Qin, "Convolution Morphological Filters," icig,, Fourth International Conference on Image and Graphics (ICIG 2007), 2007, 121-125.

[8] Z.V. Babic, D.P. Mandic , *”An efficient noise removal and edge preserving convolution filter,”*. 6th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Service, 2003. TELSIKS 2003, 538-541 vol. 2.

[9] Gupta, Manjari Pande, Akshara Rao, Rajwant Singh A.K. Tripathi,”*Design Pattern Detection by normalized cross correlation,”* 2010 International Conference on Methods and Models in Computer Science (ICM2CS), 13-14 Dec. 2010, 81.

[10] Feng Zhao Qingming Huang Wen Gao,”*Image Matching by Normalized Cross-Correlation,”* In Proceedings of 2006 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2006.14-19 May 2006, II vol. 2.

[11] A.B. Yoldemir, M. Sezgin,”*Rotation and scale invariant template matching applied to buried object discrimination in GPR data,” 2010 IEEE International* Geoscience and Remote Sensing Symposium (IGARSS), 25-30 July 2010, 3366.

[12] Heung-Ho Lee, Hee-Yong Kwon and Hee-Yeung Hwang **,”***Scale and rotation invariant pattern recognition using complex-log mapping and translation invariant neural network,”* IEEE World Congress on Computational Intelligence, 27 Jun-2 Jul 1994, 4306 vol. 7.

[13] Image understanding ,” NI Vision Concepts Manual,” June 2008, 237-238.

[14] Iris database- http://phoenix.inf.upol.cz/iris/download/

[15] J. Daugman,”*Biometric Personal Identification System Based on Iris Analysis”,* United States Patent, no.5291560, 1994

Medical Images Edge Detection Using Mathematical based Morphology

Himanshu Dutt

M.Tech ECE, Amritsar College of Engg. And Technology, Amritsar

Email: khiladi.143@gmail.com

***Abstract:* - Medical images edge detection is an important work for object recognition of the human organs and it is an important pre-processing step in medical image segmentation and 3D reconstruction. Conventionally, edge is detected according to some early brought forward algorithms such as gradient-based algorithm and template-based algorithm, but they are not so good for noise medical image edge detection. In this paper, basic mathematical morphological theory and operations are introduced at first, and then a novel mathematical morphological edge detection algorithm is proposed to detect the edge of lungs CT image with salt-and-pepper noise. The experimental results show that the proposed algorithm is more efficient for medical image denoising and edge detection than the usually used template-based edge detection algorithms and general morphological edge detection algorithms.**

Keywords- Medical image, edge detection, mathematical morphology, denoising

I. INTRODUCTION

Medical images edge detection is an important work for object recognition of the human organs such as lungs and ribs, and it is an essential pre-processing step in medical image segmentation. The work of the edge detection decides the result of the final processed image. Conventionally, edge is detected according to some early brought forward algorithms like Sobel algorithm, Prewitt algorithm and Laplacian of Gaussian operator , but in theory they belong to the high pass filtering, which are not fit for noise medical image edge detection because noise and edge belong to the scope of high Frequency. In real world applications, medical images contain object boundaries and object shadows and noise. Therefore, they may be difficult to distinguish the exact edge from noise or trivial geometric features. Mathematical morphology is a new mathematical theory which can be used to process and analyze the images. It provides an alternative approach to image processing based on shape concept stemmed from set theory, not on traditional mathematical modeling and analysis. In the mathematical morphology theory, images are treated as sets, and morphological transformations which derived from Minkowski addition and subtraction are defined to extract features in images. As the performance of classic edge detectors degrades with noise, morphological edge detector has been studied. In this paper, a novel mathematical morphology edge detection algorithm is proposed to detect lungs CT medical image edge. It is a better method for edge information detecting and noise filtering than differential operation, which is sensitive to noise. And it is a better compromise method between noise smoothing and edge orientation, but the computation is more complex than general morphological edge detection algorithms.

MATHEMATICAL MORPHOLOGICAL OPERATION-

Mathematical morphology is developed from set theory. It was introduced by atheron as a technique for analyzing geometric structure of metallic and geologic samples. It was extended to image analysis Based on set theory, mathematical morphology is a very important theory, whose operation must be defined by set arithmetic. Therefore, the image which will be processed by mathematical morphology theory must been changed into set. Mathematical morphology uses structuring element, which is characteristic of certain structure and feature, to measure the shape of image and then carry out image processing. Based on set theory, mathematical morphology is the operation which transforms from one set to another. The aim of this transformation is to search the special set structure of original set. The transformed set includes the information of the special set structure and the transformation is realized by special structuring element. Therefore, the result is correlative to some characteristics of structuring element. The basic mathematical morphological operators are dilation and erosion and the other morphological operations are the synthesization of the two basic operations. In the following, we introduce some basic mathematical morphological operators of grey-scale images. Erosion is a transformation of shrinking, which decreases the grey-scale value of the image, while dilation is a transformation of expanding, which increases the grey-scale value of the image. But both of them are sensitive to the image edge whose grey-scale value changes obviously. Erosion filters the inner image while dilation filters the outer image. Opening is erosion followed by dilation and closing is dilation followed by erosion. Opening generally smoothes the contour of an image, breaks narrow gaps. As opposed to opening, closing tends to fuse narrow breaks, eliminates small holes, and fills gaps in the contours. Therefore, morphological operation is used to detect image edge, and at the same time, denoise the image.

PROPOSED MORPHOLGICAL EDGE DETECTION ALGORITHMS

Morphological edge detection algorithm selects appropriate structuring element of the processed image and makes use of the basic theory of morphology including erosion, dilation, opening and closing operation and the synthesization operations of them to get clear image edge. In the process, the synthesized modes of the operations and the feature of structuring element decide the result of the processed image. Detailedly saying, the synthesized mode of the operations reflects the relation between the processed image and origin image, and the selection of structuring element decides the effect and precision and the result. Therefore, the keys of morphological operations can be generalized for the design of morphological filter structure and the selection of structuring element. In medical image edge detection, we must select appropriate structuring element by texture features of the image. And the size, shape and direction of structuring element must been considered roundly. Usually, except for special demand, we select structuring element by 3×3 square. The effect of erosion and dilation operations is better for image edge by performing the difference between processed image and original image, but they are worse for noise filtering. As opposed to erosion and dilation, opening and closing operations are better for filtering. But because they utilize the complementarity of erosion and dilation, the result of processed image is only correlative with the convexity and concavity of the image edge. Accordingly, what we get is only the convex and concave features of the image by performing the difference between processed image and original image, but not all the features of image edge. In this paper, a novel mathematical morphology edge detection algorithm is proposed. Opening-closing operation is firstly used as preprocessing to filter noise. Then smooth the image by first closing and then dilation. The perfect image edge will be got by performing the difference between the processed image by above process and the image before dilation.

EXPERIMENTAL RESULT AND ANALYSIS

In this section, the proposed morphological edge detection algorithm is compared with a variety of existing methods for edge detection. Fig.1 is the original lungs CT image with salt-and-pepper noise. Fig.2 and Fig.3 are the results of processed lungs CT images after respectively applying Laplacian of Gaussian operator and Sobel edge detector. Fig.4 and Fig.5 are the lungs CT images processed by morphological gradient operation and dilation residue edge Detector.

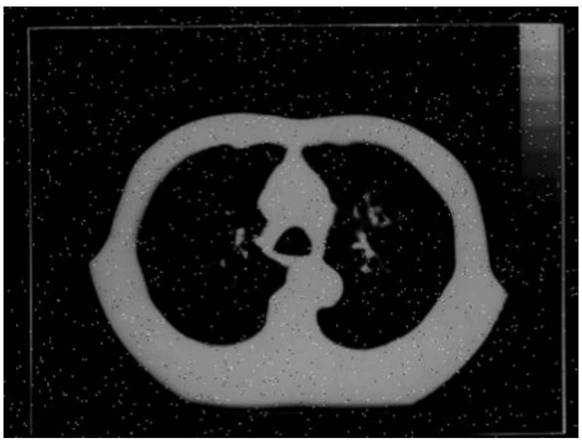


Fig.1. Original lungs CT image with salt-and- pepper noise.

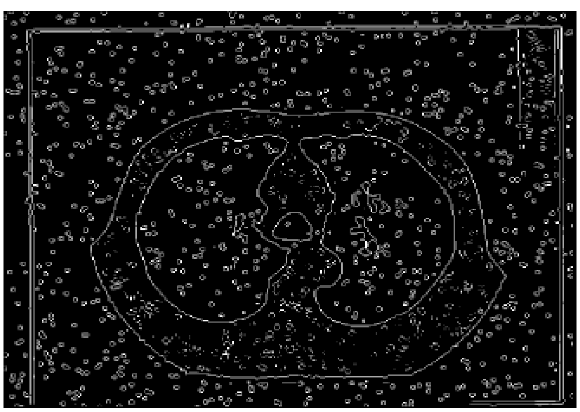


FIG.2. Lungs CT image processed by Laplacian of Gaussian operator.

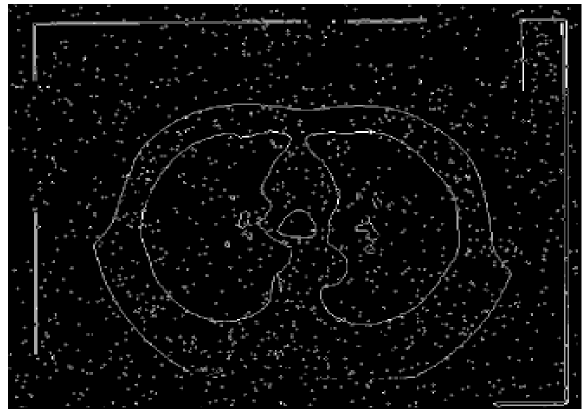


Fig.3. Lungs CT image processed by Sobel detector

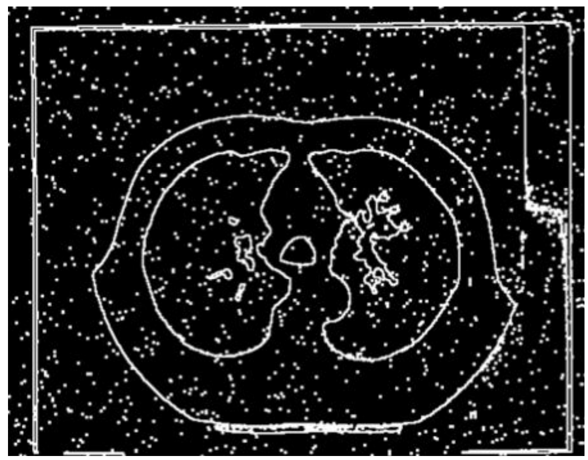


Fig.4. Lungs CT image processed by Morphological Gradient Operation.

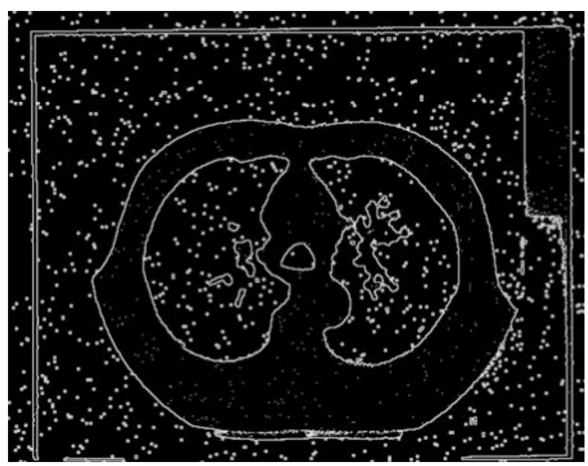


Fig.5. Lungs CT image processed by dilation residue edge detector.

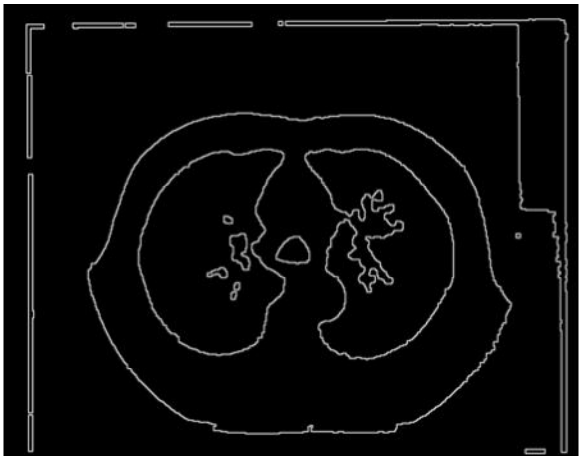


Fig.6. Lungs CT image processed by novel morphological edge detector.

According to the experiment results shown in Fig.2 and Fig.3, Laplacian of Gaussian operator and Sobel edge detector detect the lungs edges successfully, but Sobel edge detector fail to detect the outer edge of body, and both of them can't filter the noise. By Fig.4 and Fig.5, the morphological gradient operation and dilation residue edge detector are succeed in lungs and body edges detection, and the detected edges are clearer than the edges detected by Laplacian of Gaussian operator and Sobel edge detector. But both of them fail to filter the noise in despite of the latter is better for noise filtering than the former. By Fig.6, the novel morphological edge detector proposed in this paper is succeed in lungs and body edges detection, but more important than template-based edge detection algorithm and general morphological edge detection algorithm mentioned before, it also filters the noise successfully.

CONCLUSION

In this paper, a novel mathematic morphological algorithm is proposed to detect lungs CT medical image edge. The experimental results show that the algorithm is more efficient for medical image denoising and edge detecting than the usually used template-based edge detection algorithms such as Laplacian of Gaussian operator and Sobel edge detector, and general morphological edge detection algorithm such as morphological gradient operation and dilation residue edge detector.

REFERENCES

[1] M.I. Rajab, M.S. Woolfson, and S.P. Morgan, "Application of region-based segmentation and neural network edge detection to skin lesions," Computerized Medical Imaging and Graphics, vol. 28, pp.61–68, 2004.

[2] H. Tang, E.X. Wu, Q.Y. Ma, D. Gallagher, G.M. Perera, and T. Zhuang, "MRI brain image segmentation by multi-resolution edge detection and region selection," Computerized Medical Imaging and Graphics, vol 24, pp. 349–357, 2000.

[3] Huertas, A. and Medioni, G., "Detection of intensity changes with sub pixel accuracy using Laplacian-Gaussian masks," IEEE Trans. On Pattern Analysis and Machine Intelligence, PAMI, vol. 8, pp. 651–664, 1986.

[4] Maragos P., "Differential Morphology and Image Processing," IEEE Trans Image Processing, vol. 5, pp. 922–937, June 1996.

[5] Ortiz F, and Torres F, "Vectorial Morphological Reconstruction for Brightness Elimination in Colour Images," Real-Time Imaging, vol. 10, pp. 379–387, June 2004.

[6] Jing Xiao-jun, Yu Nong, and Shang Yong, "Image Filtering Based on Mathematical Morphology .

Electronic-Nose

Sukhwinder kaur
M.Tech ECE,Amritsar College of Engg. & Technology
malhi_sukh@ymail.comn

***Abstract*-Until now, online communication involved only two of our senses, sense of sight & sense of hearing. Soon it will involve the third, the sense of smell using an e-nose. Digital scent technology is the main application of e-nose. With digital scent technology, it is possible to sense, transmit & receive smell through internet.There is complete software and hardware solution for it. When applied to communications, scent becomes a new information channel. It allows us to perceive products and irate a previously unimagined emotionality and product credibility. Scents extend the myriad of multimedia possibility towards a new level. Scent communication will be one of the most important information tools of the future.**

***Keywords*— Electronic nose, optical fiber sensors, mosfet sensors, quartz sensors, olfaction, electronic aroma detection.**

I. INTRODUCTION

Until now online communication involved only two of our senses, sense of hearing and sense of sight. Soon it will involve the third, the sense of smell. A new technology is being developed to appeal to our sense of smell. Bringing alive our experience, technology now targets on the sense of smell. An electronic nose (e-nose) is a device that identifies the specific components of an odor and analyses its chemical makeup to identify it. An electronic nose consists of a mechanism for chemical detection, such as an array of electronic sensors, and a mechanism for pattern recognition, such as a neural network Using Electronic-nose we can sense a smell and with a technology called Digital scent technology it is possible to sense, transmit and receive smell through internet, like smelling a perfume online before buying them, sent scented E-cards through scent enabled websites, and to experience the burning smell of rubber in your favourite TV games etc. If this technology gains mass appeal no one can stop it from entering into virtual world. Just imagine you are able to smell things using a device connected to your computer. With Digital scent technology this can be made a reality. There is complete software and hardware solution for scenting digital media and user .Before we describe the possibilities of olfactory displays, we should take a glance at the physiological aspects of smell. How does the nose work and what is its function? Naturally we can breath, smell and additionally taste with our nose. First of all we are interested in the anatomy of the nose. Odor consists of many different molecules, for e.g. the aroma of coffee is made up of 20 various molecules. Nonetheless our nose perceives only 15 odors which is enough to identify the smell as coffee. At first the odor molecules reach the olfactory mucosa. The receptors for the molecules are placed at the olfactory hairs. When the molecules reach there ceptors, an electric impulse is sent directly to the brain to the olfactory bulb. Then the information gets to the olfactory glomerli a part of the olfactory hairs. When the molecules reach there ceptors, an electric impulse is sent directly to the brain to the olfactory bulb.Then the information gets to the olfactory glomeruli, a part of the olfactory bulb.The glomeruli is able to associate the information to the intensity. The olfactory bulb consequently processes the odor and can send the impulse to the olfactory brain. We notice that we have a direct connection between our sense of smell in land our brain. Those scent impulses reach the area of our brain that handles emotions and memories. That explains the link between smelling and being reminded of something.We percept smell very individually. Every human perceive a difference between a pleasant and unpleasant odor. Humans are not capable to distinguish odors in terms of intensity. Roughly we can only distinguish between three concentrations of some odor whereas we should actually be able to differentiate 1000 types of odors. Another problem for olfactory display is the fast acclimatization of humans to scents. What makes it even more difficult to construct olfactory display is that a set of primary odors has not really been found. There was an attempt to define seven such of primary odors but had to be extended to 100 odors. For vision, three base colours are sufficient to display any colour. Unfortunately this cannot be applied to olfaction as our nose has thousands of receptors and apart from that the odors are not orthogonal. That means you will not necessarily get a new one by mixing two odors.

II. ELECTRONIC NOSE

An electronic nose can be a modular system comprising of active materials which operate serially on an odorant sample. These active material scan be classified into two: an array of gas sensors and a signal processing system. The output of the electronic nose can be the identification of the odorant, an estimation of the concentration of the odorant or the characteristic of the odor as might be perceived by the human. Fundamental of artificial nose is that each sensor in the array has different sensitivity. The pattern of response across the sensors is distinct for different odors. This distinguish ably allows the system to identify the unknown odor from the pattern of sensor responses. The pattern of response across all the sensors in the array is used to identify the odor. Different enoses use different types of gas sensors which form heart of e-nose.

In a typical e-nose, an air sample is pulled by a vacuum pump through a tube into a small chamber housing the electronic.

III. COMPARISION OF ELECTRONIC NOSE WITH BIOLOGICAL NOSE

Each and every part of the electronic nose is similar to human nose. The function of inhaling is done by the pump which leads the gas to the sensors. The gas inhaled by the pump is filtered which in the human is the mucus membrane. Next comes the sensing of the filtered gas, which will be done by the sensors i.e., olfactory epithelium in human nose. Now in electronic nose the chemical retain occurs which in human body is enzymal reaction. After this the cell membrane gets depolarised which is similar to the electric signals in the electronic nose. This gets transferred as nerve impulse through neurons i.e., neural network and electronic circuitries sensor array. Next the sampling handling units exposes the sensors to the odorant, producing a transient response as the VOCs interact with the surface and bulk of sensor's active material. A steady state condition is reached in a few seconds to a few minutes. During this interval, the sensor's response is recorded and delivered to the signal processing unit. Then a washing gas such as alcoholic vapour is applied to the array so as to remove the odorant mixture from the surface and bulk of sensor's active material. Finally a reference gas is again applied to the array to prepare it for a new measurement cycle. The period during which odorant is applied is called the response time of the sensor array. The period during which washing and reference gases are applied is called the recovery time.The sensor's response is converted into electronic signal by using a transducer and is processed by using the signal processing unit.

IV. TYPES OF SENSORS

Polymer sensors, Quartz sensors, MOSFET sensors, Optical fiber sensors.

A) polymer sensors:

The working of polymer sensors is based on the change in conductivity of the polymer when the odorant is applied. Response time is inversely proportional to the polymers thickness. The main drawback of this method is that it is difficult and time consuming to electro polymerize the active material, so they exhibitor desirable variations from one batch to another.

B) quartz sensors:

Here the vibration of the quartz is changed by a contact between the molecules and the surface. The response and recovery times are minimized by reducing the size and mass of quartz crystal along with the thickness of the polymer coating. The main disadvantage is that they have more complex electronics than of polymer sensors.

C) mosfet sensors:

These are based on the principle that VOCs in contact with a catalytic metal can produce a reaction on the metal. The reaction products can diffuse through the gate of the MOSFET to change the electrical properties of the device. The sensitivity and selectivity of the device can be optimized by varying the type and thickness of the metal catalyst and operating them at different temperatures. The advantage is that they can be made with IC fabrication so that batch to batch variations can be minimized.

D) optical fiber sensors:

A light source of single frequency is used to interrogate the active material, which in turn responds with color change in the presence of VOCs to be detected and measured. The active material contains chemically active fluorescent dyes immobilized in an organic polymer matrix. As VOCs interact with it, polarity of the fluorescent dyes is altered and they respond by shifting their fluorescent emission spectrum. These sensors are cheap and easy to fabricate. The disadvantage is that fluorescent dyes are slowly consumed by the sensing process.

V. APPLICATIONS OF ELECTRONIC NOSE

The electronic nose has been used in a variety of applications and could help solve problems in many fields. The electronic nose can be applied by food manufacturers to such tasks such as freshness testing, quality screening of incoming raw material, and monitor for accidental or intentional contamination.In the medical field, e nose has a variety of application such as rapid diagnosis of acute infection through breath analysis and screening of bacteria cultures for early detection of pathogens. E-nose can serve in safety and security applications such as hazard alarm for toxic and biological agents, screening airline passengers for explosives and drugs. Its military applications include land-mine detection, biological and chemical agent detection etc.

VI. ADVANTAGES AND DISADVANTAGES

ADVANTAGES: It can be used without fall over hours, days, weeks and even month's andcan even circumvent problems associated with the use of human panels such as individual variability, adoption, fatigue mental state and exposure to hazardous material. The e-nose is a compact device and so it is portable and reliability is very high. It can identify simple molecules which cannot be accomplished by human nose. It can identify a smell objectively.

DISADVANTAGES: There are a few disadvantages to the e-nose technology which includes the price. The cost of an e-nose ranges from \$5000 to \$100,000. Another disadvantage has been the delay between successive tests, the time delay ranging between 2to 10 minutes during which time, the sensor is to be washed by a reactivating agent, which is applied to the array so as to remove the odorant mixture from the surface and bulk of the sensors active material.

VII. APPLICATIONS

A) Application for the scent on the web

In addition to revolutionizing gaming, digital scent technology will bring consumers more life like and memorable experiences with scented movies and music, websites, e-mail, e-commerce and online advertising.

B) e-commerce

Scent will bring the online shopping experience to life. Scent-enabled shopping sites will be more compelling if you can actually smell perfumes, flowers, food and beverages, cigars and exotic place.

C) advertising

Vendors of food, cosmetics, home care products and travel related services can use scent to make advertisements more engaging and memorable. Eventually, like musical jingles and graphical logos, scented banner ads will make it possible to communicate the key feature of scented products or to simply evoke a certain feeling.

D) communication

Scent offers developers as well as consumers another medium for creativity and self-expression. For eg: scented websites, electronic greeting cards ande-mail. With smell technology you can travel anywhere in the world or to any time period in the past.

E) education

Scent is an effective teaching tool for subjects such as Geography, History and Sciences. With digi scents technology, you can travel anywhere in the world or to any time period in the past.

F) medicine

Aromatherapy is a kind of curing certain diseases by using different types of smell. It helps in identifying dementing brain disorders including Huntington's and Parkinson's and for differentiating them from other mental disorders. This method is based on detecting the olfactory defaults that are diagnostic of the dementing diseases.

G) entertainment

Scent will make music, movies and interactive games life like and immersive. Scent tracks will be synchronized with movies, much liked musical score and sound track, in order to create foreshadowing and to establish atmosphere, mood, sense of place and character.

VIII. FUTURE WORK

They proposed a naval configuration of an olfactory display that does not require user's to put anything on the face and that localizes the effective space of the displayed scent. The technical key to realizing this concept is to transfer a clump of scented air from a place near the nose, and we confirmed that this is possible by using air cannon. The constructed prototype system successfully displayed the scent to the target user, even if the user moved his head. They are going to propose another choice in methods to enjoy scent in interactive applications. The wider the variety of olfactory displays, the wider the variety of applications will emerge to make our VR experience rich and realistic. Improvement of scent generation is necessary to extend the variety of displayed scent and we can learn a lot from preceding research efforts on scent blending and generation. Also precise theoretical analysis of a toroidal vortex might be effective for optimal design of the air cannon. They are planning a step by step in order to construct a transparent, easy to use olfactory display system.

CONCLUSION

New medium in the world of communication: scent. Scents have an immediate and compelling effect. They touch our soul, consciously or unconsciously, and allow us to fell deeply. Scents are unambiguous and unmistakable. The integration of all the senses, in how we create and imagine the universe of goods, is becoming more important. The consumer is always searching for experiences. Enter into a new form of dialogue with your customers. Offer him a cache of new impulses to purchase, through the power of scent. When applied to communication, scent becomes a new information channel. It allows us to perceive products and create a previously unimagined emotionality and product credibility. At the same time seeing and hearing, scents extend the myriad of multimedia possibilities to a new level? Scent communication will be one of the most important information tools of future.

REFERENCES

[1] Hand book on applications and uses of arrays by Jeffery.H

[2] www.scienceweek.com

[3]Electronic nose and their applications By P.E. Keller, L.J.Kangas, L.H. Liden

[4] http://en.wikipedia.org/wiki/Operational_data_store

[5] http://en.wikipedia.org/wiki/electronic_nose

[6]http://seminarprojects.com

[7]http://WWW.IIT.EDU

[8]http;//www.scribd.com

Optimization of DWDM ring using Different Data Formats

Er.Anurag Sharma[1], Er.Pankaj Sharma[2], Er.Vikrant Sharma[3]
[1,2,3] *Assistant Professor ECE*
[1,2,3] *CT Intstitute of Technology Jalandhar, Punjab, India*
[1] er.anurags@gmail.com, [2] er.pankajsharma@rocketmail.com, [3] vikrant.sharma@yahoo.com

Abstract**----** The performance of different modulation formats over 250km OADM ring are discussed, Eye Diagram Q factor and BER has been monitored for RZ, NRZ & Manchester Loading schemes.

Keywords**---**OADM (Optical Add Drop Multiplexer), NRZ, RZ**.**

I. INTRODUCTION

This paper demonstrates an OptSim design for DWDM ring with Optical Add-Drop Multiplexers (OADM). The configuration of DWDM ring is in

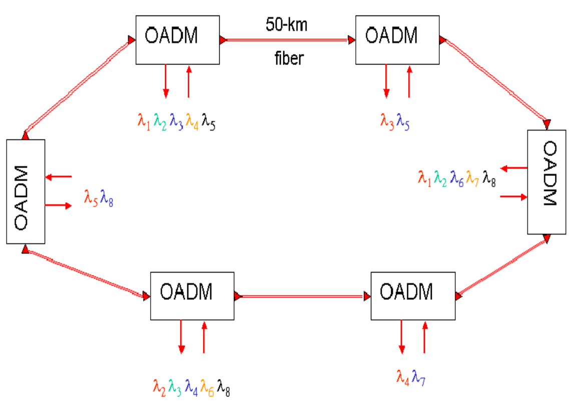


Fig 1: DWDM Ring configuration study

Fig 1. The DWDM ring consists of six nodes and six fiber spans. Total number of wavelengths used is eight, with 3 head-end nodes adding/dropping five Channels at the time and 3 serial OADM nodes adding/dropping two channels. Here for simplicity we used only uni-directional fiber link with signals propagating in clock-wise direction. Clockwise means from one OADM to another and so on in one direction only not in other direction. The project can be easily extended to full fiber pair with signals propagating in both directions.

The schematic consists of six OADM nodes connected by fiber spans of specified length and type of the fiber (Corning LEAF fiber). For simplicity, we assumed that all nodes are equidistant and all six fiber spans are 50-km long. To compensate for the fiber attenuation in fiber spans we inserted fixed gain amplifiers after each fiber span. The power per channel of -9 dBm was used at transmitters. We used 8 wavelengths at 50 GHz (0.4 nm) spacing starting at 1550 nm wavelength. After each node we put also a plotter block MultiPlot to observe an optical spectrum evolution along the ring. We use Time Delay block to connect signal from the last node back to the first node. This way we can provide steady-state solution for ring simulation.

The more details of OADM node on the example of serial OADM with 2 wavelengths adding/dropping at $\lambda 3$ and $\lambda 6$. OADM block is modeled as a compound component (CC) or hierarchy block, i.e. is composed inside from other OptSim library blocks. Each OADM node has one input and one output for line signal (in- and out-transport interfaces), and also 8 inputs/outputs for added/dropped wavelengths (in- and out-clients interfaces). Here two single-channel NRZ-modulated transmitters modulated at 10 Gbps bit rate (OC-192 rate) are connected to add-ports 3 and 6. Since the OptSim requires that all input ports for CC models to be connected we use Null Signal model and connect it to the other 6 ports. At the output of CC we can connect only the ports of interest – in this case port 3 and 6. Output from these ports are connected to optical receiver blocks and then to plotter model MultiPlot which will provide plots for electrical signal output waveform, spectrum, and eye diagram.

The OADM CC block can specify following parameters: crosstalk level, switching configuration, first channel wavelength and channel spacing, optical filter bandwidth used in demultiplexing. The properties dialog window for OADM CC block. The switching configuration is given as an N-size array (N is a number of channels in a link) where “1” stands for channel to be dropped at that port and “0” – for channel to go through. For example, in the case of Node 2 where we are adding/dropping λ3 and λ6, the switch array is given as {0,0,1,0,0,1,0,0}.

To understand how the OADM CC block works let us look inside it – see Figure 3. The OADM CC block consists of one 1x8 Demultiplexer, 8x1 Multiplexer, and 8 optical switches. The input from transport interface (Input1) is demultiplexed into 8 wavelengths and each of them goes to a switch with

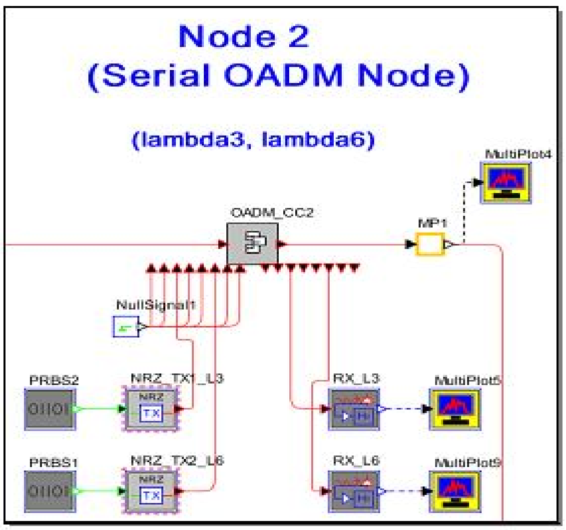


Fig 2: one of the OADM nodes with 2 wavelength being added and dropped

corresponding input from client interface (Input2-Input9). 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-Output9) and the other output is being multiplexed with 7 other outputs and then sent to transport interface output (Output1).

To understand how the OADM CC block works let us look inside it – see Figure 4. The OADM CC block consists of one 1x8 Demultiplexer, 8x1 Multiplexer, and 8 optical switches. The input from transport interface (Input1) is demultiplexed into 8

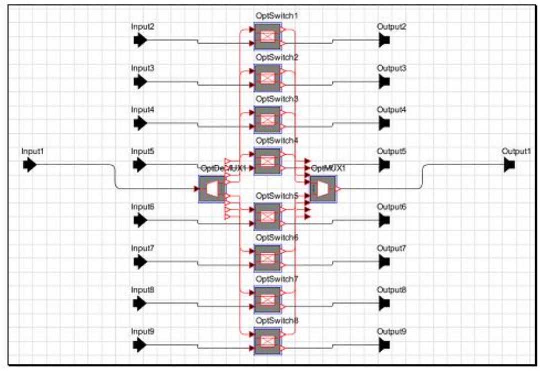


Fig 3: Inside configuration of CC for OADM

wavelengths and each of them goes to a switch with corresponding input from client interface (Input2-Input9). 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-Output9) and the other output is being multiplexed with 7 other outputs and then sent to transport interface output (Output1).

Finally we can run simulations and review the results. Figures 6 and 7 demonstrate some of the results for DWDM ring simulation. Figure 3 shows optical spectrum after two of nodes: (a) Node 1, and (b) Node 6. Figure 4 shows a few examples of eye diagrams for dropped channels at (a) Node 1 λ3 and (b) Node 6 λ8. We expect that the difference in eye diagrams comes from the difference in distance traveled before being dropped and from the difference in cumulative dispersion experience by these channels. So, eye diagram for channel 3 at Node 1 is after 100-km (coming from Node 5), and eye diagram for channel 8 at Node 6 is after 50-km (also coming from Node 5).

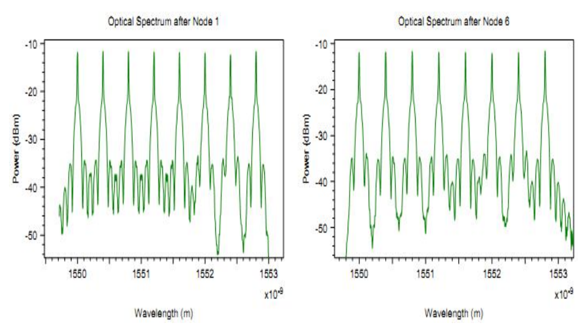


Fig 4: optical spectrum at the node output

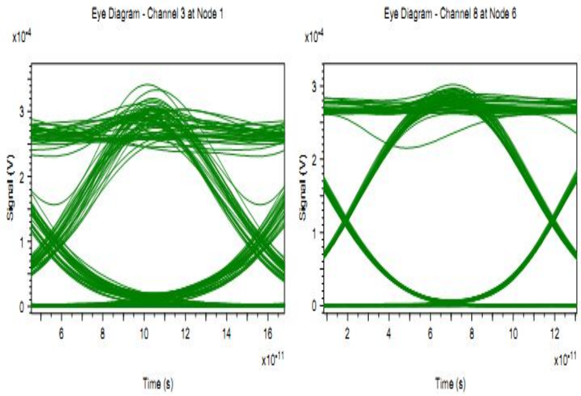


Fig 5:Receiver eye diagram for dropped channels

II.RESULTS

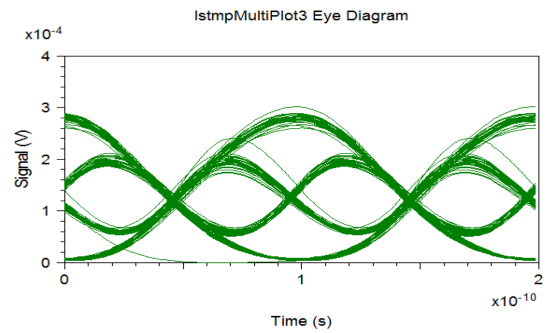
The modulation format plays an important role in the design of the system as it can minimize the effect of SPM in high speed dispersion managed optical system .We have started in this system with comparison between three type of modulation formats Manchester, RZ and NRZ and note down the eye diagram for each wavelength .Lets see the results and summarize which modulation format is the best suited for the system. As we can see from the eye diagrams RZ and NRZ both are good wide open eye patterns we have to check for BER and quality factor specifications for selection of the modulation format .The Manchester modulation does not give good results and are out of scope for further BER and quality factor checking . from BER investigation of RZ & NRZ. It has been seen that NRZ is the best modulation format. One more format which is known as CRZ modulation format is better than CRZ method. Because in CRZ format.

The lowest power level, which is set by ASE noise is nearly the same for NRZ & CRZ.

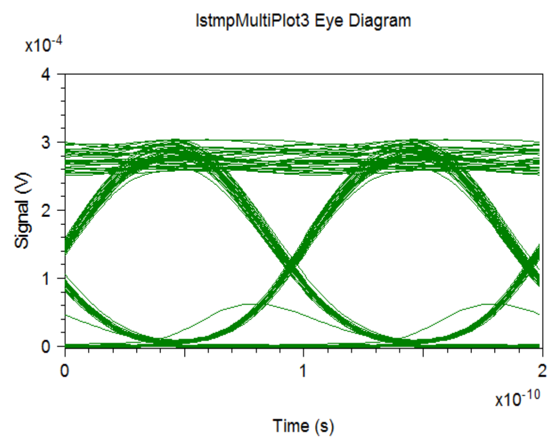
By Constant the upper level which is set by single channel non linearity is substantially higher for CRZ modulation formats.

In figure X .The comparison of eye diagrams is given we have taken for one wavelength i.e.; lambda 5 which is added at first OADM & dropped from last OADM after 250Km. can easily judge that NRZ eye diagram is much better as compared to other two & it can further improved by using CRZ.

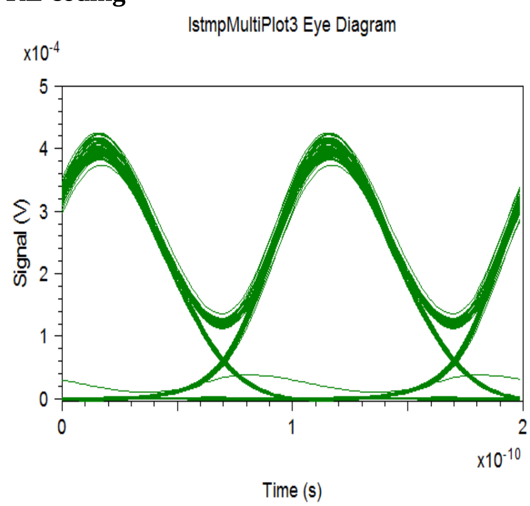
Manchester Coding



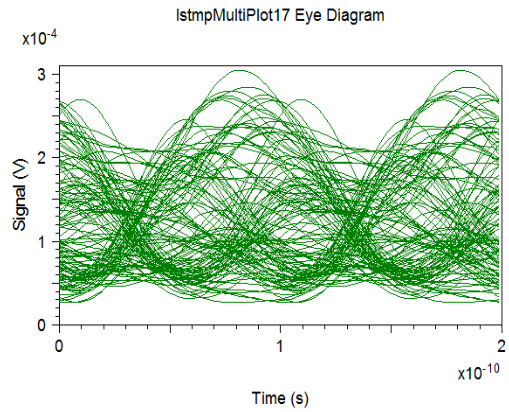
NRZ coding



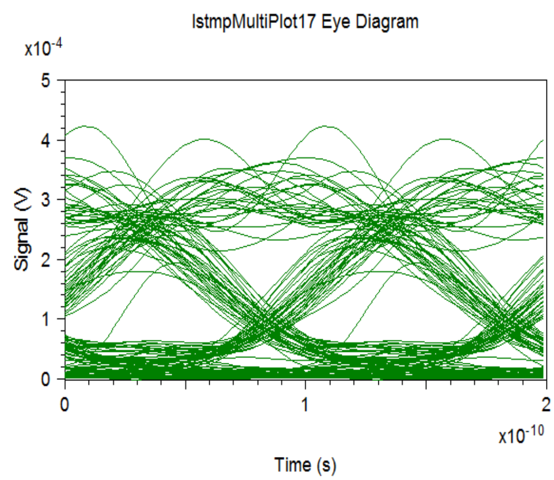
RZ coding



Manchester Coding



NRZ coding



RZ coding

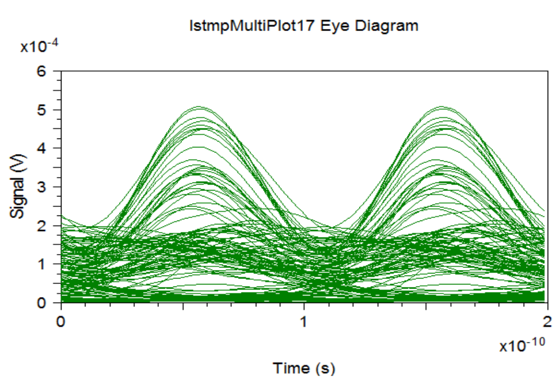


Table. 1
BER table for different formats and for different nodes

Node 1(λ 5): RZ Modulation

BER	BER_lo	BER_hi	Q^2(dB)
3.6274e-080	9.2530e-092	5.5571e-067	2.5540e+001

Node6(λ 5): RZ Modulation

BER	BER_lo	BER_hi	Q^2(dB)
1.1146e-003	4.5571e-004	2.5609e-003	9.7084e+000

Node 1(λ 5): NRZ Modulation

BER	BER_lo	BER_hi	Q^2(dB)
3.6274e-080	9.2530e-092	5.5571e-067	2.5540e+001

Node 6 (λ 5): NRZ Modulation

BER	BER_lo	BER_hi	Q^2(dB)
1.0698e-065	8.7808e-117	3.1106e-071	2.6329e+001

In table 1. There is a comparison of NRZ & RZ BER & Q^2 (dB) for lambda 5 which shows that NRZ is much better than RZ.

REFERENCES

[1] Emma Lazzeri, Truong Nguyen, Nobuyuki Kataoka, Naoya Wada, Antonella Bogoni, and Luca Potì,' All Optical Add and Drop Multiplexing Node for Hybrid Topology Networks, JOURNAL OF LIGHTWAVE TECHNOLOGY,VOL. 29, NO. 24, DECEMBER 15, 2011

[2] J.R. Yeh, M. Selker, J. Trail, D. Piehler and 1. Levi, "DWDM Architectures for Video on Demand Transport and Distribution" 1998 IEEE LEOS Summer Topical Meetings Digest, p5 (1998)

[3] Sorin Tibuleac and Mark Filer,' Transmission Impairements in DWDM Networks with Reconfigurable Optical Add-Drop Multiplexers, JOURNALOF LIGHTWAVE TECHNOLOGY, VOL. 28, and NO. 4, FEBRUARY 15, 2010

[4] www.ijser.org/researchpaper/Performance-investigations-on-DWDM-OADM-Optical-Ring-Network-for-CRZ-Data-Format-at-Different-Fiber-Length.pdf

Image Processing using Segmentation: Graph, Medical and Color Image base Segmentation Techniques

Rahul Sharma[1] and Himanshu Sharma[2]
M.Tech, ACET Amritsar

Abstract— **The general field of Image Processing is still in infancy. Image Processing represents complex field in which the input to the system is image in any form and the output of the processing obtained after using different filters and tools is the processed image in the form of any parameters related to the image. Particularly, the challenge of providing efficient image is one of the most critical task. This literature review attempts to provide a brief overview of some of the most common segmentation techniques, and a comparison between them.**

I. Introduction

The area of image processing is becoming vast due to the need of storage of documents in the electronic form and all information that was conventionally stored on paper, is now being converted into digital form. Further Different techniques are used in this field. Image segmentation is one of them.This article provides an overview of segmentation techniques including its methods such as Graph Based Methods, color image segmentation, video segmentation, Medical image based research.

Image segmentation in general is defined as a process of partitioning an image in which it subdivides an image into homogenous groups such that each region is homogenous but the union of no two adjacent regions is homogenous[1] .The level to which the subdivision is carried depends on the problem being solved. Image segmentation has been interpreted differently for different applications. Types of segmentation includes Point, Line And Edge Segmentation . Efficient image segmentation is one of the most critical tasks in automatic image processing.

SEGMENTATION METHODS

Graph Based Methods:

A family of graph-theoretical algorithms based on the minimal spanning tree are capable of detecting several kinds of cluster structure in arbitrary point sets. In the year 1971,C. T. Zahn presented a paper,where brief discussion is made of the application of cluster detection to taxonomy and the selection of good feature spaces for pattern recognition[2].

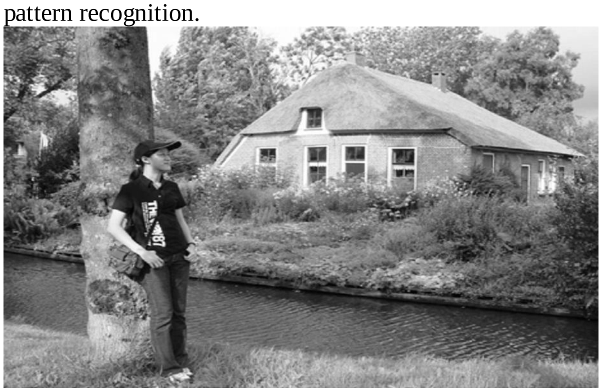
COLOR IMAGE SEGMENTATION:

Among various strategies, the evaluation of the colour attributes in succession was one of the most popular directions of research. A segmentation algorithm has been proposed by Mirmehdi and Petrou [3]. In this paper the authors introduced a colour–texture segmentation scheme where the feature integration is approached from a perceptual point of view. Here, convolution matrices are calculated using a weighted sum of Gaussian kernels and are applied to each colour plane of the opponent colour space (intensity, red–green and blue–yellow planes, respectively) to obtain the image data that make-up the perceptual tower. A related segmentation approach was proposed by Huang et al. In this paper the authors applied Scale Space Filters (SSF) to partition the image histogram into regions that are bordered by salient peaks and valleys. A well-known technique of colour–texture feature integration approach was proposed by Deng and Manjunath and this algorithm is widely regarded as a benchmark by the computer vision community. The proposed method is referred to as JSEG and consists of two computational stages, namely colour quantisation and spatial segmentation.

Video Segmentation:

Video Segmentation is required To support object-oriented video compression technology, like MPEG-4 standard (Core, Main Profile) and To support content-based video processing application. Luciano et al. performed an experiment on real and synthetic sequences in April 2010. In this paper they experimented with real and synthetic sequences suggest that their method also could be used in other image processing and computer vision tasks, besides video coding, such as video information retrieval and video understanding. Their work described an approach for object-oriented video segmentation based on motion coherence.

A family of graph-theoretical algorithms based on the minimal spanning tree are capable of detecting several kinds of cluster structure in arbitrary point sets [7][8]. In the year 1971,C. T. Zahn presented a paper,where brief discussion is made of the application of cluster detection to taxonomy and the selection of good feature spaces for pattern recognition.



Original Image

Segmented Image

Figure :Showing image segmentation [Moving toward region-based image segmentation techniques: a study by Dr.s.v.kasmir raja, Dr.s.s.riaz ahamed.]

MEDICAL IMAGE BASED RESEARCH:

In the year 1990, Lawrence M. Lifshitz et al.[4] presented a paper on amultiresolution hierarchical approach to Image Segmentation Based on Intensity Extrema for abdominal CT images. Lawrence & Stephen presented that Stack-based image segmentation correctly isolates anatomical structures in abdominal CT images. The aim of their research has been to create a computer algorithm to segment gray scale images into regions ofinterest (objects). These regions can provide the basis for scene analysis (including shape parameter calculation) or surface-based shaded graphics display. The algorithm creates a tree structure for image description by defining a linking relationship between pixels in successively blurred versions of the initial image. In the year 2000,Dzung, Chenyang & Jerry [5] presented critical appraisal of the current status of semi-automated and automated methods for the segmentation of anatomical medical images and the methods in medical image segmentation.

SOFTWARES USED IN IMAGE SEGMENTATION

As per [6], Several open source software packages are available for performing image segmentation.Open source software packages for performing image segmentation are listed below:

1. ITK - Insight Segmentation and Registration Toolkit (Open Source).
2. ITK-SNAP is a GUI tool that combines manual and semi-automatic segmentation with level sets.
3. GIMP which includes among other tools SIOX (Simple Interactive Object Extraction).
4. VXL is a computer vision library.
5. ImageMagick segments using the Fuzzy C-Means algorithm.
6. 3DSlicer includes automatic image segmentation.
7. MITK has a program module for manual segmentation of gray-scale images.
8. OpenCV is a computer vision library originally developed by Intel.
9. GRASS GIS has the program module i.smap for image segmentation.
10. Fiji - Fiji is just ImageJ, an image processing package which includes different segmentation plugins.
11. AForge.NET - an open source C# framework.

There is also software packages available free of charge for academic purposes:
1. GemIdent
2. CVIPtools
3. MegaWave

Open source software for medical image analysis[6]:

Software packages for performing analysis of medical images are listed below:
1. ImageJ
2. 3D Slicer
3. ITK
4. OsiriX
5. GemIdent
6. MicroDicom
7. FreeSurfer
8. ClearCanvas
9. Seg3D
10. NumPy + SciPy + MayaVi/Visvis
11. InVesalius

CONCLUSION

The major objective of this paper was to analyse the main directions of research in the field of colour image segmentation and to categorise the main approaches with respect to the colour image segmentation process. After evaluating a large number of papers, we identified three major trends in the development of colour image segmentation, namely algorithms based on feature integration, approaches that integrate the colour and texture attributes in succession and finally methods that extract the colour image features on independent channels and com- bine them using various integration schemes. As we discussed, the methods that fall in the latter categories proved to be more promising when viewed from algorithmic and practical perspectives.

REFERENCES

[1] V. Dey a , □, Y. Zhang a, M. Zhong "a review on image segmentation techniques with remote sensing perspective" Vol. XXXVIII, Part 7A

[2] Patel Janakkumar Baldevbhai1, R.S. Anand2 Review of Graph, Medical and Color Image base Segmentation **Techniques**

[3] M. Mirmehdi, M. Petrou, Segmentation of colour textures, IEEE Transactions on Pattern Analysis and Machine Intelligence 22(2)(2000) 142–159.

[4] Lifshitz, L. and Pizer, S., ―A multiresolution hierarchical approach to image segmentation based on intensity extrema,□ IEEE Transactions on Pattern Analysis and Machine Intelligence, 12:6, 529 - 540, 1990.

[5] Dzung L. Pham, Chenyang Xu, and Jerry L. Prince, ―Current Methods in Medical Image Segmentation□, Annual Review of Biomedical Engineering, volume 2, pp 315-337, 2000.

[6] www.wikipedia.org

[7] C. T. Zahn, "Graph-theoretical methods for detecting and describing clusters,□ IEEE Transactions on Computers, pp.68-86, Vol. 20, No. 1, 1971.

[8] Heng-Da Cheng and Ying Sun,□ A Hierarchical Approach to Colour Image Segmentation Using Homogeneity□, IEEE transactions on image processing, vol. 9, no. 12, December 2000.

Evolution of Image Capturing and Processing in Portables

Loveleen kaur
Department of Electronics and Communication, ACET, Amritsar

Abstract - **The primary application for solid state camera modules is in portable electronics, such as mobile phones, digital cameras, computer accessories, automotive driver aids and safety devices to surveillance cameras and toys. Typically more than two million CMOS imagers are manufactured every day of the year solely for just mobile phone application. The consumer demand for vast quantities of camera modules at very low price is forcing a search for new approach to camera module architecture, materials choices and assembly techniques. This paper will review the changing trends in camera module with particular emphasis on a new wafer-level chip size package that comfortably exceeds mobile phone and automotive reliability standards. The handset manufacturer is more interested in reduced cost and height, as thin, portable electronics products is currently the fashion. However, the existing evolutionary roadmap of smaller and cheaper camera modules, without degraded picture quality, has come to an end. Achieving a step-wise reduction in cost and height, while simultaneously boosting optical performance, requires a radical new approach.**

Index Terms — Camera Modules, Wafer Level, Assembly

I. Introduction

Solid-state imaging began with the invention of the charge coupled device (CCD) in 1969 at the Semiconductor Components Division of Bell Laboratories. The first portable digital still camera (DSC) using a CCD sensor was demonstrated by engineers at Eastman Kodak in 1985. An alternative imaging technology based on *Complementary Metaloxide Semiconductors* (CMOS) emerged in the 1970s out of work at the National Aeronautics and Space Administration Jet Propulsion Laboratory. This effort eventually led to the development of CMOS active-pixel sensors. CMOS imagers are able to furnish many on-chip functions, allowing for reduced size, lower power consumption and simplified assembly. Consequently, CMOS now dominates solid state imager technology except for niche applications where optical performance or imager resolution is paramount. Today, image sensor die are manufactured by many semiconductor companies. The smallest standardised area imager is the quarter common intermediate format (QCIF), with a resolution of 25,344 pixels, while the largest commercially available imager has 111M pixels. A camera module has three main components:

(i) a solid state image sensor that converts photons to electrons
(ii) a lens train that comprises all the optical parts of the camera
(iii) a housing that holds the lens train in the correct physical location with respect to the imager.

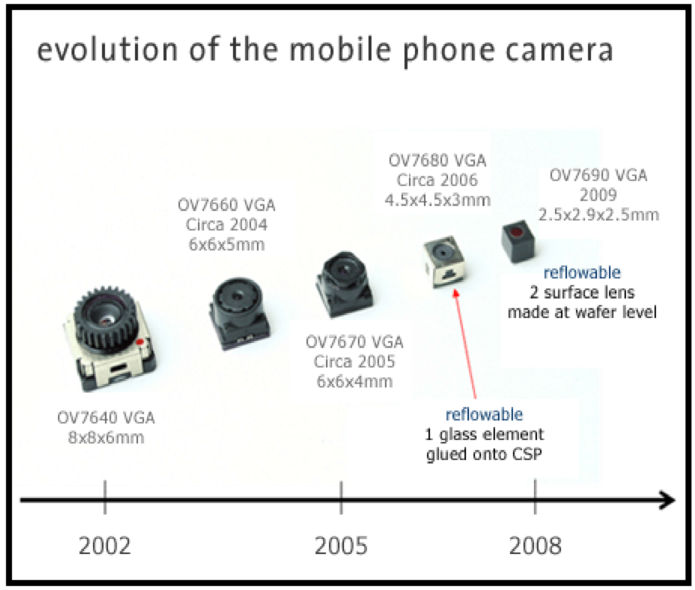


Figure 1 Evolution of Mobile phone camera

A cut-away drawing of a camera module is given in Figure 2. The imager die is mounted on a substrate and connected to it by wire bonds. Over the imager is lens train mounted in a turret. The turret is held in the correct location with respect to the imager by the lens housing, which attaches to the substrate. The optically sensitive area of the imager is covered with an array of micro lenses, one per pixel. The lens train also has three key components:

(i) The lenses that focus the scene onto the image
(ii) Several apertures
(iii) an infrared filter

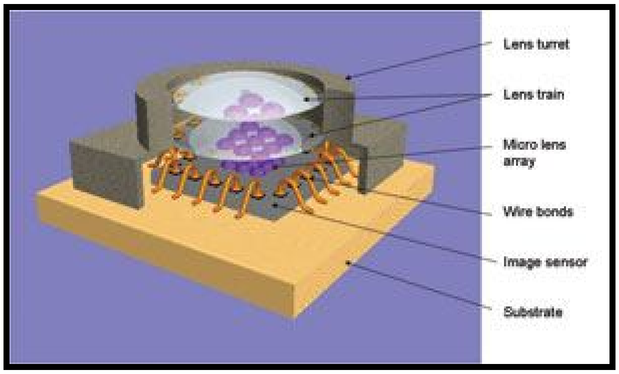


Figure 2:Cut away drawing of a solid state camera module

There can be as few as one lens on a QCIF camera module, or as many as four on one camera module with mega pixel resolution. Usually, the lenses are plastic for reasons of cost, but higher quality cameras sometimes use glass for the first lens because of its superior optical properties. The optical train inevitably contains at least one stop to limit the field of view and, frequently, several apertures to control optical aberrations and set the exposure speed (F-number). These are simply opaque films with a clear central region that are applied to lens surfaces. For normal applications, the lens train will always contain an infrared filter because silicon imagers are sensitive to longer wavelengths than the human eye can perceive.

The lens train is mounted in a sub-assembly called the lens turret that screws into the lens housing. The lens housing holds the lens train in approximately the right location with respect to the imager and adjustment of the screw thread of the lens turret is used to set the focus.

II. EMERGING REQUIREMENTS OF PORTABLE CAMERAS

Solid-state camera modules remained a specialist product until 2001, when a common intermediate format (CIF) camera debuted on a cell phone. Within eight years, the number of image sensors produced annually for hand-held platforms rose from thousands to approximately 100 crore (1 billion). In 2008, more than 80 per cent of all cell phones had at least one camera, many having two, with market saturation eventually occurring at around 150 crore (1.5 billion) units per annum (Techno Systems Research Co. Ltd, November 2007). Other applications that use cell phone cameras include low-end DSC, automotive driver aids, web cams and toys, which together could result in an additional 100 crore (1 billion) camera modules per annum by 2015.Consumers appreciate cell phones with cameras because the camera is integrated and hence they automatically carry it with them most of the time. However, cell phone cameras are unable to deliver the same quality of pictures as a DSC with the equivalent number of pixels. Consumers are increasingly recognising this performance gap and that raw pixel count does not relate to image quality. The trend seems to be that future cell phones will have a Video Graphics Array (VGA) resolution camera for video conferencing and a higher resolution camera in the 5 to 10 Mega pixel range for photography. Given the potential ability of a mega-pixel camera to obtain good quality images, consumers are demanding DSC-like quality and more DSC-like features on camera phones. Some of the most sought after characteristics include:

- Low light sensitivity, especially the ability to take photographs indoors without flash. Flash photography is very power hungry in a product where battery life is often a major consideration for the consumer. Flash photography can also introduce undesirable image artifacts, such as red-eye, that then need correction.
- Focus, image stabilisation and optical zoom. Clearly, photographs need to be in focus. Image stabilisation is, in many ways, analogous to focus because it is also image blurring, but in a lateral direction. Because the distance of the object to the camera varies, what is required is the means of either adjusting the focus to suit or extending the depth of field. Optical zoom allows the user to get closer to the subject of the photograph and details of the scene to be magnified. Digital zoom degrades image quality, so is undesirable if the captured quality is already low.
- Higher resolution. Although higher resolution does not directly translate into higher picture quality, acquisition of additional information by a higher resolution imager facilitates effective image enhancement by software.
- Size reduction. The height of camera modules is one factor limiting the thinness of cell phones where the current fashion is for extreme thinness. Camera modules are typically around 5mm high, but would ideally be less than 1.5mm tall.
- Cost reduction. Camera modules and the associated image processor are relatively expensive components and contribute to the overall handset price. This is especially true for cell phones that have two cameras, where the cost of the cameras is around 12 per cent of the total handset bill of materials. A long-term goal of the industry is a cheaper VGA camera module. The consumer demand for camera phones is high-quality image capture under a wide range of conditions, accomplished at a single button push. Unlike a DSC, the most desirable “feature” of a camera phone is fully automatic operation of the camera with no menus or settings to navigate. Delivering that in a highly compact package translates to system design challenges where cost is the ultimate arbitrator.

III. ECONOMICAL MODULE CONSTRUCTION

The fundamental cost-driver of a camera module is size. The cost of an imager is proportional to the square of the die diagonal and the cost of an optical train to the cube of its diameter. Decreasing the silicon die size boosts the number of die per wafer and hence reduces unit imager cost. However, decreasing the die size comes with the major penalty of an inevitable reduction in pixel size. This has the knock-on effect of decreasing quantum efficiency and exacerbating poor low light performance. The cost of optical components is largely determined by the volume of material used to manufacture lenses and the area of planar components like apertures and filters. To decrease cost, injection-moulded plastic lenses are widely used in camera modules. However, compared to glass lenses, plastic has inferior colour purity and transparency, particularly for short wavelength (blue) radiation. Because the area and height of a camera module are related, decreasing the area of the sensor and optics helps the trend towards thinner phones. The second major cost-driver of camera modules is the manufacturing process. Unlike conventional semiconductor components, the cost of assembling camera modules is not driven by the materials and process, but by yield. The principal cause of yield loss for camera modules is optical defects. Because the camera module cannot be tested for optical performance until the assembly is complete, a defective unit represents an appreciable cost penalty. Most optical defects are caused by particulate contamination of the imager.

The traditional way of assembling camera modules, using chip-on-board assembly, is prone to low yields because the imager die is completely exposed until the lens housing is attached and lens train finally set in place. It is therefore not surprising that the proportion of camera modules manufactured using chip-onboard assembly is rapidly waning and being replaced by wafer level packaging. It is estimated that by 2012 more than 65 per cent of imagers will be protected in wafer-level packages. The key to the success of wafer-level packaging of imagers is a glass wafer that is attached to the front face of the semiconductor wafer as the very first step of the camera module manufacturing process. Thereafter, the delicate optical area of each die is protected and any contamination that lands on that lands on the glass can be easily removed without damaging the imager.

The major cause of field returns of cell phones with defective cameras is failure of the flexible circuit and connector used to join the camera module to the main circuit board of the handset. These components are only necessary because the choice of materials traditionally used for the optical housing and lens train has made conventional camera modules incompatible with the thermal excursion involved in surface mount assembly; virtually all other components of cell phones are surface mounted.

Wafer-level packages are easily provided with a ball grid array interface, making them compatible with surface mount assembly. This has engendered a trend towards sourcing higher quality materials for the optical components to achieve a cost reduction through eliminating the flexible circuit and connector, and permitting the camera module to be surface mounted like most other components of the phone.

IV. MODULE HEIGHT REDUCTION

A wafer-level package for an imager is less than 500µm thick. The majority of camera module height is accounted for by the optical components. Reducing the number of lenses in the lens train has a dramatic influence on this important metric and also decreases cost. There is, however, a direct relationship between the number of optical surfaces and image quality. Camera module designers need to be aware of the opportunity to relax certain specifications to achieve the cost/height/image quality the customer requires. For example, the specifications of a VGA camera that is used as the main picture camera on a low-end phone are very different to those of a VGA camera that is used as the secondary, video conferencing camera on a high-end phone. The video camera does not need to resolve the same detail and hence can use a smaller imager with smaller and cheaper optics.

Because conventional camera modules cannot be reduced significantly in height or deliver higher quality pictures through engineering optimisation, alternative strategies must be used. The options are completely different hardware architectures for the camera module and/or the utilisation of non-linear optics implemented through a combination of hardware and software.

V. DRAWBACKS OF DISCRETE ASSEMBLY

Manufacture of a conventional camera module involves serial assembly from discrete parts. This has several fundamental limitations. First, the costs of assembly increase with every part manufactured. Once a high-volume line has been established, further gains in productivity can only be realised by diluting the administrative overhead for the facility by installing a parallel line with all the attendant capital costs involved.

Second, the precision of assembly increases with the quality of the image required. Image quality and imager resolution are linked, so manufacturing a megapixel camera module generally requires far greater mechanical precision than one of VGA resolution. The specifications of camera module assembly are already far more exacting than for conventional electronic components, necessitating more expensive machines, and generally compounded by slower throughput.

Third, each component must be manufacturable at the lowest possible cost. This means, for example, lenses must be radically symmetrical about the optical axis because this is the least expensive shape to produce and is compatible with using a screw thread mechanism to set camera focus. Imager die, however, always have a rectangular format, so there is clearly a mismatch between the aspect ratio of the lens train and the imager, which usually manifests as reduced image quality in the picture corners.

Finally, there are the consequences arising from the inability to make the product right the first time. While it is possible to set the optical axis of the lens train sufficiently lose to that of the imager so as to not require adjustment, the same is not true of its height. The multiple glue line thicknesses and other variables cannot be controlled with sufficient accuracy to fix the lens turret in position during manufacture. This means that every camera module must be placed in a test fixture, powered up and a series of images acquired while the focus is adjusted and then set. Clearly, this is a slow undertaking requiring a known good die tester equipped with an optical head, making final set-up of conventional camera modules a significant and unavoidable contributor to the manufacturing cost.

VI. WAFER LEVEL PACKAGING

Wafer-level packaging of imagers is not a new architecture for camera modules, but, as it will be explained, it continues to enable design advancements.

The original basis of the move to wafer-level packaging of imagers was the need to decrease cost. While the adoption of wafer-level packaging actually increases assembly costs, it helps eliminate the major cause of yield loss from camera module manufacturing—particle contamination of, or damage to, the optically sensitive area of the imager. The benefit comes from attaching a cover glass to the face of the image sensor wafer as the first process step. From that point on, the imager is totally protected from environmental and mechanical damage so the vast majority of good optical die on the wafer can be converted into yielded camera modules.

A wafer-level package can be provided with a ball grid array interface on its rear surface. This allows the imager to be placed, with other components, on a common printed circuit board and attached by a single reflow soldering cycle, further minimising assembly cost. The reliability of a ball grid array interface is also superior to the flexible circuit and connector of a conventional camera module built using chip-on-board assembly.

The wafer-level package must provide connectivity between the bond pads on the front face of the imager die and the ball grid array interface on the rear face of the package. There are a number of means by which this may be accomplished, and one has gained commercial acceptance in the form of a family of imager package solutions. More than 100 crore (1 billion) imagers have been included in these wafer-level packages since the technology was introduced in 2001.

Wafer-level packaging fails to be a low-cost solution if the materials and equipment set are derived from standard semiconductor practice. To achieve substantial cost reduction requires innovation in both of these areas. One approach that has proved highly successful is the adoption of mature materials produced in extremely high tonnage for an entirely different industry and purpose, and a tool set developed for the PCB industry. By this means, the packaging cost per die, including depreciation of the equipment, works out to a few cents per die, which is two orders of magnitude cheaper than discrete packaging.

The cost reduction achieved from wafer-level packaging of the imager is beneficial, but does nothing to address the cost of the optical train or alter the height of the camera module. That requires an alternative approach to discrete optical assembly.

Wafer-scale packaging of imagers is economically attractive because the fixed process costs are divided among the number of good parts on the wafer. The same argument holds for manufacturing lenses at the safer scale. By using several wafers of lenses, these can be easily and accurately stacked and bonded together. These stacks can then be singulated. The net result is an optical part with the same functionality as a conventional camera optic, but with greatly reduced costs and very precise and reproducible alignment. Switching to wafer-scale manufacturing techniques provides new freedoms in materials choice and lens shape. The materials used to make wafer-scale lenses an have a higher refractive index than injection-moulded lenses, whereby aiding size reduction. The materials chosen for the construction of the lenses can be selected for high temperature durability, facilitating surface mounting of the camera module and accompanying cost benefits. The precision of layer-to-layer and rotational alignment made possible by wafer-scale manufacturing means the lenses can have re-entrant profiles, diffractive surface features and be asymmetric in shape to match the rectangular format of the imager. These advances in optics permit the lens stack to be reduced in height without sacrificing optical performance.

VII. WAFER LEVEL CAMERA MODULES

While wafer-level packaging of imager die was originally developed as a means to decrease the cost of camera modules (by eliminating assembly yield loss) a highly fortuitous result is that the cover glass provides an exceptionally uniform surface, spaced an exact distance from the imager die. It therefore makes an ideal substrate or platform onto which a wafer-level optical stack can be attached. Owing to the high precision of the wafer-level imager package and also the wafer-level lens stack, they can be permanently joined without the need for costly, live adjustment of focus. The absence of moving parts and small size makes for an extremely robust product. Because the wafer-level stack has a considerably lower profile than the housing and lens turret it replaces, the resultant wafer-level camera module is appreciably smaller.

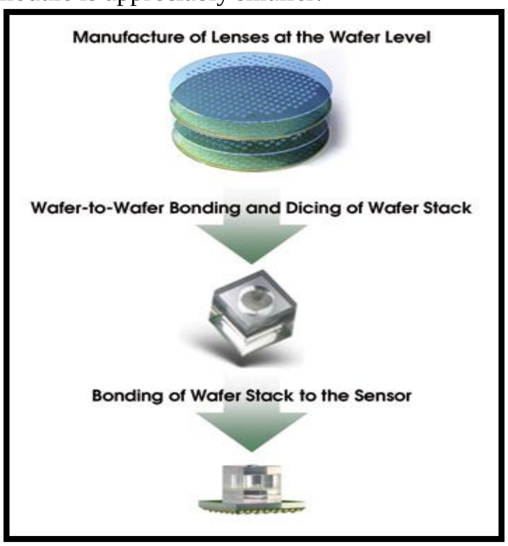


Figure 3:Wafer level package

VIII. CONCLUSION

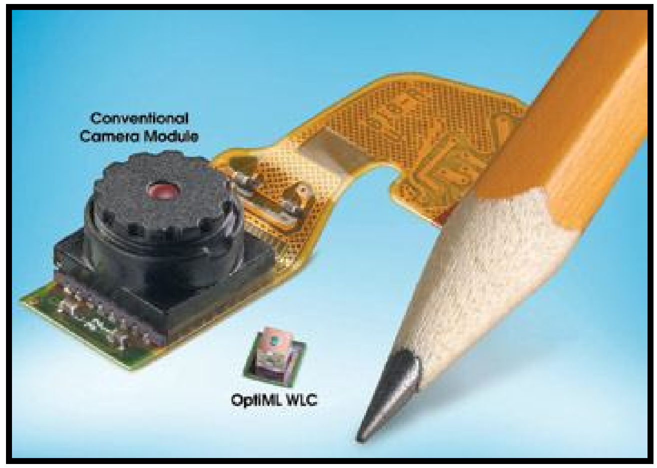


figure 4 : A wafer level camera module and conventional camera module of same resolution

Innovation in imager and optics wafer-level packaging has lead to the development of extremely compact camera modules that can be manufactured at low cost. The wafer-level camera is 30 to 50 per cent less expensive has a form factor 50 per cent smaller and is surface mountable. Currently the development in this arena is towards the *software enhanced optics* which has taken the image quality by portable cameras to a completely new horizon making it equivalent to DSC.

REFERENCES

[1] Giles Humpston Tessera: “Address image capture, processing challenges in portabls".

[2] I Hecht, B Kidron and M Kriman **“**Cost Reduction and Reliability Enhancement of Solid State Image Sensors by Wafer Level Chip Size Packaging Technology using New Materials **“.**

[3] Rafael C. González, Richard Eugene Woodsl, “digital image processing”, Prentice Hall.

[4] Giles Humpston, “Wafer-level Packaging of Image Sensors”.

[5] Nick Efford, “digital image processing a practical introduction using java” ,Pearson Education Ltd.

Emotion Detection from Speech

Gurpreet kaur
M.tech ECE student,Amritsar college of enginering & technology
Email:gurpreetkaurchohan@gmail.com

Abstract—**In emotion detection from speech,we selected pitch, log energy, formant, mel-band energies, and mel frequency cepstral coefficients(MFCCs) as the base features, and added velocity/acceleration of pitch and MFCCs to form feature streams. We extracted statistics used for discriminative classifiers, assuming that each stream is a one-dimensional signal. Extracted features were analyzed by using quadratic discriminant analysis (QDA) and support vector machine (SVM).**

Keywords— **Feature for speech, segmentation of speech signal.**

I. INTRODUCTION

Emotion detection from speech is relatively new field of research, it has many potential applications. In human-computer or human-human interaction system, it is an important factor in human communication. Emotion recognition systems could provide users with improved services by being adaptive to their emotion. In virtual worlds, emotion recognition could help simulate more realistic avatar interaction. Performance of emotion recognition largely depends on how can extract relevant features invariant to speaker, language, and contents. We decided not to use any linguistic knowledge in this work for generality. To reduce the variability in F0,format frequencies and duration due to speaker and phonetic dependency, it is crucial to normalize the feature. When a back-end classifier working with a fixed-length feature vector is used for classification, it is also important to convert variable length feature vectors into a representative fixed-length feature vector. Because the resulting feature vector usually has very large dimension, selecting good feature is needed to improve accuracy as well as to reduce te feature dimension.

In a parallel way, advances on the topic made it possible to successfully and accurately detect several basic emotion coming from these mentioned audio or video data sets. However, detecting any expressions of human affective behaviour in less constrained setting still proved to be a very challenging problem. The major problems regarding is to be attributed to the fact spontaneous affective behaviour differ in visual appearance, audio profile, and even timing; on controllable parameters that are eliminated with deliberate, acted, affective behaviour data sets.

An increasing number of studies regarding the automatic analysis of spontaneously displayed affective behaviour began to emerge,explaining how these kind of facial and vocal expression could be detected.It was also shown that integrating the information from audio video simultaneously leaded to an improved performance of affection recognition. Variation of head position,eye and brows movement,clutter or slightly appreciable face muscles movment proved that by making use of the complementary informati of these two channels-audio and video-would improve the reliability of affective behaviour recognition.

II. EMOTION RECOGNITION

Emotion recognition in this work has three stages: Feature extraction,feature selection and classification. Base features and statistics were computed in feature extraction. Feature components were analyzed in feature selection. Classification was made by using various classifiers based on dynamic models or discriminative models

A. *Feature extraction*

Figure 1 shows the block diagram of feature extraction. We selected the pitch, log energy, formant, band energies, and melfrequency cepstral coefficients (MFCCs) as the base features based on the previous study results and our preliminary results. The frame shift in feature extraction was 10ms. We first segmented only speech parts from an input utterance by using an endpoint detector based on zero crossing rate (ZCR) and frame energy. For each frame of speech signals, we estimated F0, log energy, three formant frequencies (F1, F2, F3), five mel-band energies, and two MFCCs [8]. The fundamental frequency was estimated by finding the time shift that minimizes the average mean difference function (AMDF). The formant frequencies were estimated by finding the poles of the autoregressive transfer function obtained from the linear predictive coding (LPC) coefficient.

We also added velocity and acceleration information for pitch and MFCCs, respectively, to take the rate of speaking into account and model the dynamics of the corresponding temporal change of pitch and spectrum. Hence we have 15 streams of features including velocity and acceleration components. These streams were used as input vectors of HMM-based classifiers.

For discriminative classifiers, the feature streams were converted into a fixed-length vector for each utterance by computing statistics to represent the streams. Note that conversion to a fixed-length vector was performed only for discriminative classifiers such as SVM, LDA and QDA.

To obtain a fixed-length vector, we computed 11 statistics from pitch, 7 from energy, 5 for each formant frequency,4 for each MFCC, and 1 for mel-band energies. The statistic sincluded the 90th percentile, range, mean, standard deviation, skewness, and so on. For pitch we added the pitch of the first frame, the pitch of the last frame, the mean and theregression coefficient of the first/last segments. Regarding log energy, we subtracted the mean log energy to normalize amplitudechange according to the speaker volume or the distance of a speaker from a microphone. Two duration-dependent componentswere added.

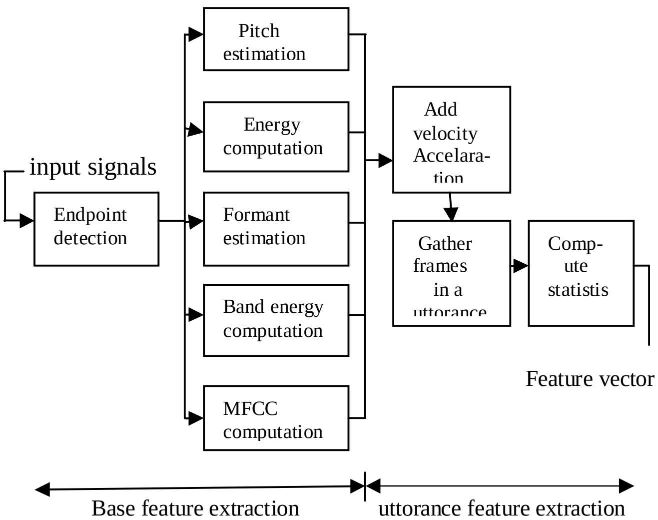


Figure 1: Block diagram of the feature extraction module.

One is the mean MFCC distance between adjacent frames and the other is the duration in frames divided by the mean MFCC distance. The dimension of one feature vector, the input vector of the feature selection, was 59 for each utterance. The pitch and formants in silence regions were interpolated from adjacent frames so that there are no discontinuities on the contours of the pitch and formants. To obtain the relevant quantities, we used only the voiced region of the input utterance.

B. Pitch estimation

Bäzinger et al. argued that statistics related to pitch conveys considerable information about emotional status. Yu et al. have shown that some statistics of the pitch carries information about emotion in Mandarin speech. For this project, pitch is extracted from the speech waveform using a modified version of the RAPT algorithm for pitch tracking implemented in the VOICEBOX toolbox. Using a frame length of 50ms, the pitch for each frame was calculated and placed in a vector to correspond to that frame. If the speech is unvoiced the corresponding marker in the pitch vector was set to zero.

C. MFCC and related feature

MFCCs are the most widely used spectral representation of speech in many applications, including speech and speaker recognition. Kim et al. argued that statistics relating to MFCCs also carry emotional information. For each 25ms frame of speech, thirteen standard MFCC parameters are calculated by taking the absolute value of the STFT, warping it to a Mel frequency scale, taking the DCT of the log-Melspectrum and returning the first 13 components .
The variation in three MFCCs for a female speaker uttering “Seventy one” inemotional states of despair and elation. It is evident from this figure that the mean of the first coefficient is higher when “Seventy one” is uttered in elation rather than despair, but is lower for the second and third coefficients. In order to capture these and other characteristics, we extracted statistics based on the MFCCs. For each coefficient and its derivative we calculated the mean, variance, maximum and minimum across all frames. We also calculate the mean, variance, maximum and minimum of the mean of each coefficient and its derivative. Each MFCC feature vector is 112-dimensional.

D. Formants estimation

Tracking formants over time is used to model the change in the vocal tract shape. The use of Linear Predictive Coding (LPC) to model formants is widely used in speech synthesis. Prior work done by Petrushin suggests that formants carry information about emotional content. The first three formants and their bandwidths were estimated using LPC on 15ms frames of speech. For each of the three formants, their derivatives and bandwidths, we calculated the mean, variance, maximum and minimum across all frames. We also calculate the mean, variance, maximum and minimum of the mean of each formant frequency, its derivative and bandwidth. The formant feature vector is 48-dimensional

E. Feature selection

Out of the many derived features given to the classifiers, we want to identify those that contribute more in the classification.This tells us what features and properties of the speech are important in distinguishing emotions. We could then derive more relevant features accordingly to improve classification accuracy. However, it is forbiddingly time consuming to perform exhaustive search for the subset of features that give best classification. Instead, we used forward selection and backward elimination to rank the features and identify the subset that contributes more in classification. Forward selection sequentially adds one featureat a time,choosing the next one that most increases or least decreases classification accuracy. Backward elimination startswith all input features and sequentially deletes the next feature that most decreases or least increases classification accuracy.

III. FEATURES FOR SPEECH EMOTION RECOGNITION

An important issue in the design of a speech emotion recognition system is the extraction of suitable features that efficiently characterize different emotions.Since pattern recognition techniques are rarely independent of the problem domain , it is believed that aproper selection of features significantly affects the classification performance. Four issues must be considered in feature extraction.The first issue is the region of analys isused for feature extraction.While some researchers follow the ordinary frame work of dividing the speech signal in to small intervals,called frames, from eachwhicha local feature vector is extracted, other researchers preferto extract global statics from the whole speech utterance.Another important question is what the best feature types for this taskare,e.g.pitch, energy, zero crossing, etc.? A third question is what is the effect of ordinary speech processing such a spost-filtering an dsilence removal on the overall performance of the classifier ? Finally, whether it suffices to use acoustic features for modeling emotions or if it is necessary to combine them with other types of features such a slinguistic, discourse information,or facial features. The above issues are discussed in detail in the following sub sections. In Section 3.1,a comparison between local features and global features is given.Section 3.2 describes different types of speech features used in speech emotion recognition.This subsection is concluded with our recommendations for the choice of speech features.

A. Local features versus global feature

Since speech signals are not stationary even in wide sense, it is common in speech processing to divide a speech signal into small segments called frames. Within each frame the signal is considered to be approximately stationary. Prosodic speech features such as pitch and energy are extracted from each frame and called local features. On the other hand, global features are calculated as statistics of all speech features extracted from an utterance. There has been a disagreement on which of local and global features are more suitable for speech emotion recognition. Global features have another advantage over local features; their number is much less. Therefore, the application of cross validation and feature selection algorithms to global features are executed much faster than if applied to local features. However, researchers have claimed that global features are efficient only in distinguishing between high-arousal emotions, e.g. anger, fear, and joy, versus low-arousal ones, e.g. sadness. They claim that global features fail to classify emotions which have similar arousal, e.g. Anger versus Joy. Another disadvantage of global features is that temporal information present in speech signals is completely lost. Moreover, it may be unreliable to use complex classifiers such as the hidden Markov model (HMM) and the support vector machine (SVM) with global speech features since the number of training vectors may not be sufficient for reliably estimating model parameters. On the other hand, complex classifiers can be trained reliably using the large number of local feature vectors and hence their parameters will be accurately estimated. This may lead to higher classification accuracy than that achieved if global features are used.

B. Categories of speech features

An important issue in speech emotion recognition is the extraction of speech features that efficiently characterize the emotional content of speechand at the sametime donot depend onthe speaker or the lexical content. Although many speech features have been explored in speech emotion recognition, researchers have not identified the best speech features for this task. Speech features can be grouped into four categories: continuous features, qualitative features, spectral features, and TEO (Teager energy operator)-based features. The main purpose of this section is to compare the pros and cons of each category. However, it is common in speech emotion recognition to combine features that belong to different categories to represent the speech signal.

IV. CLASSIFICATION

We compared the performance of classifiers based on discriminative and generative models by using SVM, LDA, QDA and HMM. Hsu and Lin compared various methods proposed to extend the binary SVM to multi-class and found that the “one against- one” scheme is more suitable for practical use. We used their scheme for the multi-class problem. For binary classification, we used our MATLAB implementation of the sequential minimal optimization SVM. The HMM-based classifier has the advantages over other static discriminative classifiers that frame length normalization is not necessary and temporal dynamics of the base features can be reflected by using the state transition probability. Short-time temporal dynamics is implicitly modeled through the addition of velocity and acceleration components. However, the HMM classifier is still weak at modeling long-time temporal dynamics. There are many other classifiers that have been applied in many other studies to the problem of speech emotion recognition such as k-NN classifiers , fuzzy classifiers, and decision trees. However, the above-mentioned classifiers, especially the GMM and the HMM, are the most used ones on this task. Moreover, the performance of many of them is not significantly different from the above mentioned classification techniques. One might conclude that the GMM achieves the best compromise between the classification performance and the computational requirements required for training and testing. However, we should be cautious that different emotional corpora with different emotion inventories were used in those individual studies. Moreover, some of those corpora are locally recorded and inaccessible to other researchers. Therefore, such a conclusion cannot be established without performing more comprehensive experiments that employ many accessible corpora for comparing the performance of different classifiers.

V. CONCLUSIONS

In this paper, a survey of current research work in speech emotion recognition system has been given. Three important issues have been studied: the features used to characterize different emotions, the classification techniques used in previous research, and the important design criteria of emotional speech databases. There are several conclusions that can be drawn from this study. The first one is that while high classification accuracies have been obtained for classification between high-arousal and low- arousal emotions, N-way classification is still challenging. More- over, the performance of current stress detectors still needs significant improvement. The average classification accuracy of speaker-independent speech emotion recognition systems is less than 80% in most of the proposed techniques. In some cases, such as , it is as low as 50%. For speaker-dependent classification, the recognition accuracy exceeded 90% only in few studies. Many classifiers have been tried for speech emotion recognition such as the HMM, the GMM, the ANN, and the SVM. However, it is hard to decide which classifier performs best for this task because different emotional corpora with different experimental setups were applied. Most of the current body of research focuses on studying many speech features and their relations to the emotional content of the speech utterance. New features have also been developed such as the TEO-based features. There are also attempts to employ different feature selection techniques in order to find the best features for this task. However, the conclusions obtained from different studies are not consistent. The main reason may be attributed to the fact that only one emotional speech database is investigated in each study. Using statistics computed from the base features, we analyzed the effects of the features in emotion recognition. The pitch and energy were shown to play a major role in recognizing emotion, which matches insights. We performed classification using SVM, LDA, QDA and HMM with the SUSAS and AIBO databases. Both SVM and HMM classifiers yielded classification accuracy significantly better than the previous results in the SUSAS database. For the AIBO speech database, we evaluated classification accuracy for 5 emotion classes. Further study is needed to explore new features better representing prosody and timbre, improve the pitch and formant tracking algorithms, and develop a new sophisticated approach to model dynamics of feature streams.

REFERENCES

[1.] poseidon.csd.auth.gr/papers/PUBLISHED/JOURNAL/pdf/Ververidis06a.pdf

[2.] C. Lee, S. Yildrim, M. Bulut, A. Kazemzadeh, C. Busso, Z. Deng, S. Lee, S. Narayanan, Emotion recognition based on phoneme classes, in: Proceedings of ICSLP, 2004, pp. 2193–2196.

[3.] J. Nicholson, K. Takahashi, R. Nakatsu, Emotion recognition in speech using neural networks, Neural Comput. Appl. 9 (2000) 290–296.

[4.] O. Pierre-Yves, The production and recognition of emotions in speech: features and algorithms, Int. J. Human–Computer Stud. 59 (2003) 157–183.

[5.] L. Rabiner, R. Schafer, Digital Processing of Speech Signals, first ed., Pearson Education, 1978.

[6.] R.W. Picard, E. Vyzas, J. Healey, Toward machine emotional intelligence: analysis of affective physiological state, IEEE Trans. Pattern Anal. Mach. Intell. 23 (10) (2001) 1175–1191.

[7.] T.L. Pao, Y.-T. Chen, J.-H. Yeh, W.-Y. Liao, Combining acoustic features for improved emotion recognition in Mandarin speech, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 3784, 2005, pp. 279–285, cited by (since 1996) 1.\

[8.] V. Hozjan, Z. Kacic, Context-independent multilingual emotion recognition from speech signal, Int. J. Speech Technol. 6 (2003) 311–320.

[9.] B. Schuller, G. Rigoll, M. Lang, Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture, in: Proceedings of the ICASSP 2004, vol. 1, 2004, pp. 577–580.

[10.] M.T. Shami, M.S. Kamel, Segment-based approach to the recognition of emotions in speech, in: IEEE International Conference on Multimedia and Expo, 2005. ICME 2005, 2005, 4pp.

[11.] D. Ververidis, C. Kotropoulos, Emotional speech classification using Gaussian mixture models and the sequential floating forward selection algorithm, in: IEEE International Conference on Multimedia and Expo, 2005. ICME 2005, July 2005, pp. 1500–1503.

[12.] C. Lee, S. Narayanan, Toward detecting emotions in spoken dialogs, IEEE Trans. Speech Audio Process. 13 (2) (2005) 293–303.

Decoding Of Individuated Finger Movements Using Variance-Covariance analysis For Surface Electromyogram Signals

Adhiti Gupta
Deptt. Of Electrical Engg.Baba Bandha Bahadur Engg.College Fatehgarh Sahib.

E-mail: adhitimbs@gmail.com

Abstract— **In this paper study on the anatomy of muscles of the forearm has been documented in the field of finger movements. The fingers of the hand were movable in four direction flexion (bending), extension (straightening) and abduction (moving sideways from the body), and adduction (moving sideways towards the body).**

The Surface Electromyogram (SEMG) signal was a biomedical signal that measures electrical currents generated in muscles during its contraction representing neuromuscular activities. In the present research the SEMG signals were recorded non-invasively by placing surface electrodes at the forearm muscles (flexor digitorum superficialis and extensor digitorum superficialis) and for two operations (back side near elbow and front mid side) of forearm, for two channels, for the different movements of finger. The SEMG signals were acquired from the forearm muscles required advanced methods for detection, decomposition, processing and analysis. For the acquisition the SEMG amplifier was built consisted of differential amplifier and non-inverting amplifier. The bipolar electrode was designed that was connected to the SEMG amplifier was placed on the forearm. The signals were carried out by the electrodes and transmitted to the amplifier. Both amplifiers were designed by using LM-324 chip. The acquisition circuit was built in Lab View. The SEMG signals were acquired from SEMG amplifier and then signal was transmitted to the interfacing device cRIO (Compact reconfigure input/output) and at last the interfacing device sends SEMG signals to the Lab View acquisition circuit. After acquiring data from six subjects analysis was done on the best selected 1000 SEMG samples and then the parameters (RMS, Standard deviation, Variance, Min and Max) were calculated. Further analysis was made on Variance-Covariance matrix for the different movements of fingers for two channels.

Keywords- Surface Electromyogram (SEMG), muscles, LabView,NeuralNetworkclassifier(NN-Classifier),cRIO(Compact reconfigure Input/output).

I. INTRODUCTION

The human hand finger movements are mostly determined by musculator activity in the forearm [1]. The finger movement is carried out by several groups of muscles. The muscles that flex the fingers primarily flexor digitorum superficialis and flexor digitorum profundus are located in the palmar aspect of the forearm. The addition of finger movement would enhance the quality of life of physically disabled people in order for them to participate in the cultural events such as, playing sports or musical instruments, finger motion are essential [2] .

The SEMG signal is a biomedical signal that measures electrical currents generated in muscles during its contraction representing neuromuscular activities. The nervous system also controls the muscle activity (contraction /relaxation).This signal is normally a function of time and is describable in terms of its amplitude, frequency and phase. The main reason of the interest in SEMG signal analysis is in biomedical applications [3].

The recent researches that have been used for the analysis of finger movements are SEMG based controllers, matlab, Auto regression and wavelet and pattern recognition [10]. Analysis for SEMG signal can be done by using the statistical approaches. In this research paper we study the finger movements using labview developed acquisition system using cRIO. In the end the analysis was done with the help of variance-covariance matrix.

II. ANATOMICAL STUDY FOR FOREARM AND FINGER MOVEMENTS

The forearm muscles are functionally divided into two equal groups: those causing wrist movements and those moving the finger and thumb. The fleshy portions contribute to the roundness of the proximal forearm and then they taper to the long insertion.

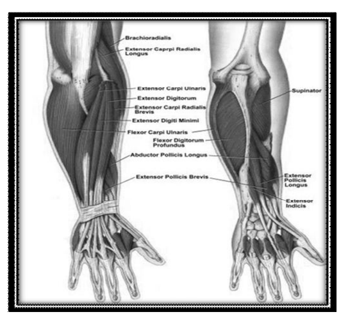


Figure 1. Muscles of forearm [9]

These insertions are securely anchored by ligaments called flexor and extensor retinacula. Forearm muscles arise from the humerus; their actions on the elbow are slight. Flexion and Extension are movements typically effected at both the wrist and finger joints.

Location of Electrodes

The electrodes placement determines the electrical view of a muscle, and then it is important in SEMG measurements, to be consistent in the placement of the electrodes for a subject over consecutive recording sessions.

Electrodes placement for finger movements are Flexion (bending), and Extension (straightening).For flexion movement the electrodes are placed on flexor digitorum superficialis and flexor digitorum profundus. For extension movement the electrodes are placed on the extensor digitorum.

There are three kinds of muscles that generated from the arm participate in flexure of fingers. They are flexure digitorum superficialis muscle, flexure digitorum profundus muscle and flexure pollicis longus muscle. The flexure pollicis longus muscle participates in flexure of the thumb, where as flexure digitorum superficialis muscle and flexure digitorum profundus muscles participate in flexure of the index finger, middle finger, ring finger and small finger. These various muscle movements are shown below:

TABLE 1. FUNCTION OF VARIOUS MUSCLES

S.No	Muscle	Function
1.	*Flexor pollicis longus*	*Flexion of the thumb*
2.	*Flexor Digitorum superficial*	*Flexion of the index gives the highest response*
3.	*Flexorcarpiulnaris*	*Flexion of pinkie*
4.	*Flexorcarpiradialis, Palmaris longus*	*Flexion of the middle and the ring*
5.	*Extensorpollicis longus*	*Extension of the thumb*

The movements that are detected by various researchers for electrode placements for SEMG signals are shown below:

TABLE 2. DIFFERENT LOCATIONS OF SEMG ELECTRODE PLACEMENT

S No	Researchers	Location of SEMG	Movement detected
1.	*Uchida et al.[4]*	*Flexordigitorum superficialis andFlexordigitorumprofundus*	*Flexion of all fingers, and relaxation of all fingers.*
2.	*ChristianAntfolk and Fredrik sebelius[5]*	*Superficial flexor and Extensor*	*Flexion and extension of each individual finger and thumb adduction/abduction and a rest class*
3.	*FrancescoTenore andAnderRamos [6]*	*Olecranon and clavide*	*The movement consisted of all finger individually and of the middle, ring and little finger.*
4.	*Jun-uk Chu, In Hyuk Moon,Shin-Kikim, and Mu-Seong Mun[7]*	*Extensor carpi radialis,Extensordigitorum,Flexor carpi ulnaris, Palmaris longus*	*Flexion and Extension ,radial and ulnar flexor of wrist, pronation and supination of wrist open and grasps of the finger and relaxation.*

In this study we have selected two sides for two channels because many have used and anatomically analyse the data. The selected movements will be hand completely at rest, hand closing, thumb, index finger, middle finger, ring finger and little finger because day to day these activities varies.

III. METHODOLOGY

There were three male and three female subject aged from 23 to 27 and in a healthy condition participated in the experiment**.** In the experiment, the SEMG signal was recorded by three pairs of self designed electrodes for two channels.

For channel 1 the electrode was placed on the extensor digitorum superficialis muscles on the back side of the forearm for acquiring the signals. For channel 2 the electrodes were placed on the flexor digitorum superficialis muscles of the forearm on the front mid side. The signal was acquired by placing electrodes on the muscles sites.

The signal that was acquired from the forearm muscles was in uv (microvolt) to convert that signal to mv (milivolt) the amplifier and filter circuit was used. In the experiment the differential mode operational amplifier with LM324 having a gain of 1000 was used. The signal was again amplified by non-inverting amplifier with a gain of 10.A real-time SEMG acquisition system was also developed.

For analysis and capturing PC-based system was used, in computer the acquisition circuit was build in Lab

View. The acquisition circuit consists of various modules. The A10 module and A11 module was shared variable node, which acts as an interface between a Lab View acquisition circuit and cRIO (Compaq reconfigure Input/output). The A10 and A11 modules drag from the cRIO chassis to input the raw signal to PC. The A10 and A11 modules were further connected to the numeric indicator and waveform chart. The wait express waits the specified count time (5miliseconds).The numeric number 5000 represent the number of sample for which SEMG signals was recorded. i-represents the iteration. The comparator was used to record the 5000 samples then wait for 5milisecond and then repeated the process. For repeating process, while loop was used. The statistics block provides the statistical calculation of selected parameter. The statistics block was connected to the filter, having a band pass frequency of 110-240 Hz. The filtered signal was further connected to the spectral measurement block. This block provides the FFT-based spectral measurements, such as the averaged magnitude spectrum and phase spectrum on a signal.

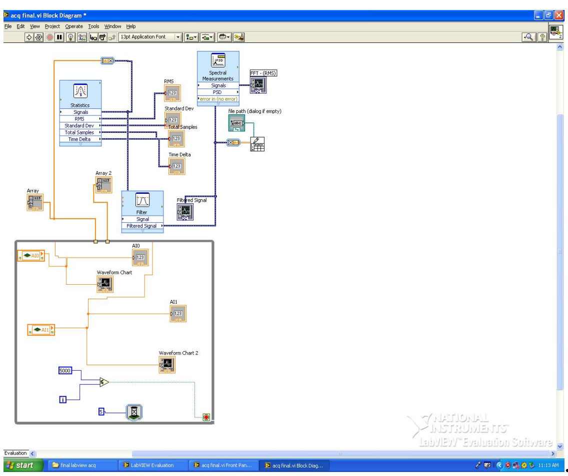


Figure 2. Block diagram of acquisition circuit

The SEMG signals were acquired by placing the electrodes on the motor unit action potential of the forearm muscles during different movements of fingers for two channels. The acquired signals was amplified with a hardware module consisted of amplifier and pre-amplifier circuit. The amplified SEMG signal was send to the cRIO which acts as an interfacing device between PC and hardware module. The interfacing devices send signals to the Lab View acquisition system. The Lab View acquisition system provides the SEMG signals readings in terms of time and amplitude. The data for SEMG signals were recorded and store for each subject in the excel sheet. The parameters (root mean square, standard deviation, variance, min, max, average) were calculated by applying formulas in excel.

The Parameter for the analysis of best SEMG signals for 1000 sample were Root mean square, Standard deviation, Variance, Min, Max. The formulas that were used to calculate the selected parameters in the excel were:-

RMS=SQRT(SUMSQ(A1:A1000)/COUNTA (A1:A1000))
Standard Deviation type =STDEVA (A1:A1000)
Variance =VAR (A1:A1000)
MIN (A1:A1000)
MAX (A1:A1000)
AVERAGE (A1:A1000)

The desired outcome of the study was to analyze the different operation of fingers for SEMG signals. For the analysis the variance-covariance matrix was used.

IV. Results and Discussion

The analysis for the interpretation of SEMG signals for finger movements was done with the help of variance-covariance matrix. The variance-covariance matrix was generated for raw data that was calculated with the help of matlab.The equation that were used for calculating the variance-covariance matrixes are:-

a= raw data;
One=ones (n, n);
a1=a-((one*a).*(1/n));
v= (a1'*a).*(1/n);

Where one is an n*n column vectors of ones.
a1 is an n*n matrix of deviations.
a is an n*n matrix of raw data.
a1’*a is the deviation sums of squares and cross product matrix.
n is the number of data in each column of the original matrix.
v is n*n variance-covariance matrix.

The above interpreted data show the calculation of variance-covariance matrix in statics v, which analyzed the various operations of finger.

In the Figure 4 the variance of SEMG signals was analyzed from variance-covariance matrix for six subjects for channel 1 and 2 were represented in below bars. The channel 1 were represented by blue bar, for subjects the blue bar indicates the maximum variance in case of open hand operation and minimum variance in case of closed hand operation. The channel 2 were represented by red bar, for each subjects the variance was maximum in case of closed hand operation and minimum variance in case of open hand operation.

Figure 4 (a) Variance of SEMG signals for channel 1and 2 for subject 1

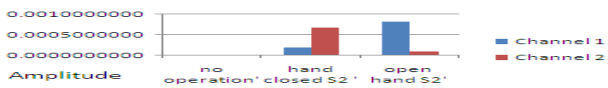


Figure 4 (b) Variance of SEMG signals for channel 1and 2 for subject 2

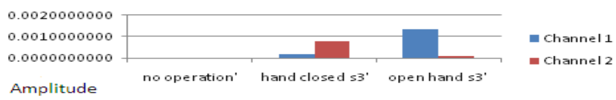


Figure 4 (c) Variance of SEMG signals for channel 1and 2 for subject 3

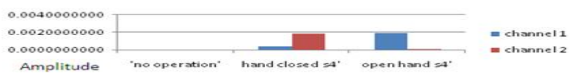


Figure 4 (d) Variance of SEMG signals for channel 1and 2 for subject 4

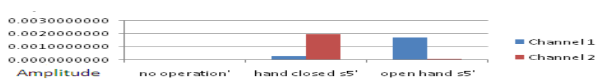


Figure 4 (e) Variance of SEMG signals for channel 1and 2 for subject 5

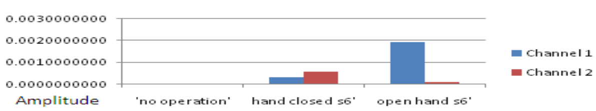
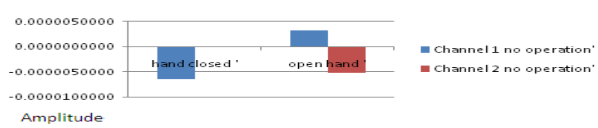


Figure 4 (f) Variance of SEMG signals for channel 1and 2 for subject 6



In the Figure 6 the variance of SEMG signals was maximum in channel 1 as compared to the rest of fingers for six subjects. The comparison of variance was done with open hand operation with respect to thumb, index finger, middle finger, ring finger, little finger. The middle finger has maximum variance for channel 1.

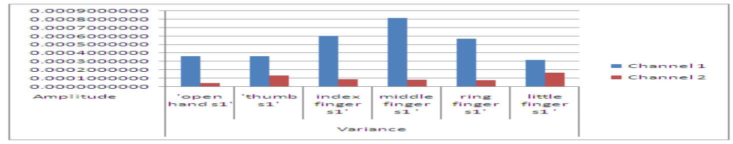


Figure 4.5(a) Variance of SEMG Signals for different operation of Fingers in Subject 1.

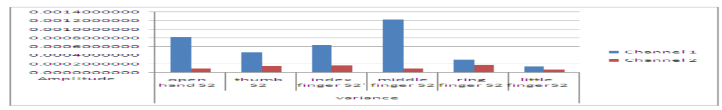


Figure 4.5(b) Variance of SEMG Signals for Different Operations of Fingers in Subject 2

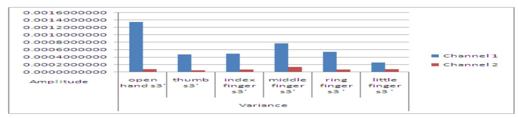


Figure 4.5(c) Variance of SEMG Signals for Different Operations of Fingers in Subject 3

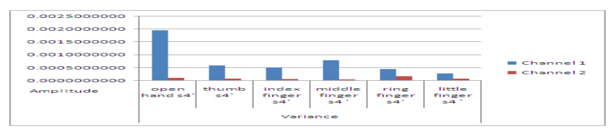


Figure 4.5(d) Variance of SEMG Signals for Different Operations of Fingers in Subject 4

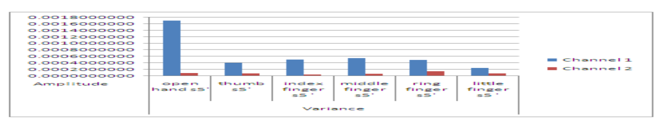


Figure 4.5(d) Variance of SEMG Signals for Different Operations of fingers in subject 5

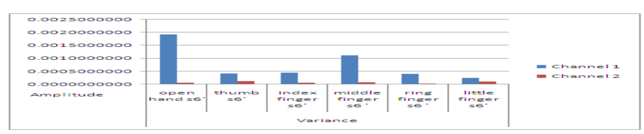


Figure 4.5(f) Variance of SEMG Signals for Different Operations of Fingers in Subject 6

In the Figure 4.6 the covariance in opening of hand with thumb, index finger, middle finger, ring finger and little finger operation was represented for channel 1 and 2 for six subjects. In channel 1 the covariance in middle finger was maximum and minimum in little finger. The covariance in thumb, index finger and ring finger varies but less varied as compared to middle finger. In channel 2 there could be less variation of covariance in opening of hand with thumb, index finger, middle finger, ring finger and little finger as compared to channel 1.

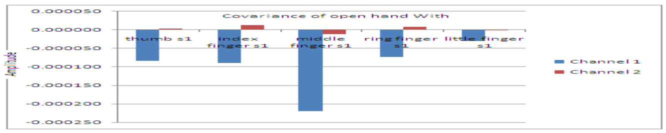


Figure 4.6(a): Covariance of Open Hand with all Fingers for Subject 1

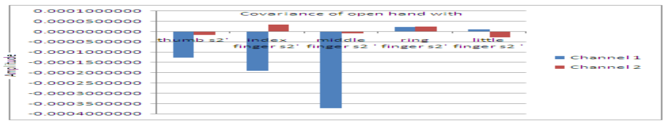


Figure 4.6(b): Covariance of Open Hand with all Fingers for Subject 2

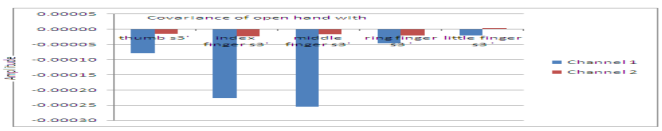


Figure 4.6(c): Covariance of Open Hand with all Fingers for Subject 3

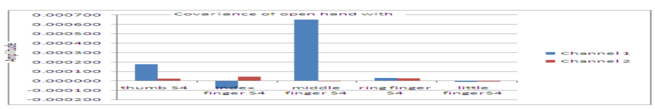


Figure 4.6(d): Covariance of Open Hand with all Fingers for Subject 4

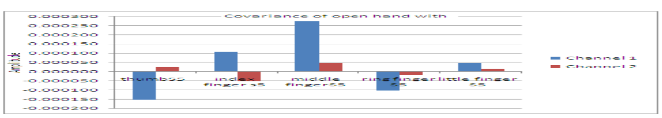


Figure 4.6(e): Covariance of Open Hand with all Fingers for Subject 5

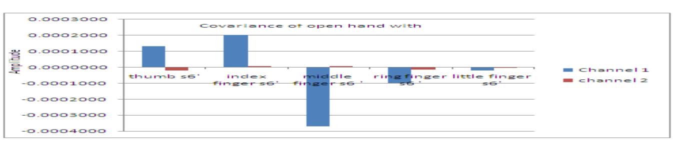


Figure 4.6(f): Covariance of Open Hand with all Fingers for Subject 6

CONCLUSION

SEMG was the more common method of measurement, since it was non-invasive and could be conducted by personnel other than medical doctors, with minimal risk to the subject. For forearm limbs, there were many muscles to control its movements. It was important to find out functional muscles. In this research, the SEMG were analyzed different movement of fingers during human forearm movement (flexion and extension).

The desired outcome of SEMG signals were recorded by placing electrodes on the extensor digitorum superficialis on the back side near elbow of the forearm for channel 1 operation and flexor digitorum superficialis on the mid front side near elbow of the forearm for channel 2 operation for different movements of finger. The analysis of SEMG signals were made in term of selected parameters (RMS, Standard deviation, Variance, min, max) and also with the variance-covariance matrix

The forearm and fingers were the important part of body. Forearm and finger amputation happened very often, so this study reveals that variance gave the best analysis among the selected parameters (Root mean square, Standard deviation, Min, Max.The selected movements i.e. hand completely at rest, hand closed, hand open, thumb, Index finger, middle finger, ring finger and little finger were statistically analyzed, thus being the prominent. In future studies the work must be extended for multisite-multimovement analysis. Further this study could be extended to SEMG frequency based approach along with the root mean square value for better analyses and classification.

ACKNOWLEDGMENT

The author would like to thanks the guide for their support in learning and explaining the Lab View functionality. The author would also like to thanks the faculty of the college for their guidance and support.

REFERENCES

[1] JV. Basmajian, CJ. de Luca, “Muscles Alive – The functions revealed by electromyography”.The Williams & Wilkins Company; Baltimore,1985.

[2] R. Kato, H. Yokoi, T. Arai, “Competitive learning method for robust EMG-to-motion classifier”, Intelligent Autonomous Systems, vol. 9, IOS Press, pp.956-953, 2005.

[3] Shahid S,“Higher Order Statistics Techniques Applied to EMG Signal Analysis and Characterization,”Ph.D. thesis, University of Limerick, Ireland, 2004.

[4] A. Andrews, “Finger Movement classification Using Forearm EMG Signals,”M.Sc. dissertation, Queen’s University, Kingston, On, Canada, 2008.

[5] C. Cipriani, F. Zaccone, S. Micera and M. C. Carrozza, "On the shared control of an EMG-controlled prosthetic hand: Analysis of user-prosthesis interaction," IEEE Trans. Robot.,vol.24,pp.170-184, 2008.

[6] F. V. G. Tenore, A. Ramos, A. Fahmy, S. Acharya, R. Etienne-Cummings and N. V. Thakor, "Decoding of individuated finger movements using surface electromyography," IEEE Trans. Biomed. Eng.,vol. 56,pp.1427-1434, 2009.

[7] C. Jun-Uk, M. Inhyuk, L. Yun-Jung, K. Shin-Ki and M. Mu-Seong, "A supervised feature-projection-based real-time EMG pattern recognition for multifunction myoelectric hand control," IEEE/ASME Trans. Mechatron., vol.12,pp.282-290, 2007.

[8] Jones,F.W., "The Principles of Anatomy as Seen in the Hand” , 2nd ed., Baillier, Tindall, and Cox,London,readable classic, 1942.

[9] Elaine N. Marieb, Katja Hoehn, “Science Human Anatomy and Physiology,” Google Book Result, June 2010, pp. 357–60.

[10] LanlanYu, Tianxing Meng, Jian Hu, “The finger Movement Identification Based on Fuzzy Clustering and BP Neural Network Performance,” 24-26 July 2009.

[11] Heather M. Seifert1 and Andrew J. Fuglevand, “Restoration of Movement Using Functional Electrical Stimulation and Bayes

A Review on Different Categories of Image Segmention

Sukhchain Singh, Kawaldeep Singh
M.Tech, ACET, Amritsar
Email: er.sukhchain87@gmail.com, deep_kawal17@yahoo.com

***Abstract*—Today with the evolution of research on the image segmentation it is important to categorize the image using image segmentation techniques. In this paper, different image segmentation techniques applied on optical remote sensing images are reviewed. The selection of papers include sources from image processing journals, conferences, books, dissertations and thesis out of more than 3000 journals, books and online research databases available at UNB. The conceptual details of the techniques are explained and mathematical details are avoided for implicitly. Both broad and detailed categorizations of reviewed segmentation techniques are provided. The state of art research on each category is provided with emphasis on developed technologies and image properties used by them. The categories defined are not always mutually independent. Hence, their interrelationships are also stated. Finally, conclusions are drawn summarizing commonly used techniques and their complexities in application.**

Keywords: Image,Segmentation, Measurement

I. INTRODUCTION

Image segmentation in general is defined as a process of partitioning an image into homogenous groups such that each region is homogenous but the union of no two adjacent regions is homogenous (Pal and Pal, 1993). Efficient image segmentation is one of the most critical tasks in automatic image processing (Pavlidis, 1988; Haralick and Shapiro, 1985; Pal and Pal, 1993; Zhang, 1997; Cheng et al., 2001). Image segmentation has been interpreted differently for different applications. For example, in machine vision applications, it is viewed as a bridge between low level and high level vision subsystems (SpirKovska, 1993), in medical imaging as a tool to delineate anatomical structure and other regions of interest whose a priori knowledge is generally available (Pham et al.,2000) and in statistical analysis, it is posed as a stochastic estimation problem, with assumed prior distributions on image structure, which is widely used in remote sensing (Kerfoot etal., 1999). In remote sensing, it is often viewed as an aid to landscape change detection and land use/cover classification.

Aforementioned examples state that image segmentation is present in every kind of image analysis. This constitutes aplethora of literature on the image segmentation. This necessitates the organized categorisation of them. In order to present an organized review on image segmentation techniques, this review paper limits its analysis to optical remote sensing image analysis. This is essential because radar image segmentation is another horizon in remote sensing image analysis. From now onwards, remote sensing image would refer only to optical satellite remote sensing images.

Optical remote sensing imagery has been to a paradigm shift in the decade after year 1999. Landsat 7 launched in 1999 (with Multispectral (MS), 30m spatial resolution; Panchromatic (Pan), 15m spatial resolution), IKONOS launched in 1999 (MS, 4.0m; Pan, 1.0m), Quickbird launched in 2001 (MS, 2.44m; Pan, 0.61m), WorldView-1 launched in 2007 (Pan, 0.5m), GeoEye-1 launched in 2008 (MS, 1.65m; Pan, 0.42m), and WorldView-2 launched in 2009 (MS, 1.8m; Pan, 0.46m) are evidence of this shift. The spatial resolution has been changed so considerably that pixel size has become smaller than a size of car which was earlier bigger than two or three buildings. This led to research on new classification algorithms for high and very high resolution remote sensing images because traditional pixel based analysis was proved to be insufficient due to its incapability to handle the internal variability of complex scenes (Schiewe, 2002; Blaschke and Strobl, 2001; Carleer et al.,2005). These also propelled object based approach or Object Based Image Analysis (OBIA) for very high resolution image segmentation (Hay and Castilla, 2006). Detailed applications and discussion on the development trends of OBIA can be found in Blaschke (2010). However, in this paper applications based on OBIA are not the concern. This paper deals with technological aspect of image segmentation, which concern about identification of objects but not much related to further analysis of the object. Still object analysis is required for assessment of segmentation accuracy.

According to the aforementioned definition of segmentation, themajor thrust is on determining the suitable homogeneity measure which can discriminate the objects from each other. Some examples of the measures may be spectral, shape, texture and contexture. Most of the methods applied on remote sensing imageries are imported from other fields (Color image segmentation, Medical Image segmentation etc) and they work well because the underlying principal is same. For example, Cheng et al. (2001) extended the application of monochrome (single band) segmentation method, which was originally used on medical imagery, to colour image segmentation (threebands). With the numerous recent developments of new segmentation methodologies, the requirement of their categorisations based on successful applications have become essential. Therefore, the first objective of this paper is to categorise the technologies of image segmentation by conceptualising the implementation details. Image segmentation techniques which are applied on optical remote sensing image segmentation are included whereas those applied on active remote sensing satellite imagery like SAR imagery are excluded because of the reason already mentioned. However, in order to state the technological development some of the non-remote sensing applications are stated too. The second objective of this paper is to give an insight to the readers about the state of art of technological aspects of image segmentation and aid in deciding the mathematical form for image segmentation.

The rest of the paper is organized as follows. Section 2 discusses about the development of segmentation as per the existing review papers on image segmentation. Section 3 describes the categorisation of image segmentation from broad to fine level. Section 4 states the conclusion of the performed review. In order to state the development in a particular technology, similar methods are grouped and presented in a paragraph in rest of this literature.

I. DEVELOPMENT OF SEGMENTATION

One of the early application of image segmentation on remote sensing points to ECHO (Extraction and Classification of Homogeneous objects) classification by Kettig and Landgrebe (1976). This states that association of segmentation with remote sensing imagery was not much later than the operation of the first remote sensing satellite Landsat-1 in 1972. There have been many developments in remote sensing image processing techniques after that. Haralick and Shapiro (1985), Reed and Buf (1993), Spirkovska (1993) and Pal and Pal (1993) did extensive review on early stage of image segmentation techniques existed used in various applications along with remote sensing. Developments of image segmentation algorithms for remote sensing imageries have been drastically increased after the availability of high resolution imagery (Schiewe, 2002; Blaschke, 2010). This is obvious with the failure of pixel based techniques on high resolution imageries as discussed in the introduction section. Further, the commercially available software eCognition, since 2000 based on Fractal Net Evaluation Approach (FNEA), incorporating similarity of objects at hierarchical scale, has revolutionised the research on image segmentation and is still influencing the research very substantially (Baatz and Schäpe, 2000; Blaschke, 2010). This is why most of the review papers before the period of the year 2000 don't specifically cover remote sensing applications. After that we do have a few good review papers. For example, Schiewe (2002) categorised the available remote sensing technologies for high resolution imagery, Carleer et al. (2005) evaluated qualitatively some of the most widely used image segmentation technologies for very high spatial resolution satellite imagery, Shankar (2007) presented various techniques with mathematical details of image segmentation techniques and Blaschke (2010) on OBIA.

II. CATEGORISATION OF SEGMENTATION

The abundance of literature on image segmentation makes the categorisation both necessary and challenging. The approach of categorisation in this paper is supplementary to some earlier review papers mentioned in section 2. (Reed and Buf, 1993; Pal & Pal 1993; Spirkovska, 1993; Schiewe, 2002; Shankar, 2007). Most of the earlier literatures have categorised them as a) Edge based b) Point/Pixel based c) Region based and d) Hybrid approach. Guo et al. (2005) categorised them as colour based and texture based algorithms. However, a more clear delineation is required considering the techniques which are used to achieve segmented objects. A more general method of categorisation based on approach towards image analysis and applicable even beyond image segmentation domain are the bottom-up and top-down approaches. In image segmentation domain, they are often stated as model driven (top-down) and image driven approach (bottom-up) (Guindon, 1997). In this paper, this approach is stated as first stage of categorisation. It can also be stated as segmentation control based categorisation. However, in eCognition/Definiens developer software top-down and bottomup approach refers to hierarchy of segmentation (eCognition Elements User Guide, 2004). It can be said that bottom-up approach forms object by combining/merging pixels or group of pixels whereas top-down approach moves from splitting the whole image into image objects based on heterogeneity criteria (Benz et al., 2004). However, this is not the only definition.

The second stage of categorisation points to features or homogeneity measures based approaches used to delineate image objects. The third stage of categorisation is based on operations on image used to generate image objects. These are edge detection, region growing/splitting and may be both of them. It is important to note that these stages are highly interrelated and generally developed methods pick up one or more methods from the list at different stages to perform final segmentation. For example, Beveridge et al. (1989) used thresholding object/background model for generating initial regions and region merging algorithm with spectral, shape and connectivity as homogeneity measures. Tilton (1996) used both region growing and edge detection for Landsat TM data. A detailed description of the categorisations and their interrelationships are stated in the subsequent sections. Apart from aforementioned categorisation, image segmentation can also have supervised and unsupervised approach. Unsupervised segmentation holds its proximity to feature extraction and clustering whereas supervised segmentation incorporates segmentation accuracy as an addition to unsupervised scheme.

A. Categorisation based on homogeneity measure

Next stage of categorization corresponds to the homogeneity measures used for image segmentation. But before that it is necessary to determine the possible homogeneity measures of image features. This requires a well understanding of image objects and the final outcome of image segmentation. Image objects are real world objects represented on remote sensing image. With very high resolution satellite, image objects can be visualized by human eye. This has been addressed by some researchers. For example, Wang and Terman (1997) suggested sensory cues of segregation based on Gestalt psychology for segmentation and Fu and Mui (1981) as psycho physical perception problem for segmentation. It is similar to elements of analysis for image interpretation by human eye (pp. 67-68, Richards and Jia, 2006). Thus, the possible measures are based on similarity comprises spectral, texture, spatial, size, shape, and temporal. Some other semantic information prior knowledge, context and connectedness are also required (Wang and Terman, 1997). The primary homogeneity measure is spectral/tonal feature. Secondary homogeneity measures are spatial, texture, shape and size. Tertiary homogeneity measures are contextual, temporal and prior knowledge (pp. 67-68, Richards and Jia, 2006). As per the order, the most important is primary then secondary and then tertiary. Secondary and tertiary measures are more important when the boundaries of objects are required to be precise with very less mis-segmented pixels. In this study, more emphasis is given on secondary and tertiary measures which were not widely covered in earlier literatures. The list of measure may not be exhaustive but surely cover most of the available techniques existing for image segmentation. Subsequent sections describe the trend of techniques for different homogeneity measures used in image segmentation.

1. Spectral and Textural Features: The most primitive measures of homogeneity are spectral and textural features. Spectral values refer to grey levels or pixel values of an image. It has been long realised that using only spectral features good segmentation results are not possible but was still practiced due to the ease of incorporating them in digital format (Kettig and Landgrebe, 1976). Texture features points to spatial pattern represented by spectral values (Haralick et al., 1973). A textured image may have various texture patterns. However, quantitatively characterizing texture is not simple (pp. 128-130, Richards and Jia, 2006). Due to this fact texture segmentation has been studied widely generally in combination with other features like shape, spectral and contextual and various models till today. Chen and Pavlidis (1978) used co-occurrence matrix and a quadtree based structure to determine texture similarity for grouping pixels in a region. Cross et al. (1988) also used quadtree based hierarchical structure and applied texture measure was local difference statistics. Guo et al. (2005) usedtexture measure derived from local binary pattern and used wavelet transform to pre-process the image and derived texture from local binary pattern. They also used quadtree structure for splitting and merging. It can be seen from trend that quadtree based hierarchical image splitting has been the trusted method of texture segmentation for decades. Conners et al. (1984) used spatial grey level dependence method (SGLDM) and six texture measures namely inertia, cluster shade, cluster prominence, local homogeneity, energy and entropy in region growing algorithm based on split and merge tecnique. Ramstein and Raffy (1989) used variogram and fractal dimension measures for texture segmentation and classification. Ryherd and Woodcock (1996) used a 3x3 adaptive window to calculate texture image based on local variance and applied a multi-pass region growing algorithm which builds spatial homogeneous objects using Euclidean distance in n-dimensional space. They showed that segmentation accuracy of derived texture image is better when compared with original image, used spectral property only. Algorthm was tested with SPOT panchromatic image. Texture segmentation is one of the most sought after segmentation technique. It is evident from Reed and Buf (1993) and the above literature. This is mainly because of the presence of highly textured regions in high resolution satellite imagery. Currently, the research has shifted from texture to multiresolution model.

2. Shape and Size Features: The importance of shape and size measure could be understood when the natural object are to be identified on satellite imagery. For example, a river and a pond may has same spectral, texture and spatial properties but they differ in shape and size. It is because rivers are linear and unbounded features whereas ponds are non-linear and bounded features. Shape and size measures are mostly utilized as complementary to each other. Further, they are always applied in conjunction with the spectral and texture measures. Only some substantial algorithms based on the recent developments are mentioned. Beveridge et al. (1989), performed over-segmentation and then utilised shape, connectivity and size measure for region merging to achieve segmentation. Multi-resolution models represent the size of object through spatial scales (Bongiavanni et al., 1993).

Fractal Net Evaluation approach (explained in section 3.2.5) sed in commercial software, eCognition/Definiens developer, also uses scale, shape and compactness parameter. The state of art use of shape and size refers to multi-scale/multiresolution approach to image segmentation. Shape and size measures are especially helpful when delineating complex objects in high resolution satellite imagery.

3. Context: Context generally refers to spatial context which means relationship of pixels with its neighbourhood (Thakur and Dikshit, 1997). Contextual information is also used in conjugation with spectral or texture or both measures. Few methods are found which utilise specifically context based egmentation. Context helps in avoiding fragmentation of a segment and merging. For example in an urban image, cars in a parking lot may cause fragmentation unless context measures are applied. A good recent example of context based segmentation is Fan and Xia (2001). They deduced context information from spatial and scale space of image and modeled five context models with quadtree model for scale dependency. They called their algorithm as multi-contextual (due to five context models) and multi-scale approach to Bayesian segmentation which in mathematical terms solves context-based mixture model likelihood. They used their methods for aerial and SAR imagery. Even eCognition/Definiens Developer software has the capability of including the context information based on neighbourhood relationship measures. Benz et al. (2004) demonstrates in the paper that how eCognition integrates spatial and scale context as semantic information in identifying the appropriate image objects. Contextual constraints are used in segmentation and classification and are well modeled by Markov Random Field. This is why several context-based classifications use MRF model (Melgani and Serpico, 2003; Jackson and Landgrebe, 2002). Context is especially useful when segmentation requires bigger area to be identified as one segment e.g. land use classification.MRF models are currently the best model for implementation of contextual measures.

4. Temporal: Temporal measure refers to measurement based on images of same area and sensor characteristics in different time (pp. 67-68, Richards and Jia, 2006). Temporal measure is not directly used for segmentation but is used as an application of segmented image. Carlotto (1997) performed temporal segmentation for change detection from Landsat TM. He used total difference image to segment based on histogram thresholding. Jeansoulin et al. (1981) performed segmentation using fuzzy edge detection and region growing for segmentation and demonstrated how temporal criterion can be used to detect changes based on
objects. Hanaizumi et al. (1991) used spatial segmentation for change detection and showed result on Landsat TM imagery. They used division and detection procedure where divison/region-splitting was performed by fitting regression model on pixel scattergram. Dambra et al. (1991) fused multitemporal imagery using segmented image. SAR segmented image is also used for change detection. Several SARsegmenting methods are reviewed by Caves et al. (1996). Yamamoto et al. (2001) detected change in SPOT HRV and Landsat TM image using 3-D segmentation with time as Z axis. They applied local statistical regression model for region splitting using spatial and spectral measures. Hall and Hay (2003) used multi-object scale analysis for change detection which utilises Marker Controlled watershed segmentation (Beucher, 1992). Lhermitte et al. (2008) introduced multitemporal hierarchical image segmentation. They segmented the 10 daily data of SPOT VGT sensor by first decomposing original image time series in Fast Fourier Transform component and then performed hierarchical segmentation analogous to eCognition (Baatz and Schäpe, 2000) using Euclidean distance between FFT components of same frequency as measure of similarity.

Temporal characteristics have important application in monitoring changes like land-use change, disaster mapping, traffic flows, crop mapping etc (pp-280-81, Campbell, 2007). Temporal segmentation has been used mainly for change detection in a series of temporal image. Its application is mainly seen for large area change detection rather than small area.

Thus, more applications have been found on low resolution images than high resolution.

I

5. Prior Knowledge: Prior knowledge refers to photointerpreter knowledge regarding the regions/objects of the image (pp. 342-352, Richards and Jia, 2006). It may be the knowledge of classes of the image region or about some specific area, building or trends etc. Incorporating prior knowledge in image analysis is one steps towards developing artificial intelligence in the machine (Srinivasan and Richards, 1993). Prior knowledge may not be the primary measure for segmentation but it has the capability of utilising the location based information. For example, it is our prior knowledge which generally says that small buildings mean residential areas and large buildings means commercial or institutional areas. This indicates towards differentiation based on shape properties. In the next paragraph, few prior knowledge based segmentation or ior knowledge based homogeneity measure derivation aredescribed. Ton et al. (1991) divided segmentation techniques into two types as partial segmentation (without using a priori knowledge) and complete segmentation (using a priori knowledge). The approach for knowledge based can be further divided into histogram-oriented and cluster-oriented (Ton et al., 1991; Paudyal et al., 1994). Most of the popular method like Hierarchical split and Merge (Ojala and Pietikainen, 1999), region growing, multi-resolution used by eCognition (Baatz and Schäpe, 2000) etc are partial segmentation techniques. Ton et al. (1991) used spectral and spatial knowledge rules for supervised segmentation of Landsat TM image. They automated generation of spectral knowledge based rules based on training data and hierarchical classification. They applied both threshold and region growing for segmentation.

Liu et al. (1993) used texture measure for region uniformity and contexture information at pixel level for segmentation. They used knowledge in determining the best texture measure, which gives minimum error using multivariate Gaussian Bayesian classifier, out of the available for good segmentation. The method used is essentially supervised segmentation. Using similar concepts some researchers incorporated knowledge in textural measures (Paudyal et al., 1994; Simman, 1997).

Smits and Annoni (1999) used no prior information but derived knowledge, automatically from a selected region, to select the best feature which can distinguish object from its neighbours. Jinghui et al. (2004) also used GIS prior information to extract building from Quickbird imagery using fuzzy connectedness algorithm.

Poggi et al. (2005) used tree structured MRF model in incorporating prior knowledge for supervised segmentation. Benz et al. (2004) also showed how expert knowledge can be included in segmentation based fuzzy classification.

Prior knowledge is incorporated in mathematical models by using class distribution information. In fuzzy models, it can be incorporated as semantic rules (Benz et al., 2004). Prior knowledge is specifically useful when for segmentation of complex landscape object indistinguishable using texture and context.

CONCLUSIONS

With the numerous amounts of image segmentation techniques presented in this paper, it might be possible to get confused regarding what is presented in this paper. Thus, it is important to summarize all of those to regain the content of this paper. Image segmentation methodologies were categorized in three stages. At first stage comes model driven approach and image driven approach (mainly based on statistical analysis). The second stage corresponds to homogeneity based measure, and final category corresponds to mode of operations on an image, e.g. edge detection, region growing/splitting. In model driven approach, object background model is insufficient for remotely sensed imagery. Neural model generally suffers from complexity regarding decision of network structure, proper learning and generalization of network. Hence, neural model is not one of the liked approaches by most of the researchers. Markov Random Field model has attracted quite a decent research in image segmentation. It can utilise significant image properties namely, spectral, spatial, texture, contexture and prior knowledge. However, MRF lacks the integration of shape and size and implementation of MRF is very complex. Fuzzy model has been applied in remote sensing image segmentation mostly by means of fuzzy clustering of image or fuzzy thresholding. The strength of fuzzy model lies in ambiguity resolution. It can easily ensemble itself with neural model, MRF model and also histogram thresholding (Chen et al., 1997; Caillol et al., 1993; Shankar, 2007).

Multi-resolution (MR)/Multi-scale model is the most widely used model in remote sensing image segmentation. It has also been incorporated in a commercial software eCognition/ Definiens Developer. This model is capable identify object and object features at its intrinsic scale which makes object extraction of various scales possible (Chen at al. 2009). The problem of MR approach is scale representation and information extraction from each scale. The idea of MR approach is complex but when appropriately implemented has wide usage especially in remote sensing satellite images dealing with urban areas. Watershed model based on mathematical morphological operators is another budding technology with respect application in remote sensing image egmentation. Further, research on this approach is required. Homogeneity measures described in this paper are spectral, spatial, texture, shape, size, contextual, temporal and prior knowledge. Spectral measure is the most primitive one and quite long it has been realised that this alone wouldn't be able to deal with high resolution satellite imagery (Zhong et al., 2005).

The second most widely applied homogeneity measure is based on texture. Texture segmentation is more successful because it inherits spectral and spatial properties in itself. However, this would still not yield a perfect segmentation. A better segmentation would require a model or methodologies which utilise most of the above mentioned measures to calculate region homogeneity or heterogeneity threshold. Integration of prior knowledge and contextual information has seen quite a good research in segmentation. The selection of segmentation approach depends on what quality of segmentation is required. Further, it also depends on what scale of information is required. Few examples, based on done literature review in this paper, would be stated now to illustrate the idea. For urban GIS applications objects at different scale are required. For landuse coarse scale egmentation is required whereas for land cover fine scale. Hence, multi-resolution model would be the best choice. For highly textured image MRF Fuzzy model would be good choice to represent ambiguity of region boundaries. Neural model would be good choice no prior distribution can be assumed and not very high quality object information is required. Among homogeneity measures, spectral, shape, size, scale, compactness and texture should be concerned when complex landscapes are to be analyzed.

As a part of future recommendation, some of the mentioned approaches in this paper should be implemented to look how each behaves on same image. Behaviours with images of different spatial resolution would be quite interesting. Further, addition of existing quantitative analysis of recent segmentation evaluation techniques would be quite helpful.

REFERENCES

[1] Atkinson, P. M. and Tatnall, A. R. L., 1997. Introduction Neural networks in remote sensing. *International Journal of Remote sensing*, 18(4), pp. 699-709.

[2] Baatz, M. and Schäpe, A., 2000. Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation. In: *J.Strobl, & T. Blaschke (Eds.), Angewandte geographische informationsverarbeitung*, Heidelberg: Wichmann-Verlag, Vol. XII, pp. 12-23. Bandyopadhyay, S., 2005. Satellite image classification using genetically guided fuzzy clustering with spatial information.

[3] *International Journal of Remote sensing*, 26(3), pp. 579-593. Baraldi, A. and Parmiggiani, F., 1995. A neural network for unsupervised categorization of multivalued input patterns: An application to satellite image clustering. *IEEE Transactions on Geoscience and Remote Sensing*, 33(2), pp. 305-316.

[4] Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I. and Heynen, M., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry & Remote Sensing*, 58(3-4) pp. 239–258.

[5] Beucher S., 1992. The watershed transformation applied to image segmentation. *Scanning Microscopy Supplemen*t, 6, pp. 299–314.

[6] Beveridge, J. R., Griffith, J., Kohler, R. R., Hanson, A. R. and Riseman, E. M., 1989. Segmenting images using localized

[7] histograms and region merging. *International Journal of Computer Vision*, 2(3), pp. 311-347. Blaschke, T., 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing,* 65(1), pp. 2-16.

[8] Blaschke, T. and Strobl, J., 2001. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS.

[9] *GIS-Zeitschrift für Geoinformationssysteme,* 14(6), pp. 12-17. Bongiovanni, G., Cinque, L., Levialdi, S. and Rosenfeld, A., 1993. Image segmentation by a multiresolution approach.

[10] *Pattern Recognition*, 26(12), pp. 1845-1854. Bouman, C. A. and Shapiro, M., 1994. A multiscale random field model for Bayesian image segmentation. *IEEE Transactions on Image Processing*, 3(2), pp. 162-177.

[11] Caillol, H., Hillion, A. and Pieczynski, W., 1993. Fuzzy random fields and unsupervised image segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 31(4), pp. 801-810. Campbell, J. B., 2007. *Introduction to Remote*

Sensing. Taylor and Francis, London, pp. 280-281.

[12] Cannon, R. L., Dave, J. V., Bezdek, J. C. and Trivedi, M. M., 1986. Segmentation of a thematic mapper image using the fuzzy c-means clusterng algorthm. *IEEE Transactions on Geoscience and Remote Sensing,* 24(3), pp. 400-408.

[13] Carleer, A. P., Debeir, O. and Wolff, E., 2005. Assessment of very high spatial resolution satellite image segmentations. *Photogrammetric Engineering & Remote Sensing*, 71(11), pp. 1285–1294.

[14] Carlotto, M. J., 1997. Detection and analysis of change in remotely sensed imagery with application to wide area surveillance. *IEEE Transactions on Image Processing*, 6(1), pp. 189 – 202.

[15] Caves, R. G., McConnell, I., Cook, R. and Quegan, S., 1996. Multichannel SAR segmentation: algorithms and applications. *IEE Colloquium on image processing for remote sensing*, 27, pp. 2/1-2/6.

[16] Chehdi, K. and Liao, Q., 1993. Satellite image segmentation using edge-region cooperation. In: *Proceedings of IEEE Pacific Rim Conference on Communications, Computers and Signal*

[17] *Processing,* 1, pp. 47-50.

[18] Chen, J., Pan, D. and Maz, Z., 2009. Image-object detectable in multiscale analysis on high-resolution remotely sensed imagery. *International Journal of Remote sensing,* 30(14), pp. 3585– 3602.

[19] Chen, K. S., Tsay, D. H., Huang, W. P. and Tzeng, Y. C.,1996. Remote sensing image segmentation using a Kalman filter-trained neural network. *International Journal of Imaging Systems and Technology*, 7(2), pp. 141-148.

[20] Chen, P. C. and Pavlidis, T., 1978. Segmentation by texture using a co-occurrence matrix and a split-and-merge algorithm. In: *Proceedings of the 4th International Joint Conference on Pattern Recognition*, Kyoto, Japan, pp. 95- -3684.

Wireless and Mobile Networks

Meenakshi, Sukhwinder Kaur
M.Tech, ECE, ACET, Amritsar
Email id: er_mania88@ymail.com, Sukhwinderkaur86@yahoo.co.in

Abstract — Mobile wireless industry has started its technology creation, revolution and evolution since early 1970s. In the past few decades, mobile wireless technologies have experience 4 or 5 generations of technology revolution and evolution, namely from 0G to 4G. Wireless Communications shown great impacts on the global economy and social development throughout the world. The core technology for radio and land mobile communications progressing fast but there are maturing signs on the horizon. This paper gives the overview of wireless and mobile networks.

Introduction

Satellite technology is developing slower but it comes with the special potential of universal coverage at a greater costs and limited bandwidths. There are some technologies which are actually responsible for such facilities on our mobiles. There is numerous mobile communication technologies which makes growth of cellular industry to this extend one of the technology which is by far the best and most widely used in mobile communication is GSM, CDMA, GPRS, VOIP etc. Today's mobiles networks supports features likes SMS, GPRS, MMS, emailing facility on mobile, Bluetooth, WAP and many more depending upon how reputed and bigger mobile network company is. Mobile phones of today's age are now equal to portable PCs. These mobile phones connect to their cellular networks and these cellular networks are connected to Public switching telephone network (PSTN). Today mobile devices can be enabled to use a variety of communications technologies such as: Wireless fidelity (WiFi) Bluetooth, Global system for mobile communications (GSM), General packet radio service (GPRS).

Mobile Networks

Mobile network has already creating hype throughout the world. We are using so many features in our mobile these days which most of us have never dreamed off. We are using SMS, MMS, sending pictures and video files in amazingly quick time, finding locations, accessing high speed internet in your mobiles are the features which were not possible few years back. Mobile networks has groomed a lot in past few years, major reasons for rapid advancements in mobile network technology is requirements for being mobile or connectivity on move. There are number of mobile communication technologies which have been developed ever since from analog days and then new technology is released to cope with telephonic industry demands. Cellular companies use AMPS, D-AMPS, CMMA2000, UMTS, GSM, EVDO etc. AMPS however pretty much vanished from the scene, AMPS network system was based on analog communication technology, latest features were not supported by AMPS therefore all cellular networks worldwide have adopted digital communication methodologies to meet the need of consumers.

1G: First Generation Networks

The first mobile phone system in the market was AMPS. It was the first U.S. cellular telephone system, deployed in Chicago in 1983. The main technology of this _1st generation mobile system was FDMA/FDD and analog FM.

2G: Second Generation Networks

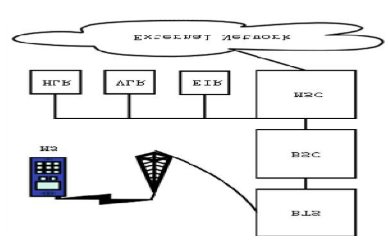
Digital modulation formats were introduced in this generation with the main technology as TDMA and CDMA . The 2G systems introduced three popular TDMA standards and one popular CDMA standard in the market. The second generation introduced a new variant of communication called SMS or text messaging. It was initially available only on GSM networks but spread eventually on all digital networks.

TDMA Standard

GSM(GLOBAL SYSTEM FOR MOBILE COMM.)

GSM leads the world as the fastest growing mobile/wireless digital technology that is most reliable and is available in the marketplace today. It provides the integrated high-speed data, fax, voice mail, paging and short messages services features. It offers the secure communication, privacy and fraud prevention.

It was the first fully digital system utilizing the 900 MHz frequency band. The initial GSM had 200 KHz radio channels, 8 full-rates or 16 half-rate TDMA channels per carrier, encryption of speech, low speed data services and support for SMS for which it gained quick popularity.



Architecture and Working of GSM Networks GSM technology had taken over mobile communication technologies and grown over to 214 countries around the world. GSM network is consist of three major systems are SS, which is known to be

The Switching System,

The Base Station System

The operation and support System

Architecture of GSM

The Switching System

The Switching system is very operative system in which many crucial operations are conducted, SS systems holds five databases with in it which performs different functions. If we talk about major tasks of SS system it performs call processing and subscriber related functions. These five databases from SS systems are HLR, MSC, VLR, AUC and EIR. HLR- Home Location Register*:*

(a) HLR is database, which holds very important information of subscribers. It is mostly known for storing and managing information of subscribers.

b) MSC- Mobile Services Switching Center**:**

MSC is also important part of SS, it handles technical end of telephony. It is build to perform switching functionality of the entire system. It's most important task is to control the calls to and from other telephones, which means it controls calls from same networks and calls from other networks.

C) VLR- Visitor Location Register:

VLR performs very dynamic tasks; it is database which stores temporary data regarding connected to MSC, when subscribe moves to different MSC location, Visitor location register

VLR integrates to MSC of current location and requests the data about subscriber or Mobile station (MS) from the Home Location Register –HLR. When subscriber makes a call the Visitor location register-VLR will have required information for making call already and it will not required to connect to Home Register Location -HRL again

d) AUC- Authentication Center:

AUC is small unit which handles the security end of the system. Its major task is to authenticate and encrypt those parameters which verify user's identification and hence enables the confidentiality of each call made by subscriber.

e) EIR – Equipment Identity Register:

EIR is another important database which holds crucial information regarding mobile equipments. EIR helps in restricting for calls been stolen, mal functioning of any MS, or unauthorized access. AUC – Authentication center and EIR- Equipment Identity registers are either Stand-alone nodes or sometimes work together as combined AUC/EIR nodes for optimum performance.

The Base Station System (BSS)

The base station system have very important role in mobile communication. BSS are responsible for connecting subscribers (MS) to mobile networks .All the communication is made in Radio transmission. The Base station System is further divided in two systems. These two systems, they are BSC, and BTS.

BTS – The Base Transceiver Station:

Base transceiver Station is the radio equipment which receives and transmits voice data at the same time. BSC control group of BTSs.

BSC – The Base Station Controller**:**

It performs the function of high quality switch by handover over the MS to next BSC when MS goes out of the current range of BTS, it helps in connecting to next in range BTS to keep the connection alive within the network. It also performs functions like cell configuration data, control radio frequency in BTS. Data moves to MSC-Mobile switching center after BSC done processing it. MSC is switching center which acts as bridge between different mobile networks.

The Operation and Support System (OSS)

It helps in managing, centralizing, local and regional operational activities required for GMS networks. Maintaining mobile network organization, provide overview of network, support and maintenance activities are other important aspects of Operation and Support System

Different Frequency Bands

There are three different frequency bands on which mobile phones are usually operates and these are Dual Band, Tri-Band and Quad Band.

Dual Band : Dual frequency band operates on 900MHz and 1800 MHz, that means mobile phone that supports dual band can be operated anywhere in the world where 900 MHz and 1800 MHz frequencies are used.

Tri-Band: As name is obvious three frequencies are supported in Tri Band, these frequencies are 900 MHz, 1800MHz and 1900 MHz. Tri band is also supported all around the world these days.

Quad-Band: Quad Band supports four frequencies which are 850 MHz, 900 MHz, 1800 MHz , 1900 MHz Quad band also enables GSM phones to road almost anywhere in the world. All countries support GSM networks hence make communication possible.

Advantages of GSM

GSM is a state of the art technology offering voice, data, fax and SMS capabilities. Most of the mobile phone operators now offer email to SMS service by which you can receive the important news, flight status and other useful information on your mobile phone. Many operators give access to the ISDN services by which fast data transmission is possible. One of the biggest advantages of the GSM network is its use of SIM, or subscriber identity module, cards to identify users' phones. SIM cards are small chips that contain information like a subscriber's phone number, contacts, preferences and other data. Users can move a single SIM card from one phone to another, making it easy to transfer service between phones without losing important data. The GSM network extends around the world, allowing users of GSM phones to place roaming calls from more locations.

CDMA (Code Division Multiple Access)

Interim Standard 95 (IS-95)

The IS-95 standard, also popularly known as CDMA One uses 64 orthogonally coded users and code words are transmitted simultaneously on each of 1.25 MHz channels. Certain services that have been standardized as apart of IS-95 standard are: short messaging service, slotted paging, over-the-air activation (meaning the mobile can be activated by the service provider without any third party intervention), enhanced mobile station identities etc. This was only one cellular telecommunications system, and the first CDMA system was launched in September 1995 .It uses of spread spectrum transmission technology .The use of CDMA offers several advantages and it is for this reason that CDMA technology has been adopted for many 3G cellular telecommunications systems. One of the chief claims for CDMA is that it gives significant improvements in network capacity. Original expectations for some of the proponents of CDMA technology were for some very significant improvements.

Advantages of CDMA include:

- Increased cellular communications security.
- Simultaneous conversations.
- Increased efficiency, meaning that the carrier can serve more subscribers.
- Smaller phones.
- Low power requirements and little cell-to-cell coordination needed by operators.
- Extended reach - beneficial to rural users situated far from cells.

Disadvantages of CDMA include:

- Due to its proprietary nature, all of CDMA's flaws are not known to the engineering community.
- CDMA is relatively new, and the network is not as mature as GSM.
- CDMA cannot offer international roaming, a large GSM advantage.
- CDMA doesn’t require any Sim card.

2.5G Mobile Network

In an effort to the 2G standards for compatibility with increased throughput rates to support modern Internet application, the new data standards were developed to be overlaid on 2G standards and this is known as 2.5G standard. There are various up gradations in 2G with which advent of 2.5G was made and that are supporting higher data rate transmission for web browsing supporting e-mail traffic enabling location-based mobile service.

2.5G networks also brought into the market some popular application, a few of which are: Wireless Application Protocol (WAP), General Packet Radio Service (GPRS), High Speed Circuit Switched Dada (HSCSD), Enhanced Data rates for GSM Evolution (EDGE) etc.

3G: Third Generation Networks

As the use of 2G phones became more widespread and people began to utilize mobile phones in their daily lives, it became clear that demand for data services (such as access to the internet) was growing. The main technological difference that distinguishes 3G technology from 2G technology is the use of packet switching rather than circuit switching for data transmission. 3G is the third generation of mobile phone standards and technology. It is based on the International Telecommunication Union (ITU) family of standards under the International Mobile Telecommunications-2000 (IMT-2000). 3G networks are wide area cellular telephone networks which evolved to incorporate high-speed internet access and video telephony. This system is called Universal Mobile Telecommunications System (UMTS).Europe, Japan, and Asia have agreed upon a 3G standard called the Universal Mobile Telecommunications System (UMTS). During the development of 3G systems, 2.5G systems such as CDMA and GPRS were developed as extensions to existing 2G networks. These provide some of the features of 3G without fulfilling the promised high data rates or full range of multimedia services. 3G networks enable network operators to offer users a wider range of more advanced services while achieving greater network capacity through improved spectral efficiency. Services include wide-area wireless voice telephony, video calls, and broadband wireless data, all in a mobile environment. Additional features also include HSPA data transmission capabilities able to deliver speeds up to 14.4Mbit/s on the down link and 5.8Mbit/s on the uplink. 3G networks are wide area cellular telephone networks which evolved to incorporate high speed internet access and video telephony. IMT-2000 defines a set of technical requirements for the realization of such targets, which can be summarized as follows:

High data rates: 144 kbps in all environments and 2 Mbps in low-mobility and indoor environments symmetrical and asymmetrical data transmission circuit-switched and packet-switched-based services speech quality comparable to wire-line quality improved spectral efficiency several simultaneous services to end users for multimedia services

global roaming. Open architecture for the rapid introduction of new services and technology.

4G (4^{th} generation)

4G is the fourth generation of cell phone mobile communications standards. It is a successor of the third generation (3G) standards. A 4G system provides mobile ultra-broadband Internet access, for example to laptops with USB wireless modems, to smart phones, and to other mobile devices. The expectations with 4G are high with respect to data rates, spectral efficiency ,mobility and integration.

Key features

The following key features can be observed in all suggested 4G technologies:

Physical layer transmission techniques are as follows:
MIMO: To attain ultra high spectral efficiency by means of spatial processing including multi-antenna and on multi-user MIMO

- Frequency-domain-equalization,
  For example *multi-carrier modulation* (OFDM) in the downlink or *single-carrier frequency-domain-equalization* (SC-FDE) in the uplink: To exploit the frequency selective channel property without complex equalization
- Frequency-domain statistical multiplexing, for example (OFDMA) or (single-carrier FDMA) (SC-FDMA, a.k.a. linearly precoded OFDMA, LP-OFDMA) in the uplink: Variable bit rate by assigning different sub-channels to different users based on the channel conditions
- Turbo principle error-correcting codes: To minimize the required SNR at the reception side
- Channel-dependent scheduling: To use the time-varying channel
- Link adaptation: Adaptive modulation and error-correcting codes
- Mobile-IP utilized for mobility

CONCLUSION

The first operational cellular system provided voice communications. The 2nd generation (2G) of the wireless mobile network used low band digital signaling which is known as Global Systems for Mobile Communication (GSM).The objective of the 3G was to develop a new protocol & new technologies. The 4G framework establish to accomplish new levels of users experience & multiservice capacity like GSM for mobile communication, GPRS-General Packet Radio Service, IMT-2000- International Mobile Communication, Wi-Fi- Wireless fidelity, Bluetooth. 4G has the following properties: Ubiquity, Multiservice Platform and Low bit cost.

References

1. “Wireless and mobile networks” Research http:// www. iwireless .com
2. http://www.wikipedia.org
3. http://www.1000projects.com
4. Wireless and mobile networks by William Stalling

Performance Review of FIR Filters

Gourav Soni[1] and Dr. V.K. Banga[2]
[12]Associate Professor, Department of Electronics & Communication Engineering, ACET, Amritsar, India
[2]Professor, Department of Electronics & Communication Engineering, ACET, Amritsar, India
Email: reetu.gaggel1204@gmail.com, sanjeev_be@indiatimes.com

Abstract: **A new approach to piecewise-polynomial approximation and recursive implementation structures for linear-phase Finite Impulse Response (FIR) filters have been recently proposed by various authors. In this paper, we describe hardware prototype implementations of the linear-phase FIR filters using a Field Programmable Gate Array (FPGA) based platform. Digital computer arithmetic is an aspect of logic design with the objective of developing appropriate algorithms in order to achieve an efficient utilization of the available hardware. Given that the hardware can only perform a relatively simple and primitive set of Boolean operations, arithmetic operations are based on a hierarchy of operations that are built upon the simple ones. Since ultimately, speed, power and chip area are the most often used measures of the efficiency of an algorithm, there is a strong link between the algorithms and technology used for its implementation. In this paper the pipelining technique and distributed arithmetic algorithm will be discussed which will target the speed of the FIR filters.**

Keywords: DA, Pipeline,Filter structure, Finite Impulse Response (FIR) digital filters, FPGA

I. INTRODUCTION

Digital FIR filters are widely used in many signal processing applications due to their many well-known favorable properties and popular design methods. Their main drawback is high arithmetic complexity, i.e., the required number of adders, multipliers and delay elements needed in conventional implementations especially when a narrow transition band is required [1]. An efficient approach to overcome the above-mentioned problem is to synthesize linear-phase FIR filters so that their impulse response is piecewise- polynomial and the implementation is performed using recursive structures [2–8]. The arithmetic complexity of these filters is proportional to the number of impulse-response pieces and the overall polynomial order rather than the actual filter order.Also, other computationally efficient approaches have been developed to reduce ar ithmetic complexity but these methods are implemented with non-recursive structures and they are mostly suited for signal-processor applications or software based implementations. The most efficient methods of this kind are Interpolated FIR-filter approach and Frequency–response-masking approach [9,10]. Mostly, the computationally efficient methods are based on the fact that the impulse response of a direct-form narrowband FIR filter generally has a smooth shape, and thus there is a strong correlation between successive coefficient values. Therefore, piecewise-polynomial approximation of the impulse response can give significant savings in arithmetic complexity. On the other hand, polynomial responses can be efficiently generated using recursive structures based on cascaded accumulators [2] High-speed digital signal processor (DSP) systems are increasingly being implemented on field programmable gate array (FPGA) hardware platforms [1]. The FPGA platform provides high performance and flexibility with the option to reconfigure. The occupied resources of FPGA chips, in the terms of multipliers and slices, can be significantly reduced by time sharing the hardware resources [2], that is, the same implemented structure is used several times between two successive input samples. An extensively used solution is the hardware folding technique to time-multiplex repeating functional blocks onto a simple functional unit [3]. FPGA has become an extremely cost-effective means of off-loading computationally intensive digital signal processing algorithms to improve overall system performance. The digital filter implementation in FPGA, utilizing the dedicated hardware resources can effectively achieve application-specific integrated circuit (ASIC) like performance while reducing development time cost and risks.

FIR Filter [10]

The impulse response, the filter's response to a delta input, is 'finite' because it settles to zero in a finite number of sample intervals. This is in contrast to infinite impulse response filters which have internal feedback and may continue to respond indefinitely. The impulse response of an Nth order FIR filter lasts for N+ 1 sample, and then dies to zero. The difference equation which defines how the input signal is related to the output signal is given as

$Y[n] = b_0\, x[n] + b_1\, x[n-1] + \ldots\ldots.. + b_N\, x[n-N]$

where $x[n]$ is the input signal, $y[n]$ is the output signal and b_i are the filter coefficients. N is known as the *filter order*; an N^{th}-order filter has $(N + 1)$ terms on the right-hand side; these are commonly referred to as *taps*. The previous equation can also be expressed as a convolution of filter coefficients and the input signal.

$$y[n] = \sum_{i=0}^{N} b_i\, x[n-i]$$

To find the impulse response we set

$X[n] = \delta[n]$

where $\delta[n]$ is the delta impulse. The impulse response for an FIR filter is the set of coefficients b_n, as follows

$h(n)=\sum_{i=0}^{N} b_i\, \delta[n-i]$

for n=0 to N.

The Z-transform of the impulse response yields the transfer function of the FIR filter

$H(z)=Z\{h[n]\}$

$= \sum_{n=-\infty}^{\infty} h[n]z^{-n}$
$= \sum_{n=0}^{N} b_n z^{-n}$

FIR filters are clearly *bounded-input bounded-output* (BIBO) stable, since the output is a sum of a finite number of finite multiples of the input values, so can be no greater than $\sum | b_i|$ times the largest value appearing in the input.

II. FUNCTIONAL BLOCK DIAGRAM FOR AN M ORDER DIGITAL FILTER[12]

A FIR filter can be described by the Z-domain functional block diagram shown in fig 1 where (in synchronous operation) each box labelled with z^{-1} denotes a register cell having a delay of one clock cycle. The diagram represents the data paths and operations that must be performed by the filter. Each stage of the filter holds a delayed sample of the input, and the connection at the input and the connections at the outputs of the stages are referred to as taps, and the set of coefficients $\{b_k\}$are called the tap coefficients of the filter. An Mth order or Nth order FIR will have M+1 taps. The samples are weighted by the tap coefficients and added together to form the output, y[n]. the adders and multipliers of the filter must be fast enough to form y(n) before the next clock, and at each stage they must be sized to accommodate the width of their datapaths. the longest signal path through the combinational logic include M stages of addition and one stage of multiplication.

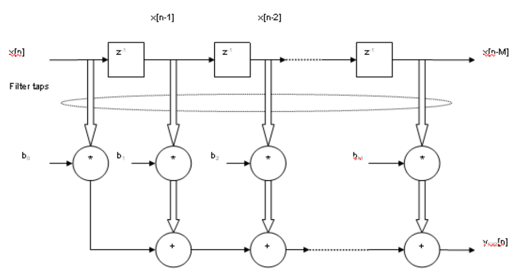


Fig 1 Functional Block Diagram for an M order Digital Filter

MAC (Multiplier Accumulator unit) Based Architecture for M order FIR filter. The architecture shown in fig 2 consists of a shift register, multipliers, and adders implementing an Mth order FIR. The datapaths must be wide enough to accommodate the output of the multipliers and adders. The samples are encoded as finite length words, and then shifted in parallel through a series of M registers. A cascaded chain of MAC's form machine

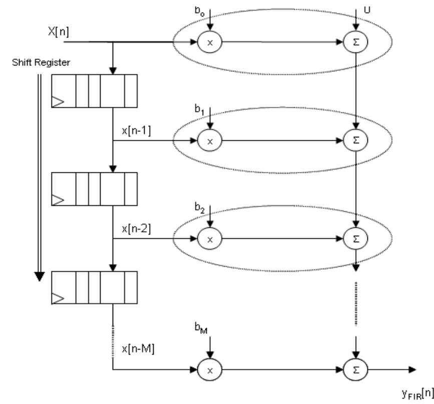


Fig 2 Mac Based Architecture for Mth order FIR filter IIR Filter

IIR filters may be implemented as either analog or digital filters. In digital IIR filters, the output feedback is immediately apparent in the equations defining the output. Note that unlike with FIR filters, in designing IIR filters it is necessary to carefully consider "time zero" case in which the outputs of the filter have not yet been clearly defined. Design of digital IIR filters is heavily dependent on that of their analog counterparts because there are plenty of resources, works and straightforward design methods concerning analog feedback filter design while there are hardly any for digital IIR filters. As a result, usually, when a digital IIR filter is going to be implemented, an analog filter (e.g. filter, Butterworth, Elliptic filter) is first designed and then is converted to a digital filter by applying discretization techniques such as Bilinear transform or Impulse invariance. An infinite impulse response (IIR) filter is a recursive filter where the current output depends on previous outputs. [3]

Transfer function derivation

Digitals filters are often described and implemented in terms of the difference equation that defines how the output signal is related to the input signal:

$$y[n] = \frac{1}{a_0}(b_0x[n] + b_1x[n-1] + \cdots + b_px[n-P] \\ - a_1y[n-1] - a_2y[n-2] - \cdots \\ - q_Qy[n-Q])$$

where

- P is the feedforward filter order
- b_i are the feedforward filter coefficients
- Q is the feedback filter order
- a_i are the feedback filter coefficients
- x[n] is the input signal
- y[n] is the output signal.

A more condensed form of the difference equation is:

$$y[n] = \frac{1}{a_0}(\Sigma_{i=0}^{P} b_i x[n-i] - \Sigma_{j=1}^{Q} a_j y[n-j])$$

which, when rearranged, becomes:

To find the transfer function of the filter, we first take the Z-transform of each side of the above equation, where we use the time-shift property to obtain:

We define the transfer function to be:

Considering that in most IIR filter designs coefficient a_0 is 1, the IIR filter transfer function takes the more traditional form:

Description of block diagram

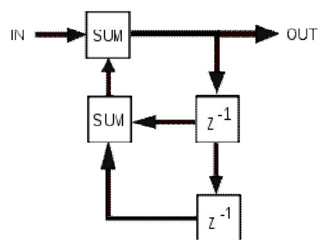


Fig 3 Simple IIR filter block diagram

A typical block diagram of an IIR filter looks like the following. The z^{-1} block is a unit delay. The coefficients and number of feedback/feedforward paths are implementation-dependent.

Stability

The transfer function allows us to judge whether or not a system is bounded-input, bounded-output (BIBO) stable. To be specific, the BIBO stability criteria require the ROC of the system include the unit circle. For example, for a causal system, all poles of the transfer function have to have an absolute value smaller than one. In other words, all poles must be located within a unit circle in the z-plane.The poles are defined as the values of z which make the denominator of H(z) equal to 0:

Clearly, if then the poles are not located at the origin of the z-plane. This is in contrast to the FIR filter where all poles are located at the origin, and is therefore always stable.IIR filters are sometimes preferred over FIR filters because an IIR filter can achieve a much sharper transition region roll-off than FIR filter of the same order.IIR filters are the most general class of linear digital filter. Their output at a given time step depends on their inputs and on previously computed outputs (i.e., they have memory). IIR filters are recursive, and FIR filters are non-recursive. The output of a IIR filter is formed in the data sequence domain as a weighted sum according to the Nth order difference equation shown below:

The filter is recursive because the difference equation has feedback. Consequently, the filter's response to an impulse may have infinite duration (i.e, it does not become 0 in a finite time).An IIR filter is modelled in the z domain by its z- domain system function, or transfer function, which is a ratio of polynomials formed as

The z-domain transforms of the output time sequences are related by

The tap coefficients of the IIR filter form the sets of the filter's tap coefficients, { } and { }, commonly referred to as feedback and feedforward coefficients, respectively. The parameters N are the order of the filter; it specifies the numberof prior samples of the output that must be saved to form the current output; it also determines the latency of the output.

III. PIPELINING TECHNIQUE

Pipelining of Multiplication[10-12]

In a filter the multiplication of a signal by a constant filter coefficient is the most time-consuming operation. By revising the conditions of shifts and additions the pipelining of the multiplications is achieved. The constant is constituted in canonical signed digit (CSD) format as to minimize the amount of shift-and-add operations for a constant multiplication.Like an binary format the CSD is represented with the difference that each digit might have an value of 0,1, or -1 (represented here as 1). Representing a constant filter coefficient in CSD significantly scales down the number of shift-and-add operations demanded to perform the multiplication by that coefficient. For example, the binary value as "11111010" stood for the decimal value 250 and in CSD as "100001010". Since the binary representation has six non-zero digits, multiplication by 250 demands the addition of six terms when a binary representation is applied:

$u \times 250 = u \times 2^1 + u \times 2^3 + u \times 2^4 + u \times 2^5 + u \times 2^6 + u \times 2^7$

The CSD representation, with only three non-zero digits, calls for the addition (or subtraction) of only three terms:

$u \times 250 = u \times 2^1 - u \times 2^3 + u \times 2^8$

In a binary tree of ripple carry adders the additions themselves are coordinated. As a whole, if A denotes the number of non-zero digits used to constitute the constant, the number of levels in the binary adder tree is given by M= [$\log_2 A$]. For our example, the CSD format demands only two levels where as the binary representation leads in a multiplier with three levels. The multiplier coefficient represented by using a value of A has a two-fold effect on the filter implementation. 1st, it checks the number of adders required for the multiplier itself. 2nd it checks the number of levels M in the consequent tree of adders, which can cause an effect on the structure of the filter. Since lower values of M, a given system throughput perhaps attained with less pipelining. This successively means that fewer registers are demanded in the reformulated system, and a lower-order filter, may be implemented. Hence applying the CSD representation to minimize A not just brings down the number of adders in each multiplier, but as well reduces the amount of multipliers required to implement the filter.

Look-Ahead Filter Forms

The reformulation of the filter in a look-ahead filter form demands the pipelining of the feedback loop in an IIR filter. Here it is exemplify by the process upon a second-order digital filter constituted by the transfer function

$$G(z) = \frac{az^2 + bz + c}{z^2 + dz + e}$$

and represents the difference equation

$$y(k) = au(k) + bu(k-1) + cu(k-2) - dy(k-1) - ey(k-2)$$

In an look-ahead form of the abovedifference equation, the term $y(k-1)$ is rewritten in terms by older values of y, such $y(k-2)$ and $y(k-3)$ which are variable earliest. As, for example, $y(k-2)$ is available one clock cycle earlier than $y(k-1)$, the reformulation allows for the insertion of one another level of pipelining in the feedback computation. Not every look-ahead forms preserve the stability of the original filter. One that make so is the Scattered Look-Ahead (SLA) form Beginning from the transfer function of the original filter in (above equation the transfer function of the SLA filter comprises

$$G_s(z) = \frac{az^2 + bz + c}{z^2 + dz + e} \cdot \frac{z^2 - dz + e}{z^2 - dz + e}$$

$$= \frac{\hat{a}z^4 + \hat{b}z^3 + \hat{c}z^2 + \hat{f}z + \hat{g}}{z^4 + \hat{d}z^2 + \hat{e}}$$

where $\hat{a} = a, \hat{b} = b - ad, \hat{c} = ae + c - db, \hat{d} = 2e - d^2, \hat{e} = e^2, \hat{f} = be - cd, and\ \hat{g} = ce.$

Two newer poles have been acquainted; these poles feature the same magnitudes as the original poles, and alter by them alone in their angles. This implies that the SLA form is stable when the original system was stable (all poles inside the unit circle). The transfer function equates to the difference equation

$$y(k) = \hat{a}u(k) + \hat{b}u(k-1) + \hat{c}u(k-2) + \hat{f}u(k-3) + \hat{g}u(k-4) - \hat{d}y(k-2) - \hat{e}y(k-4)$$

This is to be noted that in the difference equation (x), $y(k-1)$ has been eliminated, and $y(k)$ is calculated from $y(k-2)$ and $y(k-4)$. It is likewise possible to eliminate $y(k-2)$, and so forth, more levels of pipelining are demanded.

IV. DA FUNDAMENTAL

Distributed Arithmetic[8]

Distributed Arithmetic (DA) is a different approach for implementing digital filters. The basic idea is to replace all multiplications and additions by a table and a shifter-accumulator. DA relies on the fact that the filter coefficients are known, so multiplying c[n]x[n] becomes a multiplication with a constant. This is an importance difference and a prerequisite for a DA design.Distributed Arithmetic (DA) can be used to compute sum of products. Many DSP algorithms like convolution and correlation are formulated in a sum of products (SOP) fashion. Consider the following sum of products:

$$y = \langle c, x \rangle = \sum_{n=0}^{N-1} c[n]x[n] = c[0]x[0] + c[1]x[1] + \ldots + c[N-1]x[N-1]$$

Further assume that the coefficients c[n] is known values and that the variable x[n] can be represented by

$$x[n] = \sum_{b=0}^{B-1} x_b[n] \times 2^b \text{ with } x_b[n] \in [0,1],$$

where $x_b[n]$ represents the b[th] bit position of the number's binary representation. The SOP can be represented as:

$$y = \langle c, x \rangle = \sum_{n=0}^{N-1} c[n] \sum_{b=0}^{B-1} x_b[n] \times 2^b$$

Expanding the summations yields to:

$$y = \langle c, x \rangle = c[0] \times \left(x_{B-1}[0]2^{B-1} + x_{B-2}[0]2^{B-2} + \ldots x_0[0]2^0\right)$$
$$+ c[1] \times \left(x_{B-1}[1]2^{B-1} + x_{B-2}[1]2^{B-2} + \ldots x_0[1]2^0\right)$$
$$\vdots$$
$$+ c[N-1] \times \left(x_{B-1}[N-1]2^{B-1} + x_{B-2}[N-1]2^{B-2} + \ldots x_0[N-1]2^0\right)$$

In more compact form:

$$y = \langle c, x \rangle = \sum_{b=0}^{B-1} 2^b \sum_{n=0}^{N-1} c[n] \times x_b[n]$$

The key is to realize that the second summation can be mapped to a Look Up Table (LUT). The coefficients c[n] are known and the $x_b[n]$ values are either 1 or 0 then each SOP is just a combination of the c[n]'s for which a true table can be constructed. Suppose we have:

$$\left(c[0]x_{B-2}[0] + c[1]x_{B-2}[1] + \ldots c[N-1]x_{B-2}[N-1]\right) \times 2^{B-2}$$

Where each x_{B-2} digit belongs to a different x[n] variable.Multiplication by a power of 2 is no more that a bit shift, so what need to do is to slice and concatenate the bits of the different x[n] in order to build a table given that the c[n] are all known. What is left is to show how we can deal with signed implementations of DA. A minor modification needs to be introduced when working with signed two's complement numbers. In two's complement, the MSB is used to determine the sign of the number. We use, therefore, the following B-bit representation:

$$x[n] = -2^{B-1} \times x_{B-1}[n] + \sum_{b=0}^{B-2} x_b[n] \times 2^b$$

Then, the output y[n] is defined by:

$$y[n] = -2^{B-1} \times \sum_{n=0}^{N-1} c[n] \times x_{B-1}[n] + \sum_{b=0}^{B-2} 2^b \times \sum_{n=0}^{N-1} c[n] \times x_b[n]$$

Finally, a block diagram for the DA implementation of a FIR filter is shown in figure 4

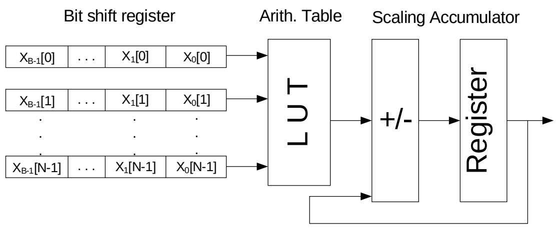


figure 4: block diagram for the DA implementation of a FIR filter

Design & Implementation Flow for Digital Filters

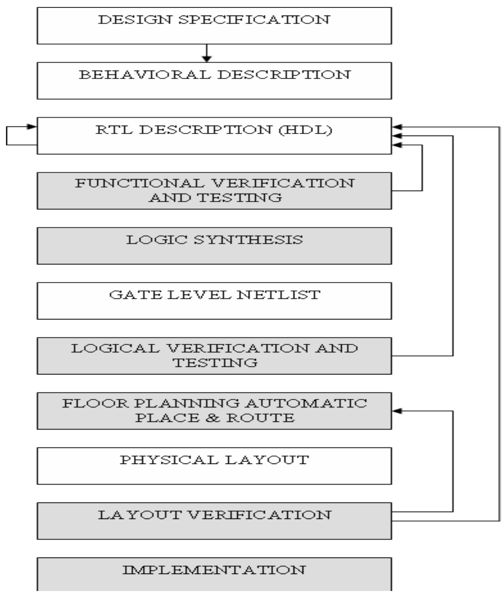


Fig 5 Typical Design Flow of FPGA[9]

A typical design flow for designing for VLSI circuits. Unshaded blocks show the level of design representation shaded blocks show process in the design flow.

V. CONCLUSION

In this paper we have reviewed the performance of FIR filter using pipelined and DA algorithms. It is concluded that better performance level is achieved using DA algorithm FIR Filter using DA algorithm has minimum delay and can operate over high frequency ranges as compared to simple FIR without DA algorithm.

REFERENCES

[1] N. Karaboga, A. Kalinli, D. Karaboga "Designing digital IIR filters using ant colony optimisation algorithm" Engineering Applications of Artificial Intelligence 17 (2011) 301–309, Elsevier, 2011.

[2] V.M. Poucki, A. Zemva, M.D. Lutovac, T. Karcnik "Elliptic IIR filter sharpening implemented on FPGA" Digital Signal Processing 20 (2010) 13–22,Elsevier,2010.

[3] ZhenbinGao , XiangyeZeng , Jingyi Wang , Jianfei Liu "Fpga Implementation Of Adaptive iir Filters With Particle Swarm Optimization Algorithm" School of Information Engineering, Hebei University of Technology Tianjin 300401, P. R. China

[4] Sami Khorbotly, Joan E. Carletta, Robert J. Veillette "A Methodology for Implementing Pipelined Fixed-Point Infinite Impulse Response Filters" 41st South eastern Symposium on System Theory University of Tennessee Space Institute Tullahoma, TN, USA, March 15-17, 2009

[5] Mohamed Al Mahdi Eshtawie, and Masuri Bin Othman " An Algorithm Proposed for FIR Filter Coefficients Representation" D. International Journal of Applied Mathematics and Computer Sciences 4;1 2008.

[6] L. Zhao, W. H. Bi, F. Liu, "Design of digital FIR bandpass filter using distributed algorithm based on FPGA," Electronic Measurement Technology,vol. 30, pp.101-104, 2007.

[7] K.N. Macpherson and R.W. Stewart "Area efficient FIR filters for high speed FPGA implementation" IEEE Proc.-Vis. Image Signal Process., Vol. 153, No. 6, December 2006

[8] T. Salim, J. Devlin, J. Whittington, M. I. Bhatti **"**An efficient serial distributed arithmetic algorithm for FPGA implementation of digital up conversion**"** ComplexityVolume 11,Issue 1, Pages: 24-29, 2005

[9] Mariusz Rawski, Pawel Tomaszewicz, Henry Selvaraj, and Tadeusz Luba, "Efficient Implementation of Digital Filters with Use of Advanced Synthesis Methods Targeted FPGA Architectures", 8th Euromicro conference on digital system design (DSD'05), IEEE Computer Society, 2005.

[10] Chao-Huang Wei, Hsiang-Chieh Hsiao, Su-Wei Tsai "FPGA Implementation of FIR Filter with smallest Processor" IEEE, 2005

[11] L. Narinder,B. Taylor, Designing control circuits for high performance DSP systems, DSP 55–57, 2005

[12] A. Shaw and M. Ahmed, "Pipelined recursive digital filters: a general look-ahead scheme and optimal approximation," IEEE Trans. on Circuits and Systems II: Analog & Digital Signal Processing, vol. 46, no. 11, pp. 1415– 1420, Nov. 1999.

Network Security by Using IP (Internet Protocol)

Minakshi[1], Tanu Preet Singh[2]

[1]Lecturer Department of Computer and Science, Sai Institute of Engineering and Technology, Amritsar.
[2]Associate Professor ,Department of Computer Science & Engineering,Amritsar College of Engineering & Technology,Amritsar.
Email: er.mina08@yahoo.com,tanupreet.singh@gmail.com

***Abstract* - Network security is a complicated subject. It is tackled by well-trained and experienced experts. However, as more and more people become "wired", an increasing number of people need to understand the basics of security in a networked world. This document was written with the basic computer user and information systems manager in mind, explaining the concepts needed to read through the hype in the market place and understand risks and how to deal with them.**

I. INTRODUCTION TO NETWORKING

A basic understanding of computer network is in order to understand the principles of network security. A network has been defined as a set of interlinking lines resembling a net, a network of alliances. This definition suits our purpose well where computer network is simply a system of interconnected computers[1]. Authentication Header is to ensure authentication and integrity for IP data gram packets and to provide protection against replay attacks. It is mainly related to authentication process.

The AH protocol is dealt with different types of algorithms, like Message Digest (MD) 5 that produces a data representation which is 128 bit fixed size long and unique and it will be used for authentication. There is a Sequence Number of 32-bit long field that denotes values used as counter, which is used to give protection from replay attack. The format of IP AH header is given in Figure 4. The 'Next' field is an 8-bit long that is used to identify the type of the next payload. The Payload has length of 8-bit. There is a 6-bit reserved field for future purpose use. The Security Parameter Index, which is called SPI, is 32-bit long value that denotes the Security Association (SA) for this datagram, which is unique. Authentication Data field as shown in figure 4t is variable length field and it has the Integrity Check Value (ICV). For better understanding, Figure 4 also illustrates how an authenticated packet format will change in tunneling. ESP deals with different authentication and encryption algorithms, and fig 5 shows the use of the DES-CBC transform. After packet is encrypted, only authenticated and authorized users could decrypt it. The Internet Key Exchange (IKE) mechanism or protocol is used to exchange or negotiate some important parameters and finalize keys between two communicating nodes and the setup of a Security Associations (SA). It is a one sided agreement between two entities. Fig6 shows the table of SA.

There are different kinds of firewalls exist for secure communication. Generally, it is implemented in security gateways. The most sophisticated and important part for mobile IP used as a firewall, which is known as secure tunneler, which is as shown in Figure 7. This firewall is using AH and ESP protocols mentioned above.

MOBILE IP IN CAMPUS INTRANET

We are taking campus intranet as security model to understand different things of Mobile IP attacks and security. In this part we are using the network security model shown in Figure 8 below. We have to make some assumptions like a network having no links to the Internet, and no firewalls installed anywhere with secure access.

Mobile nodes and network itself are quite vulnerable to attack from insiders-in many cases they are own employees of the company, and perform the malicious purposes. Here security threats of the intranet discussed above also included the view with attack to the particular mobile node, which is outside of the intranet i.e. in public network. So protection for mobile node as well as how this mobile node accesses the intranet in secure manner is discussed in this section. We will then discuss the problem of mobile nodes receiving packets after passing the firewall and they are located outside of the private network i.e. somewhere in public network or Internet.

TCP/IP: The Language of the Internet

TCP/IP (Transport Control Protocol/Internet Protocol) is the "Language" of the Internet. Anything that can learn to "Speak TCP/IP" can play on the internet. This is functionality that occurs at the Network (IP) and Transport (TCP) layers in the ISO/OSI reference model.

Consequently, host that has TCP/IP functionality (Such as UNIX, Os/2, MacOS, or Window NT) can easily support applications (Such as Netscape's Navigator) that use the network[2][3].

Open Design

One of the most important features of TCP/IP is not a technological one: The protocol is an "open" protocol, and anyone who wishes to implement it may do so freely. Engineers and scientist from all over the world participate in the IETF (Internet Engineering Task Force) working groups that design the protocols that make the Internet work[4] .Their time is typically donated by their companies,
and the result is work that make the internet work. Their time is typically donated by their companies, and the result is work that benefits everyone.

IP (Internet Protocol):

IP is network layer protocol. This is the layer that allows the host to actually "talk" to each other. Such things as carrying datagram, mapping the internet address (such as 10.2.3.4) to physical network address (such as 08:00:69:0a:ca:8f), and routing, which takes care if making sure that all of the devices that have Internet connectivity can find the way to each other. IP has number very important features which make it an extremely robust and flexible protocol [6]. For our purposes, though, we're going to focus on the security of IP, or more specifically, the lack thereof.

Attacks against IP

A number of attacks against IP are possible. Typically, these exploit the fact that IP doesn't perform a robust mechanism for authentication, which is proving that a packet came from where
it claims it did. A packet simply claims to originate from a given address, and there is not a way to be sure that the host that sent the packet is telling the truth. This is not necessarily a weakness but it is an important point, because it means that the facility of host authentication has to be provided at a higher layer on the ISO/OSI reference model. Today, applications that require strong host authentication (such as cryptographic applications) do this at the application layer.

IP Spoofing

This is where one host claims to have the IP address of another. Since many systems (such as router access control lists) define which packets may and which packets may and which packets may not pass based on the sender's IP address, this is a useful technique to attacker : he can send packets to a host. Perhaps causing to take some sort of action.

IP session Hijacking

This is very dangerous, however, because there are now toolkit available in the underground community that allow otherwise unskilled bad-guy-wannabes to perpetrate this attack. IP Session Hijacking is an attack whereby a user's session is taken over, being in the control of the attacker. If the user was in the middle of email, and then cab execute any commands he wishes as the attacked user. The attacked user simply sees his session dropped, and may simply login again, perhaps not even noticing that the attacker is still logged in the doing things.

Risk Management: The Game of Security

It is very important to understand that in security, one simply cannot say "what are the best firewalls?" There are two extremes: absolute security and absolute access. The closes we can get to an absolutely secure machine is one unplugged from the network, power supply, locked in a safe, and thrown at the bottom of the ocean, Unfortunately, this is not terribly practical, either: the internet is a bad neighborhood now, and it is not long before some bonehead will tell the computer to do something like self-destruct, after which, it is not terribly useful to you.

This is no different from our daily lives; we constantly make decisions about what risks we are willing to accept. When we get in a car and drive to work, there is a certain risk that we are taking. It is possible that something completely out of control will cause us to become part of an ancient on the highway. When we get on an airplane, we are accepting the level of risk is, and would not go beyond that in most circumstances. If I happen to be upstairs at home, and want to leave for work, I am not going to jump out the window. Yes , it would be more convenient, but the risk injury outweighs the advantage of convenience.

Every organization needs to decide for itself where between the two extremes of total security and total access they need to be. A policy needs to articulate this, and then define how that will be enforced with practices and such. Everything that is done in the name of security, then, must enforce that policy uniformly.

SECURITY MANAGEMENT ISSUES

• Ensuring the security strength of the organization is a big challenge nowadays. Organizations have some pre-defined security policies and procedures but they are not implementing it accordingly. Through the use of technology, we should impose these policies on people and process.

• Building and affirming high-quality resources for deployment and efficient management of network security infrastructure.

• Adopting technologies that are easy and cost effective to deploy and manage day-to–day network security operations and troubleshoots in the long run.

• Ensuring a fully secure networking environment without degradation in the performance of business applications.

• On a day-to-day basis, enterprises face the challenge of having to scale up their infrastructure to a rapidly increasing user group, both from within and outside of the organizations. At the same time, they also have to ensure that performance is not compromised.

• Organizations sometimes have to deal with a number of point products in the network. Securing all of them totally while ensuring seamless functionality is one of the biggest challenges they face while planning and implementing a security blueprint.

• The implementation and conceptualization of security blueprint is a challenge. Security is a combination of people, processes, and technology; while IT managers are traditionally tuned to address only the technology controls.

Network Security cuts across all functions and hence initiative and understanding at the top level is essential. Security is also crucial at the grassroots level and to ensure this, employee awareness is a big concern. Being update about the various options and the fragmented market is a challenge for all IT managers. In the security space, the operational phase assumes a bigger importance. Compliance also plays an active role in security; hence the business development team, finance, and the CEO's office have to matrix with IT to deliver a blueprint.

VoIP(Voice Over IP) Threats

As a starting point, we use the taxonomy provided by the Voice over IP Security Alliance (VoIPSA) [8]. VoIPSA is a vendor-neutral, not for profit organization composed of VoIP and security vendors, organizations and individuals with an interest in securing VoIP protocols, products and installations. In addition, we place the surveyed vulnerabilities within the traditional threat space of confidentiality, integrity, availability (CIA). Finally, we consider whether the vulnerabilities exploit bugs in the protocol, implementation or system configuration. In future work, we hope to expand the number of views to the surveyed vulnerabilities and to provide more in-depth analysis. The VoIPSA security threat taxonomy defines the security threats against VoIP deployments, services, and end users. The key elements of this taxonomy are:

- *Social threats* are aimed directly against humans. For example, configurations, bugs or bad protocol interactions in VoIP systems may enable or facilitate attacks that misrepresent the identity of malicious parties to users. Such attacks may then act as stepping stones to further attacks such as phishing, theft of service, or unwanted contact (spam).

- *Eavesdropping, interception, and modification threats* cover situations where an adversary can unlawfully and without authorization from the parties concerned listen in on the signaling (call setup) or the content of a VoIP session, and possibly modify aspects of that session while avoiding detection. Examples of such attacks include call re-routing and interception of unencrypted RTP sessions.

- *Denial of service threats* have the potential to deny users access to VoIP services. This may be particularly problematic in the case of emergencies, or when a attack affects all of a user's or organization's communication capabilities (*i.e.*, when all VoIP and data communications are multiplexed over the same network which can be targeted through a attack). Such attacks may be VoIP-specific (exploiting flaws in the call setup or the implementation of services), or VoIP-agnostic (*e.g.*, generic traffic flooding attacks). They may also involve attacks with physical components (*e.g.*, physically disconnecting or severing a cable) or through computing r other infrastructures (*e.g.*, disabling the DNS server, or shutting down power).

- *Service abuse threats* covers the improper use of VoIP services, especially (but not exclusively) in situations where such services are offered in a commercial setting. Examples of such threats include toll fraud and billing avoidance [9, 10]
- *Physical access threats* refer to inappropriate/unauthorized physical access to VoIP equipment, or to the physical layer of the network.
- *Interruption of services threats* refer to non-intentional problems that may nonetheless cause VoIP services to become unusable or inaccessible. Examples of such threats include loss of power due to inclement weather, resource exhaustion due to over-subscription, and performance issues that degrade call quality.

Types and Sources of Network Threats

Now, we have covered enough background information on networking that we can actually get into the security aspects of all of this. First of all, we shall get into the types of threats there are against networked computers, and then some things that can be done to protect you against various threats.

Focus on security

The Network security program emphasizes to secure a network. The following background information in security helps in making correct decisions. Some areas are concept-oriented[11]:

- Attack Recognition: Recognize common attacks. Such as spooling, man-in-middle, (distributed) denial of service, buffer overflow etc.
- Encryption Technique: Understand technique

CONCLUSION

Security is a very difficult topic. Everyone has a different idea of what "security" is , and what levels of risk are acceptable. The key for building a secure network is to define what security means to your organization, once that gas been define, everything that goes on with the network can be evaluated with respect to that policy. Project and systems can then be broken down into their components, and it becomes much simpler to decide whether what is proposed will conflict with your security policies and practices.

Many people pay great amount of lip service to security, but do not want to bother with it when it gets in their way. It is important to build systems and networks in such a way the user is not constantly reminded of the security policies and systems too restrictive will find ways around them. It is important to get their feedback to understand what can be improved. Security is everybody's business, and only with everyone's cooperation, an intelligent policy, and consistent practices, will it be achievable.

REFERENCES

[1] Data communication and Networking author Farozon.

[2] The book named Networking author tenaboum.

[3] The new Lexicon Webster's Encyclopedic Dictionary of the English Language, New York: Lexicon.

[4] S.M. Bellovin, security problems in the TCP/IP protocol suite, computer communication Review, volume 19, no.2, pp.32-48, April 1989.

[5] Y. Rekhter, R. Moskowitz D.Karrenberg G.de Groot, E. Lear, "Address Allocation for private internets." RFC 1918.

[6] J.P Hollbrook, J.K. Reynolds. "Site Security Handbook". RFC !244

[7] M. Cutin, "snake Oil Warning signs: Encryption software to Avoid."

[8] The Register. Two charged with VoIP fraud. http://www.theregister.co.uk/ 2006/06/08/voip fraudsters nabbed/, June 2006.

[9] The Register. Fugitive VOIP hacker cuffed in Mexicohttp://www.theregister.co.uk/2009/02/11/fugitive voip hacker arrested/, February 2009.

[10] P. Thermos and A. Takanen. *Securing VoIP Networks*. Pearson Education, 2008.

[11] VoIP Security Alliance. VoIP Security and Privacy Threat Taxonomy, version 1.0. http: //www.voipsa.org/Activities/taxonomy.php, October 2005

3DTV Technology Active Shutter 3D Glasses & Immersive Media

Aman Parkash Singh
Amritsar College of Engg & Technology Amritsar, India
Er.amanparkash@gmail.com

***Abstract*—Three-dimensional TV is expected to be the next revolution in the history of television. We implemented a 3D TV prototype system with real-time acquisition, transmission, and 3D display of dynamic scenes. We developed a distributed, scalable architecture to manage the high computation and bandwidth demands. Our system consists of an array of cameras, clusters of network-connected PCs, and a multi-projector 3D display. Multiple video streams are individually encoded and sent over a broadband network to the display. The 3D display shows high-resolution (1024 × 768) stereoscopic color images for multiple viewpoints without special glasses. We implemented systems with rear-projection and front-projection lenticular screens. In this paper, we provide a detailed overview of our 3D TV system, including an examination of design choices and tradeoffs**

***Index Terms*—3D, Video Stream, Rear-projection,3D Glasses.**

INTRODUCTION

The survey paper discusses the future of immersive telecommunication, especially video conferencing techniques needed to support such systems. A 3D or 3D (threedimensional) film or S3D (stereoscopic 3D) f ilm is a motion picture that enhances the illusion of depth perception. Derived from stereoscopic photography, a regular motion picture camera system is used to record the images as seen from two perspectives (or computer-generated imagery generates the two perspectives in post-production), and special projection hardware and/or eyewear are used to provide the illusion of depth when viewing the film. 3D films are not limited to feature film theatrical releases; television broadcasts and direct-to-video films have also incorporated similar methods, especially since 3D television and Blue-ray 3D. While attending Harvard University, Edwin H. Land conceived the idea of reducing glare by polarizing light. He took a leave of absence from Harvard to set up a lab and by 1929 had invented and patented a polarizing sheet In 1932, he introduced Polaroid J Sheet as a commercial product. While his original intention was to create a filter for reducing glare from car headlights, Land did not underestimate the utility of his newly-dubbed Polaroid filters in stereoscopic presentations.

In January 1936, Land gave the first demonstration of Polaroid filters in conjunction with 3D photography at the Waldorf-Astoria Hotel.The reaction was enthusiastic, and he followed it up with an installation at the New York Museum of Science. It is unknown what film was run for audiences at this exhibition[1]

In Real Life

Close one eye and take a walk through your home or around the block. Your sense of depth diminishes and the scene flattens out. Open both eyes, and you see your surroundings in stereo, with each eye perceiving everything from a slightly different angle (see Fig. 1). The difference between each eye's view is greatest for objects close up and tapers off for those farther away. Using that information, your brain calculates distance, helping you perceive depth.
To create a sense of depth onscreen, each eye sees video shot from slightly different angle, corresponding to the average distance between human eyes. In animated films, these "shots" come from computer models that generate the two views. In live action, two cameras are used.[2]

At the Movies

Old-fashioned 3D movies used color to separate the two images, but current theaters take advantage of polarization, a property of light rays. Though we think of light traveling in a straight line, it can actually wiggle up and down or side to side, or spin in a corkscrew manner. Polarizing filters allow just one type of light to pass.

In a theater, the 3D projector uses these filters to project two images, a right-eye perspective displayed with clockwise-polarized light and a left-eye perspective with counterclockwise light. The audience wears polarized glasses that allow only the proper polarity to reach each eye (see Fig. 2). The brain assembles the separate perspectives into a view that resembles real life.[3]

In the Living Room

3D works a bit differently at home. 3D-capable LCD or plasma TVs rapidly flash alternating left-eye and right-eye video frames. But instead of polarizing filters, your glasses have "lenses" that are actually tiny LCD panels.

These glasses are known as active shutter glasses. As a right-eye video frame flashes on the screen, the LCD over the right eye switches from opaque to clear. The right-eye

LCD darkens again and the left-eye LCD becomes clear when the left-eye video frame appears (see Fig. 3). At any moment, you see only one perspective, through one eye. But the left/right video images alternate so quickly— 120 times per second — that you perceive a full, three-dimensional view.[4]

Free Your Eyes

Glasses can produce convincing 3D imagery, but they can be awkward to wear. Next-gen 3DTVs will ditch the glasses and put lenses right on the screen. In one setup, the TV divides the left- and right-eye perspectives into alternating vertical columns. To picture this, imagine slicing two photographs into thin vertical strips and alternating them: one from the right photograph, one from the left, etc. Microscopic lenses over the screen bend the light so that slices of the right-side perspective reach your right eye, and slices of the left-side perspective reach your left eye (see Fig. 4).

But what if you aren't sitting directly in front of the screen? The new TVs will use several sets of lenses to create multiple pairs of right- and left-eye views. One set of images is for viewers directly in front of the screen. Another is for viewers way to the side of the screen, and additional pairs take care of all the viewers in between. If you do move from side-to-side, you simply transition from one pair of right-left images to the next.

The technology is very challenging. But when it's ready, TV will look nearly as real as life itself.[5]

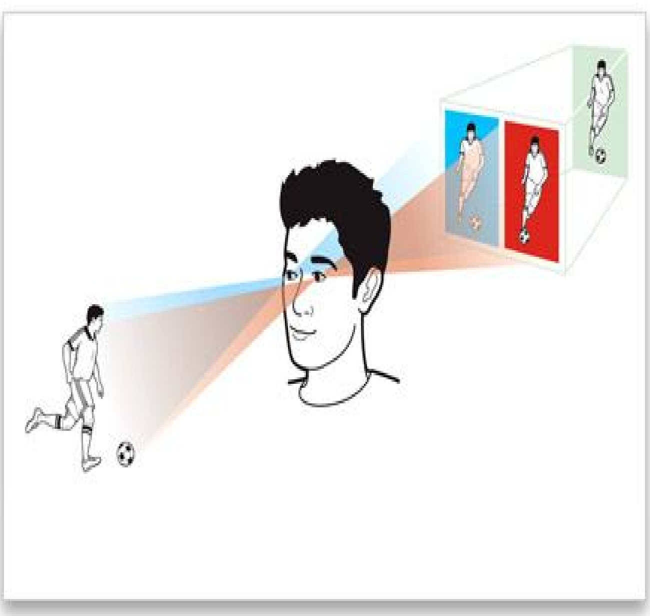


Fig. 1. In real life, each eye sees the world from a slightly different angle. Your brain compares the difference between the two perspectives to produce an impression of depth.

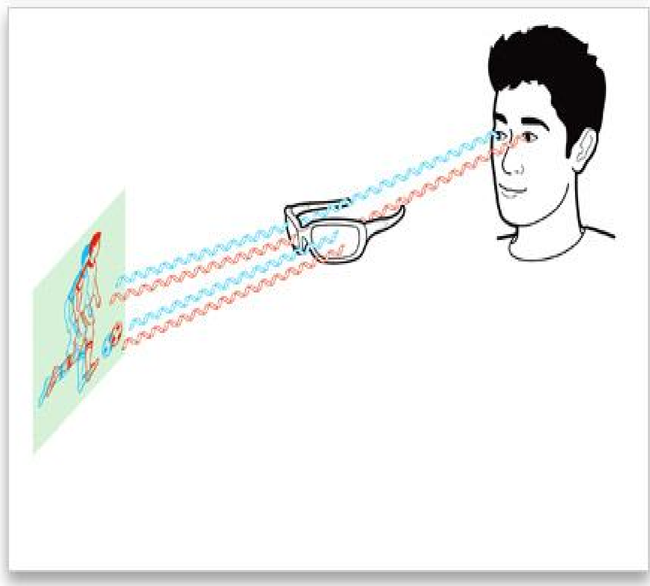


Fig. 2. In a 3D movie theater, the screen displays images shot from two different perspectives, each using light with a different polarization. Polarized glasses filter just one perspective to each eye

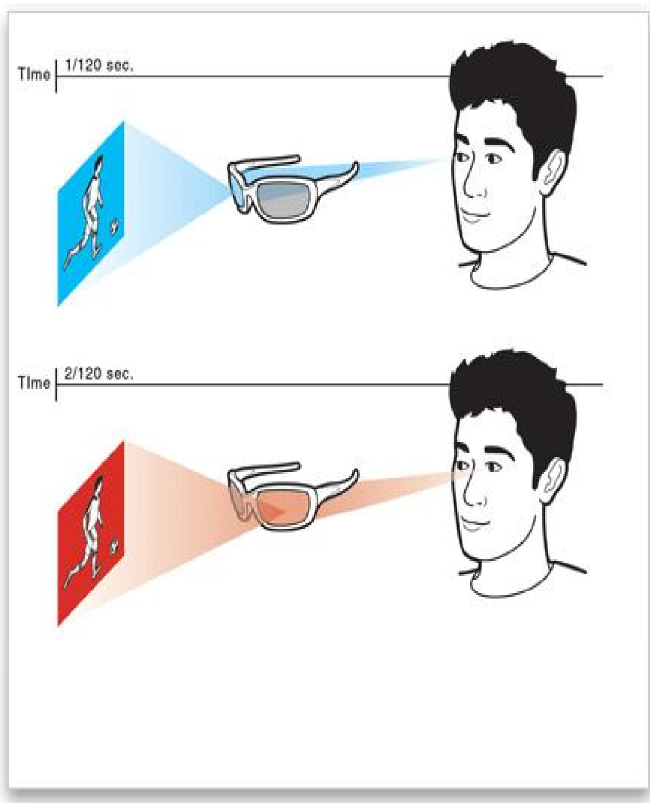


Fig. 3. 3D TVs alternate rapidly between images shot from two different perspectives. In active shutter glasses, LCDs over each eye alternate between clear and opaque in synch with the TV. You see through only one eye at any given moment, but the alternation happens fast enough that you perceive a single 3D image.

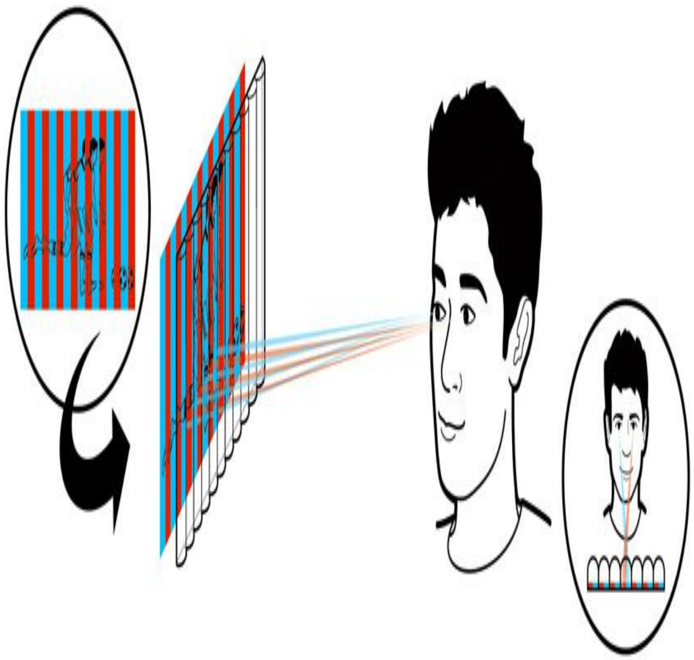


Fig. 4.Engineers are developing 3D TVs that work without glasses. Instead, they use lenses in front of the screen itself that direct the proper portion of the image to each eye

Active shutter 3D system

An active shutter 3D system (a.k.a. alternate frame sequencing, alternate image,alternating field, field sequential or eclipse method) is a technique of displaying stereoscopic 3D images. It works by openly presenting the image intended for the left eye while blocking the right eye's view, then presenting the right-eye image while blocking the left eye, and repeating this so rapidly that the interruptions do not interfere with the perceived fusion of the two images into a single 3D image.

An active shutter 3D system generally uses liquid crystal shutter glasses (also called LCS glasses, LCS 3D glasses, LC shutter glasses or active shutter glasses) Each eye's glass contains a liquid crystal layer which has the property of becoming dark when voltage is applied, being otherwise transparent. The glasses are controlled by a timing signal that allows the glasses to alternately darken over one eye, and then the other, in synchronization with the refresh rate of the screen. The timing synchronization to the video equipment may be achieved via a wired signal, or wirelessly by either an infrared, radio frequency, Bluetooth or optical transmitter. Active shutter 3D systems are used to present 3D films in some theaters. It can be used to present 3D images on CRT, plasma, LCD and other types of video displays.[6]

Advantages

LC shutter glasses mostly eliminate 3D crosstalk, which is a problem with other 3D display technologies such as linearly polarized glasses. Also, unlike red/cyan colour filter (anaglyph) 3D glasses, LC shutter glasses are colour neutral, enabling 3D viewing in the full color spectrum.

Disadvantages

Flicker can be noticeable except at very high refresh rates, as each eye is effectively receiving only half of the monitor's actual refresh rate. However, modern LC glasses generally work in higher refresh rates and eliminate this problem for most people.

Until recently, the method only worked with CRT monitors; some modern flat-panel monitors now support high-enough refresh rates to work with some LC shutter systems. Many projectors, especially DLP-based ones, support 3D out of the box.

LC shutter glasses are shutting out light half of the time; moreover, they are slightly dark even when letting light through, because they are polarized. This gives an effect similar to watching TV with sunglasses on, which causes a darker picture perceived by the viewer. However, this effect can produce a higher perceived display contrast when paired with LCD displays because of the reduction in backlight bleed. Since the glasses also darken the background, contrast is enhanced when using a brighter image.Frame rate has to be double that of an non-3D, anaglyph, or Polarized 3D systems to get an equivalent result. All equipment in the chain has to be able to process frames at double rate; in essence this doubles the hardware requirements.Shutter glasses are heavier and more expensive than other forms of stereoscopic glasses because they need electronics and batteries. Anaglyph and polarized 3D glasses can be purchased at low prices. Shutter glasses may not fit over prescription glasses.

From brand to brand, shutter glasses use different synchronization methods and protocols. Therefore, even glasses that use the same kind of synchronization system (eg. infrared) will probably be incompatible across different makers. However, efforts are being made to create a Universal 3D Shutter Glass.

Standards

In March 2011 Panasonic Corporation, together with XPAND 3D, have formulated the M-3DI Standard, which aims to provide industry-wide compatibility and standardisation of LC Shutter Glasses. This movement aims to bring about compatibility among manufacturers of 3D TV, computer, notebook, home projection, and cinema with standardised LC shutter glasses that will work across all 3D hardware seamlessly. The current standard is Full HD 3D Glasses.Field Sequential has been used in video games, VHS and VHD movies and is often referred to as HQFS for DVD's, these systems use wired or wireless LCS glasses.The Sensio format was used with DVD's using wireless LCS glasses.[7]

Active shutter 3D system providers

There are many sources of low-cost 3D glasses. IO glasses are the most common glasses in this category. XpanD 3D is a manufacturer of shutter glasses, with over 1000 cinemas currently using XpanD glasses[8]. With the release of this technology to the home-viewer market as of 2009, many other manufacturers are now developing their own LC shutter glasses, such as Unipolar International Limited, Accupix Co., Ltd, Panasonic, Samsung, and Sony.

The M-3DI Standard, announced by Panasonic Corporation together with XPAND 3D in March 2011, aims to provide industry-wide compatibility and standardisation of LC (Active) Shutter Glasses.

Samsung has developed active 3D glasses that are 2 ounces and utilize lens and frame technology pioneered by Silhouette, who creates glasses for NASA[9].

Nvidia makes a 3D Vision kit for the PC; it comes with 3D shutter glasses, a transmitter, and special graphics driver software. While regular LCD monitors run at 60 Hz, a 120 Hz monitor is required to use 3D Vision. DLP Link is a technology by Texas Instruments, used to wirelessly synchronize liquid crystal shutter glasses to 3D DLP devices, such as projectors and TVs that works optical with information between each frame

Three Qualities of Immersive Telepresence:

The goal for immersive telepresence is to reproduce the best characteristics of direct human interaction that result from a face-to-face meeting. This is accomplished when:

*Participants can all see each other on the screen no matter **who is talking***. Unless everyone is visible you miss important non-verbal expressions and reactions that provide important insight. It’s arguable that non-verbal queues are even more telling than verbal communication.

Participants can all see and hear each other in the highest clarity, no matter how they join the call. It should be true-to-life, so you should be able to see the people you are talking to as clearly as if they were sitting across the table from you.

The technology is seamless. Participants shouldn’t be aware of the technology that enables the immersive experience—they should simply be able to go about their business without worrying about how it’s being supported. This includes easy sharing of multimedia and addition and removal of participants from the call.

As with all visual communication tools, telepresence must deliver an interoperable, consistent, predictable and reliable collaboration experience. Done right, it will change the way your business gets done.

Application, Scenarios and Challenges[10]

From presence to Immersive(Generating a three-dimensional image that appears to surround the user)Telepresence

The idea of immersive media is grounded in two basic concepts i.e. presence and immersion.
The structure of presence has been studied for along time in the interdisciplinary field of human factor research.
Although several aspects are still unclear it is commonly agreed that the basic meaning of “presence” can be stated as “being virtually there”.

Immersive Videoconferencing

Effective video conferencing is an important facility for business with geographically distributed operation and high speed computer based videoconferencing is a potential killer application, gathering research efforts from major market players like VTEL Picture Tel, Sony,Teleportec,VCON and other corporate reasons for using videoconferencing systems include business globalizations or decentralization increased competition pressure for higher reactivity and short decision making increase in number of partners and reduced time and travel coasts.

According to a proprietary wainhoure research from March 2001,voice conferencinh and email are still preferred in general, videoconferencing key barriers seem to be high unit prices ,the limited business needs, the cost of ownership,and concerns about integration lack of training and user friendliness.

REFERENCES

[1] en.wikipedia.org/wiki/3D_film

[2] http://us.toshiba.com/tv/research-center/technology-guides/how-3d-tv-works/

[3] http://us.toshiba.com/tv/research-center/technology-guides/how-3d-tv-works/

[4] http://us.toshiba.com/tv/research-center/technology-guides/how-3d-tv-works/

[5] http://us.toshiba.com/tv/research-center/technology-guides/how-3d-tv-works/

[6] http://en.wikipedia.org/wiki/Active_shutter_3D_system

[7] http://www.fullhd3dglasses.com/

[8] http://www.filmjournal.com/filmjournal/content_display/news-and-features/features/technology/e3i8fb28a31928f66a5faa6cce0d6b536cc?pn=2 Aug 14, 2009-By Bill Mead

[9] Samsung crafts ‘world’s lightest’ 3D glasses in partnership with Silhouette

By:-Vlad Savov ,Posted 3 Jan 2011.

[10] 3-D Image Processing in the future of Immersive media

By:-Francesco Isgro.PhD in Computer Science from Heriot Watt University(Scotland) in 2001.

Emanuele Trucco.PhD from the University of Genoa

Peter Kauff .Reviewer of several IEEE and IEE publications.

WDM-PON: A Scalable and Flexible System for High Speed Networks

Malti[1], Amarpreet Singh[2]
[1]Lecturer Sri Sai Engineering College Manawala, Amritsar
[2]AP, ACET, Amritsar
Email:amarmandeep@yahoo.com

***Abstract*-No architecture yet deployed achieves the necessary scalability and flexibility. By running essentially passive fibre connections from the user to central offices that could be 100-150km away, large cost and energy savings are seen to be possible. The simplest architecture conceptually is point-to-point, with a physical fibre for each connection. This choice has been selected for a number of fibre-to-the-home (FTTH) deployments. Such a solution is viable where the cost of fibre deployment is not prohibitive and where the aggregation point is relatively close to the user. Wavelength division multiplexing is the key to the hardware reduction that is required, hence the birth of the 'WDM-passive optical network' or WDM-PON. Such systems can provide the economically advantageous long reach operation that is desired, either by combining WDM with time division multiplexing (WDM-TDM) or by allocating each user a full wavelength channel.**

Keywords- PON,OLT,ONU,AWG,WDM-PON.

I. INTRODUCTION

The latest development in passive optical network (PON) systems is evolving toward higher data rates and long reach transmission distances, where WDM-PONs can be very attractive solutions. Internet traffic worldwide has experienced remarkably rapid growth over the last decade. The numerous studies indicate growth rates of 50-60% per year [1], driven by new services such as peer-to-peer file sharing, social networking and game playing, as well as conventional entertainment. Personal internet use is increasing rapidly and will dominate future growth [2]. These trends have been enabled by the widespread adoption ofbroadband access technology, notably xDSL, with fibre to the home (FTTH). Passive optical networks (PONs) are receiving much interest because they represent the cheapest way to provide fiber to the home [1-3]. PON networks are basically point to multipoint (P2MP) networks, which is best for distributive services. In P2MP networks, the entire transmission capacity is divided among all subscribers for individual internet access [4-8]. The most important aspect of PON architecture is its simplicity. The Optical Line Terminal (OLT) is the main element of the network and it is usually placed in the Local Exchange. Optical Network Units (ONUs) serve as an interface to the network and are deployed on a customer's side. ONUs are connected to the OLT by means of optical fiber and no active elements are present in the link [9] [10].

II. WDM-PON ARCHITECTURE

Wavelength-division-multiplexed passive optical network (WDM PON) offers many advantages such as large capacity, easy management, network security and upgradability [11-15]. The multiple wavelengths of WDM-PONs can be used to separate optical network units into several virtual PONs co-existing on the same physical infrastructures and the wavelengths can be used collectively through statistical multiplexing to provide efficient wavelength utilization and lower delays experienced by the ONUs [16-18].

In most WDM-PON systems, a broadband light source at the central office sends a broadband seed signal into the OLT transmitters to lock their transmission to the correct wavelength as their data are transmitted down the main fiber.

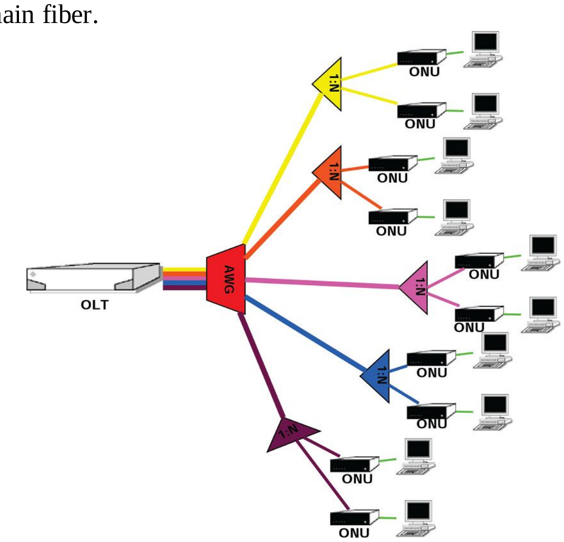


Fig1: Configuration of WDM-PON

At the 32-channel AWG de-multiplexer in the field, this signal gets split into 32 different fibers, one wavelength going to each fiber. WDM-PON networks start by replacing the 1x32 power splitter with a 32-channel athermal AWG. Rather than splitting the optical power between 32 different homes, the athermal AWG splits one wavelength to each home.

A WDM-PON has a dedicated wavelength for each ONU. It has better privacy and better scalability because each ONUs only receives its on wavelength. The wavelength Division Multiplexing Passive Optical Networks are the next generation in development of access networks. Ultimately, they can offer the largest bandwidth at the lowest cost. The term passive simply describes the fact that optical transmission has no power requirements or active electronic parts once the signal is going through the network.

III. ISSUES AND CHALLENGES

The MAC layer of WDM-PON is simplified because the P2P connection between OLT and ONUs are realized in wavelength domain so no P2MP media access is needed. In WDM-PON separate downstream wavelength is used for each subscriber. This in turn provides more bandwidth to each subscriber, more security and better operational control. Also there is no interference in downstream direction. Including all this it has following main features.

- Dedicated bandwidth
- Guaranteed Quality of service
- Physical point-to-multipoint (P2MP)
- Logical point-to-point (P2P)
- Protocol and data-rate transparency
- Simple fault localization
- Better security

Its cost is very high. The AWG filter which is used in WDM-PON is more expensive than splitters which are used in EPON, BPON, and GPON. Temperature control is another challenge because wavelength tends to drift with environmental temperature. Gbps rates of WDM-PON is too large for a single user. Big portion of the bandwidth of one wavelength is wasted. Large number of wavelengths needed, more fibers, more transceivers.

To be economically viable, it is vital that user terminals in a WDM network are all identical, i.e. 'colourless' – inventory management and deployment issues will otherwise dominate installation costs. Various solutions to this problem have been proposed, including systems based on injection-locked lasers at the user site and on reflective modulation.

CONCLUSION

Internet growth is demanding the introduction of highly scalable and flexible systems architectures such as WDM-PON, with attendant demands for high performance, 'colourless' terminals.Tunable laser based ONUs appear viable with current device technology, given appropriate systems architectures. The scalability and flexibility of WDM-PON can be enhanced by using different modulation formats and by varying bit rates, power and extinction ratio. The wavelength division multiplexed passive optical network can be made more flexible and scalable by using various types of tunable lasers.

REFERENCES

[1] Asier Villafranca Velasco, *"Novel WDM-PON Architecture Based on a Spectrally Efficient IM-FSK Scheme Using DMLs and RSOAs"* , journal of lightwave technology, vol. 26, no. 3, february 1, 2008.

[2] Peter Winterling, *"WDM-PON and CWDM network Functions and Measurement tasks"*, JDSU Publications.

[3] David Gutierrez, *"FTTHStandards, Deployment and Research Issues"*, IEEE Communication Magazine,2005.

[4] R. Monnard, *"Demonstration of a 12*155 Mb/s WDM PON Under Outside Plant Temperature Conditions"* , IEEE photonics technology letters, vol. 9, no. 12, december 1997.

[5] Chang-Joon Chae, Member, IEEE, and Nam-Hyun Oh, *"WDM.TDM-PON system employing a wavelength- selective filter and a continuous- wave shared light source"* , IEEE photonics technology letters, vol. 10, no. 9, september 1998.

[6] Carlos Bock, *"Hybrid WDM/TDM-PON using AWG FSR and featuring centralized light generation and dynamic bandwidth allocation"* , journal of lightwave technology, vol. 23, no. 12, december 2005.

[7] N.Calabretta, *"A bidirectional WDM/TDM-PON using DPSK downstream signals and a narrowband AWG"* , IEEE photonics technology letters, vol. 19, no. 16, august 15, 2007.

[8] Hyowon Kim, *"A Cost-Efficient WDM-PON Architecture Supporting Dynamic Wavelength and Time Slot Allocation"*, Feb. 12-14, 2007 ICACT2007.

[9] L. Xu and H. K. Tsang, *"WDM-PON Using Differential-Phase-Shift-Keying Re-modulation of Dark Return-to-Zero Downstream Channel for Upstream"*, IEEE photonics technology letters, vol. 20, no. 10, may 15, 2008.

[10] Yi-Hung Lin, *"Effect of AWG channel bandwidth on BER of injection-locked RSOA based ONU in WDM-PON system"* , International Conference on Photonics in Switching, 2008.

[11] Gong-Ru Lin, Senior Member, IEEE, Member, OSA, *"Comparison on Injection Locked Fabry–Perot Laser Diode With Front-Facet Reflectivity of 1% and 30% for Optical Data Transmission in WDM-PON System"* , journal of lightwave technology, vol. 27, no. 14, july 15, 2009.

[12] Shu-Chuan Lin, Ji-Ying Huang, *"WDM-PON Systems with 10-Gb/s Bidirectional Transmission Using Cross-Remodulation and Dual WavelengthLasers" ,OptoElectronics and Communications Conference,* 2009. OECC 2009.

[13] Chung-Liang Yang, *"Simultaneous ER Enhancement and Light Reuse Scheme for RSOA-Based WDM-PONs"* , IEEE 2009.

[14] Q.T.Nguyen, IEEE *"Bidirectional Transmission in Colourless WDM-PON based on Injection- Locked Fabry-Perot Laser at 2.5 Gbit/s using Low-Cost Seeding Source"* , ECOC 2009, 20-24 September, 2009, Vienna, Austria.

[15] Jianjun Yu, Zhensheng Jia *"Broadband Convergence of 60-GHz ROF and WDM-PON Systems with a Single Modulator for Bidirectional Access Networks"* , ECOC 2009 Vienna, Austria.

[16] Hanlin Feny, Fengqing Liu, *"A novel scheme of multicast WDM-PON using modified NRZ signal format"* , IEEE 2009.

[17] Jun-Ichi *"Enabling technologies for future scalable and flexible WDM-PON system"* , IEEE journal of quantum electronics, vol. 16, no. 5,october 2010.

[18] Quoc Thai Nguyen, Student Member, IEEE, Pascal Besnard, *"Bidirectional 2.5-Gb/s WDM-PON Using FP-LDs Wavelength-Locked by a Multiple-Wavelength Seeding Source Based on a Mode-Locked Laser"* , IEEE photonics technology letters, vol. 22, no. 11, june 1, 2010.

Congestion Control in Optical Burst Switching Networks

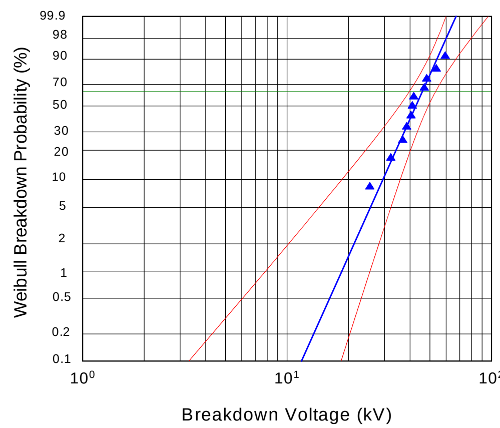
Ashok Kumar[1],Amit Wason[2], Vijay Kumar Banga[3]
PTU, Jalandhar, Punjab, India
HITM, Dheen, Ambala, Haryana
ACET, Amritsar, Punjab, India
Email: ashokcalicut1993@gmail.com, wasonamit13@gmail.com, v_banga@rediffmail.com

***Abstract*-Congestion control is the main issue in Optical Burst Switching Networks now days. Congestion occurs when the reception time is less than transmission time during the transfer of data. In this paper, we had proposed a model to control the congestion in Optical Burst Switching Networks. Simulation results show that with the control of congestion, blocking probability is controlled. When the value of congestion is more, blocking probability increases.**

Keywords-Congestion,Blocking Probability, Traffic, Routing Scheme

I. INTRODUCTION

For eliminating the limitation of all optical networks Optical Burst Switching was proposed. In Optical Burst Switching Networks [7], [8] dynamic wavelength switching of data was done which include optimal scheduling schemes like TAW and TAG [10]. These were performed for the examining the burst loss in OBS network. Another scheduling scheme was JET and JIT [11] for reducing the blocking probability with the help of complex scheduler. Here, contention resolution, QOS support, burst assembly, traffic admission control, performance monitoring, signaling and network resource dimensioning were the main issues for the development in OBS network. Blocking probability is the probability in which connection was not established due to insufficient transmission of data. Blocking Criteria was used when the design of system was based on the fraction of calls blocked. If all the devices were occupied the demand of service was initiated and blocking occurred. For a system designed on a loss basis, a suitable GOS was considered in which the percentage of calls which were lost were recovered. This happened due to unavailability of equipment at the instant of call request.Congestion occurred when the link or node carried large amount of data with which QoS of the system decreases or deteriorates. Main effects include queuing delay, packet loss or the blocking of new connections. Queuing Delay was the time job waits in



a queue until it could be executed. Packet loss occurred when one or more packets of data travelled across a computer network fail to reach their destination. The lost or dropped packets results in noticeable performance issues or jitter with streaming technologies, voice-over-IP, online gaming and video conferencing, and affected all other network applications to a degree. When network load increases, work done decreases called congestive collapse which was caused by spurious retransmissions, undelivered packets, fragments, control traffic and stale or unwanted packets. Hence, to avoid congestive collapse, congestion control concerns controlling traffic entry into a telecommunications network and various congestion control algorithms were defined. Traffic congestion increases when its use increases means when demand increases. It was characterized by slower speeds, longer trip times and increased vehicular queuing. As demand approaches the capacity of a road, it leads to congestion [3]. The demand is directly proportional to traffic intensity. Hence, Traffic Intensity (T) is as $T = \lambda / h$, demand $D = cT = c\lambda / h$.

II. CONGESTION CONTROL ALGORITHMS

To avoid congestion two routing techniques were developed [9]. First, Congestion based dynamic route selection technique using fixed alternate SP routes. Secondly, least congested dynamic route calculation technique with different weight function. After simulation, it performed better than network without any alternate routing scheme and reduces packet loss probability. [6] Studied two adaptive routing strategies, the Multipath (MP) and the Bypass (BP) in addition to SP algorithm. In MP, between each pair of source and destination nodes there were number of pre-established paths. For each burst, routing decision was made by this algorithm. If the burst was not transmitted on first route then higher or equal length burst was selected. Also, the routing algorithm could reroute the path when congestion occurs. In BP, the burst was transmitted through other node by by-pass the congested node. Hence isolatedrouting does not cope with the congestion in OBSnetwork.

For congestion control a novel congestion control scheme called Statistical Additive Increase Multiplicative Decrease (SAIMD) was introduced [5]A) It effectively solved the false congestion detection problems and significantly outperforms the conventional TCP counterparts without losing fairness. Also, to solve the false congestion detection problem in the IP over OBS networks a novel scheme called TCP with Explicit Notification Generalized Additive Increase Multiplicative Decrease (TCP-ENG) was introduced [4]. It was considered best than two proposed TCP enhancements: BAIMD and Burst TCP (BTCP). To calculate blocking probability the Erlang Loss formula was used. When network overloads, no contention resolution scheme avoids the collision and cause blocking.

To reduce the blocking probability two algorithms were proposed, Fault Tolerant Optimized Blocking Algorithm (FTOBA) and Fault Tolerant Least Congestion Algorithm (FTLCA) [2]. On comparing the performance of both techniques, FTLCA performs better than the FTOBA routing algorithm as when number of wavelength increases blocking probability reduces. These algorithms perform better when maximum value of congestion was limited.

To control the congestion, a new algorithm is proposed based on Erlang's Loss Formulae [1] known as Modified Erlang's Loss Formulae. The formula is given by equation (1.1) as:

$$B = \frac{\frac{\left(\frac{C_g}{C_1}\right)^n}{n!}}{\sum_{i=0}^{n} \frac{\left(\frac{C_g}{C_1}\right)^i}{i!}} \qquad (1.1)$$

Where, B= Blocking Probability
C_g= Congestion
C_1= Constant value
n= Number of Servers

III. SIMULATIONS AND RESULTS

The simulations are done in MATLAB software.

The relation between number of servers and blocking probability, and relation between traffic intensity and blocking probability.

(i) For n=30 and ρ= 20

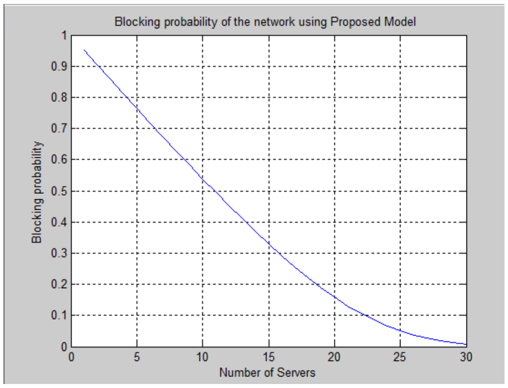


Figure 1.1: Number of Servers v/s Blocking Probability for n= 30 and ρ= 20

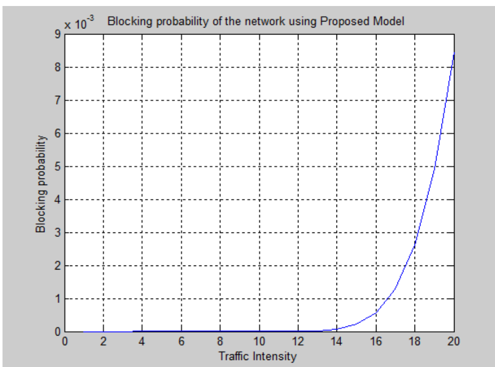


Figure 1.2: Traffic Intensity v/s Blocking Probability for n= 30 and ρ= 20

Figure 1.1 and figure 1.2 show that with increase in number of servers, the blocking probability decreases and with increase in value of traffic intensity, the blocking probability increases. For n=50 and ρ= 100

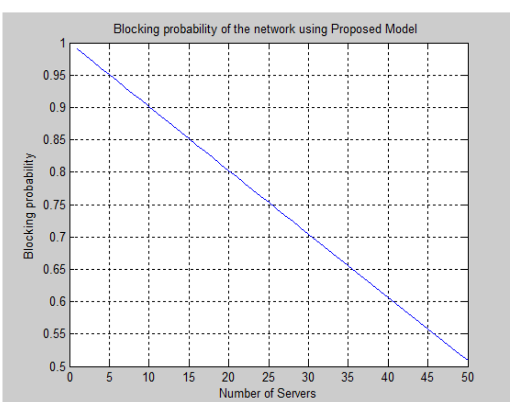


Figure 2.1: Number of Servers v/s Blocking Probability for n= 50 and ρ= 100

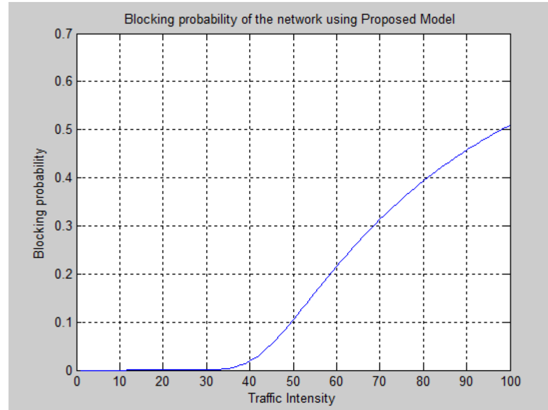


Figure 2.2: Traffic Intensity v/s Blocking Probability for n= 50 and ρ= 100

In figure 2.1 and figure 2.2 depicts that when the number of servers are less than traffic intensity, then there is little change in the performance of the graph. With increase in the value of number of servers, the blocking probability declines linearly. With increase in the value of traffic intensity, the blocking probability remains zero at some value of traffic intensity, then increase exponentially.

B) The relation between congestion and blocking probability is introduced to control the congestion.

(ii)For n=16, C_g=20 and C_1=5

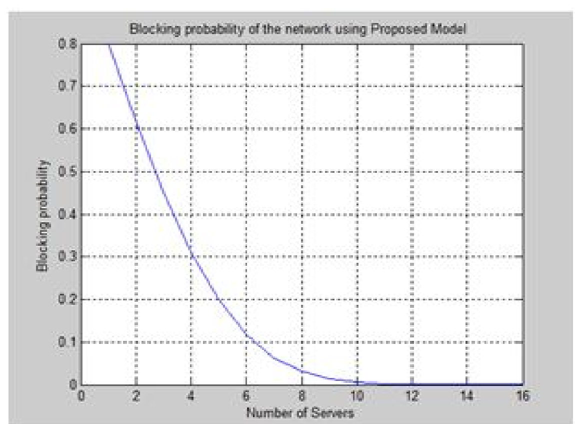


Figure 3.1: Number of Servers v/s Blocking Probability for n= 16, C_g=20 and C_1= 5

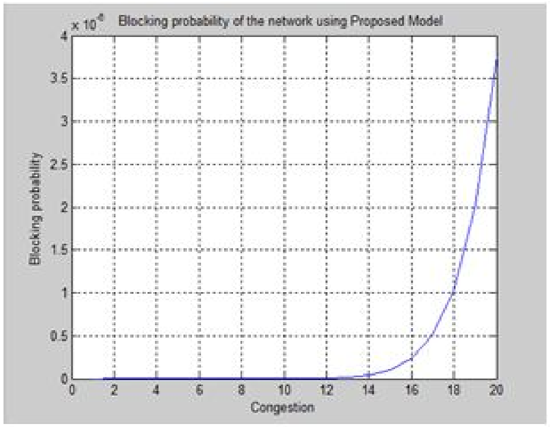


Figure 3.2: Congestion v/s Blocking Probability for n= 16, C_g=20 and C_1= 5

The results in figure 3.1 show that with increase in the number of servers, the blocking probability reduces upto an extent and then becomes approximately zero with further increase. In figure 3.2, the value of congestion is substituted in the formulae. With increase in the value of congestion, the blocking probability increases. Graph showed that, when the value of congestion is less, then blocking probability is approximately zero. When congestion increases, blocking probability increases.

CONCLUSION

Using new developed Modified Erlang's Loss Formulae, congestion in Optical Burst Switching is controlled. This control is done by relating the blocking probability parameter with congestion. From the above figures shown in part (A), with increases in number of servers, blocking probability decreases and with increases in traffic, blocking probability increases. In part (B), with increase in number of servers, blocking probability decreases and with increase in value of congestion blocking probability increases.

REFERENCES

[1] Bonald, T. and Roberts, J.W., "Internet and Erlang Formulae", ACM SIGCOMM Computer Communication Review, Vol. 42, Issue 1, pp. 22-30, 2012.

[2] Singh, H., Prakash, Jai, Arora, D. and Wason A., "Fault Tolerant Congestion Based Algorithms in OBS Network", International Journal of Engineering (IJE), pp. 350-359, Volume 5, Issue 5, Dec 2011.

[3] Iversen, V.B., "Teletraffic Engineering Handbook", June 2011.

[4] Shihada, Basem, Ho, Pin-Han and Zhang, Qiong, "TCP-ENG: Dynamic Explicit Congestion Notification for TCP over OBS Networks", Computer Communications and Networks, ICCCN 2007, pp. 516 – 521, Proceedings of 16th International Conference on 13-16 Aug, 2007.

[5] Shihada, Basem, Ho, Pin-Han and Zhang, Qiong, "SAIMD: A Novel False Congestion Detection Scheme for TCP over OBS Networks", GLOBECOM '07 IEEE, pp. 2428 – 2433, 26-30 Nov, 2005.

[6] Klinkowski, M., Careglio, D., Horta, E. and Pareta, J.S., "Performance Analysis of Isolated Adaptive Routing Algorithms in OBS networks", in Proceedings of e-Photon/ONe Summer School Workshop, Rimini, Italy, August 2005.

[7] Jue,J.P. and Vokkarane, V.M., "Optical Burst Switching Networks", Springer, 2005.

[8] Chen, Y., Qiao, C., and Yu, X., "OBS: A New Area in Optical Networking Research," IEEE Network, Vol. 18, pp. 16–23, 2004.

[9] Thodime, Guru P.V., Vokkarane, V.M., and Jue, J.P., "Dynamic Congestion-Based Load Balanced Routing in Optical Burst-Switched Networks", GLOBOCOM 2003, Global Telecommunications Conference, vol.5, pp. 2628 – 2632, 2003.

[10] Detti, A. and Listanti, M., "Application of Tell & Go and Tell & Wait Reservation Strategies in Optical Burst Switching Networks - A Performance Comparison", in proceedings of IEEE International Conference On Telecommunication (ICT), Vol. 2, 2001.

[11] Wei, J.Y. and McFarland, R.I., "Just-In-Time Signaling for WDM Optical Burst Switching Networks", Journal Of Lightwave Technology, Vol. 18, No. 12, pp. 2019-2037, December 2000.

Implementation of Wireless Power Transmission over Super Smart Grid

Er.Bimal Kumar[1] and Er.Amit Sharma[2]
[1]Associate Professor, Department of Electrical & Electronics Engineering, ACET, Amritsar, India
[2] Department of Electrical & Electronics Engineering, ACET, Amritsar, India

***Abstract*— This paper explains how wireless power transmission is possible and how it can be implemented over concept of super smart grid using ionosphere. Ionosphere can reflect waves of a particular band i.e. waves of frequency ranging from 10 MHz to 30 MHz. IMPATT diode (Fig. 2) can be used for conversion of electrical power into microwaves and vice versa at transmitting end and receiving end respectively. As the super smart grid have features of super grid and smart grid thus it'll provide very efficient, reliable and almost fault free transmission from one point of earth to another without using super conductors or other kind of wires.**

***Keywords*—** IPS – Integrated Power Systems, UPS – Unified Power Systems, CIS – Commonwealth of Independent States, WPT – Wireless Power Transmission, SSG – Super Smart Grid.

I. Introduction

Super smart grid is a wide area electricity network which will connect various renewable electrical power plants of IPS/UPS of CIS countries. IPS/UPS systems of some CIS countries have installed generating capacity of 300 GW, and produce 1200 Twh per year for their 280 million consumers [10]. But unfortunately to transmit this much high power from one place to another appropriate media are not available till date due to which super smart grid is still a hypothetical concept.

This concept can be made practically possible by using wireless power transmission which can be obtained by using ionosphere. Industrial, science and medical bands can also be used for transmission of electrical power in form of microwaves from one distant place of earth to another. Frequency of 10 MHz to 30 MHz can be used which will be reflected back by ionosphere and can be received using receiving antennas. This received energy can again be converted into electrical power by using rectenna consisting IMPATT diodes. Further on the output of IMPATT diodes band pass or band stop filters can be used to get desired frequency and quality.

II. Super Smart Grid System

The concept of a wide area "Super grid" with centralized control and the concept of small-scale, local and decentralized Smart Grid are two approaches that are often perceived as being mutually exclusive alternatives. The SSG aims at reconciling the two approaches and considers them complementary and necessary to realize a transition towards a fully decarbonized electricity system. The super grid features would deliver inexpensive, high capacity, low loss transmission, interconnecting producers and consumers of electricity across vast distances. Smart grid capabilities use the local grid's transmission and distribution network to coordinate distributed generation, grid storage and consumption into a cluster that appears to the super grid as a virtual power plant .

III. Ionosphere and Skywave Propagation

Ionosphere is a layer of ionized gases which starts from 90 kms above earth's surface and persist till 350 kms above the surface of earth. It contains positive as well as negative ions of various gases as hydrogen, nitrogen and oxygen etc. It owes its existence primarily to ultraviolet radiation from the Sun. This whole layer is divided into various layers namely F layer, E layer and D layer (Fig. 1).

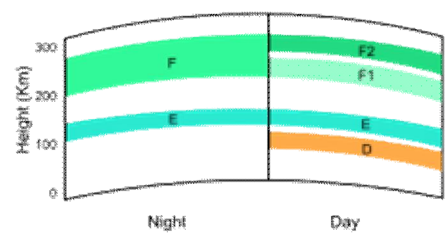


Fig. 1 different layers of ionosphere

F layer which is also known as Appleton layer extends from about 200 km to more than 500 km above the surface of Earth. It is the densest point of the ionosphere, which implies signals penetrating this layer will escape into space. Here extreme ultraviolet (UV, 10–100 nm) solar radiation ionizes atomic oxygen. The F layer consists of one layer at night, but during the day, a deformation often forms in the profile that is labeled F_1. The F_2 layer remains by day and night responsible for most skywave propagation of radio waves, facilitating high frequency (HF, or shortwave) radio communications over long distances.

D layer is the innermost layer, 60 km to 90 km above the surface of the Earth. Ionization here is due to Lyman series-alpha hydrogen radiation at a wavelength of 121.5 nm ionizing nitric oxide (NO). Recombination is high in the D layer, the net ionization effect is low, but loss of wave energy is great due to frequent collisions of the electrons (about ten collisions every msec).

The E layer is the middle layer, 90 km to 120 km above the surface of the Earth. Ionization is due to soft X-ray (1-10 nm) and far ultraviolet (UV) solar radiation ionization of molecular oxygen (O_2). Normally, at oblique incidence, this layer can only reflect radio waves having frequencies lower than about 10 MHz and may contribute a bit to absorption on frequencies above. However, during intense Sporadic E events, the E_s layer can reflect frequencies up to 50 MHz and higher. The vertical structure of the E layer is primarily determined by the competing effects of ionization and recombination. At night the E layer rapidly disappears because the primary source of ionization is no longer present. After sunset an increase in the height of the E layer maximum increases the range to which radio waves can travel by reflection from the layer.

IV. RECTENNA USING IMPATT DIODES

IMPATT stands for Impact Avalanche and Transit Time diodes. These are high power diode used in high-frequency electronics and microwave devices. If a free electron with sufficient energy strikes a silicon atom, it can break the covalent bond of silicon and liberate an electron from the covalent bond. If the electron liberated gains energy by being in an electric field and liberates other electrons from other covalent bonds then this process can cascade very quickly into a chain reaction producing a large number of electrons and a large current flow. This phenomenon is called impact avalanche and is the basic principle of working of an IMPATT diode.[1][5]

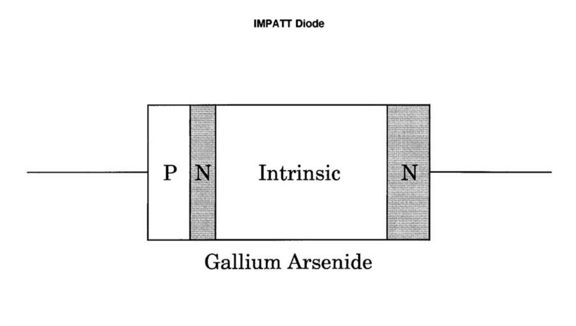


Fig. 2. IMPATT Diode

Two aluminum strips form the dipole and balanced transmission line. An added aluminum piece, located at the diode, is shown for heat sinking and securing the element to a support that suspends the element above a reflecting plane. The rectenna consists of a half wave dipole antenna, a two section input low pass filter, a GaAs IMPATT diode, and an output 30 pF capacitor for shorting RF power and tuning the diode.

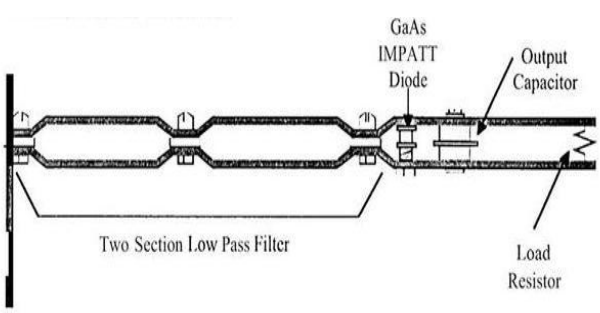


Fig. 3 Design of Rectenna Using IMPATT Diaode

A 165Ohm dc resistive load is also connected in parallel at the output to complete the dc circuit and achieve a conversion efficiency of greater than 85°/0. As a rectifier, the received microwave power is converted into dc power and measured across the 165 Ohm load. As an oscillator, bias is applied with reversed polarity and the RF power radiates from the dipole [3] [4] [7] .The nominal characteristic impedance of the low pass filter is 120 Ohm, and the cutoff frequency is 3.7 GHz for attenuating harmonic signals generated by the diode.

The oscillation frequency is dependent on the diode structure and circuit impedance. Also, the IMPATT doping concentrations and layer thicknesses influence the dc to RF efficiency. IMPATT microwave oscillators are low impedance devices with typical diode impedances of- 1 Ohm to -10 Ohm. The impedance presented to the diode in this circuit is approximately 120 Ohm which explains the low measured efficiency. In fig. 4 the measured spectrum where the oscillation frequency occurs at 3.305 GHz is shown. Sideband oscillations occurring at 560 M1lz off the carrier frequency also occurred but were eliminated by slightly moving the output capacitor away from the IMPATT diode.[2] [6]

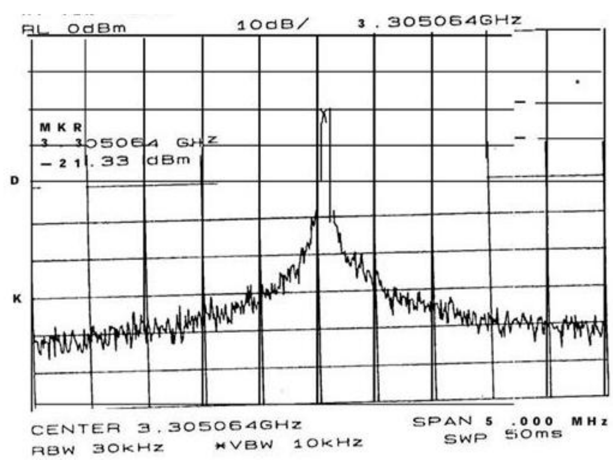


Fig. 4 The Measured Spectrum Of IMPATT Diode

V. IMPLIMENTATION OF WPT OVER SUPER SMART GRID

Now implementation of wire less power transmission technique over super smart grid phenomenon can make this concept practically possible, for which ionosphere can be used. By collecting electrical power generated by different renewable power plants at a centralized station and then convert that electrical power into microwaves of frequency ranging from 10 MHz to 30 MHz and will filter that microwave using band pass filters to get desired frequency output with desired quality.

After which we will further transmit that microwave using transmitting antennas at an angle of 45 degree (max angle at which wave will travel max. Distance). At this angle we can transmit wave from 750 kms to 3500 kms depending upon the extent of penetration (which is a function of frequency), the angle of incidence, polarization of the wave, and ionospheric conditions, such as the ionization density. The band, frequency and wavelength for used frequency is as per Table 1.

Table 1.

BAND	FREQUENCY	WAVELENGTH
HIGH FREQUENCY	3–30 MHZ	100–10 M
VERY HIGH FREQUENCY	30–300MHZ	10–1 M

The band shown in Table 1. Can be used for transmission of electrical power in form of microwaves and can be reflected back from ionosphere. However the frequencies of the wave will vary during night and day time as layers will vary during these timings. As we have mentioned that the distance will also vary from 750 kms to 3500 kms. Electrical power can be transmitted from one distant place to another using single hop or by using multiple hopping technique for long distances (Fig 5.).

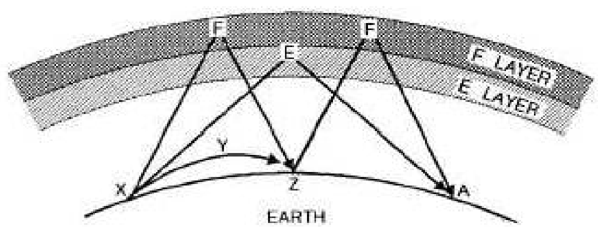


Fig. 5 Multiple hopping technique

Also the power density is inversely proportional to the square of the distance thus the distance will be decreased as the power density of wave will increase thus for high power transmission we will have to use repeaters after a certain distance. The other method is to modulate the high frequency waves of frequency ranging 6 GHz to 7GHz over wave having frequency 10MHz to 30MHz. By using this method the wave can carry more power as compared to the wave of frequency 10 to 30 MHz and will travel more distance too. This method is however increasing the power carrying capacity and distance of transmission but will increase the cost of the equipment as in this case Doppler effect will also be introduced along with tropospheric ducting. So in order to eliminate the ill effects of such kind of disturbances or losses more equipment will be required which will decrease the net efficiency also. But as amount of transmitting power is very high so this method can also be adopted as in this much high level of power such losses can be neglected or can be sought out using efficient equipments. Now As this system is a combination of super grid and super smart grid thus this system will have features of both the systems as self healing, consumer participation, resist attack, high quality power, accommodation generation option, enable electricity market, optimize assets, load adjustment, demand response support, greater resilience to loading, decentralization of power generation, price signalling to consumer etc. All these features will make the system secure, efficient and very accurate.

VI. WIRELSS POWER TRANSMISSION USING LASER AND GROUND BASED WIRELESS VS INOSPHERIC PROPAGATION

Both microwave and laser beams are attenuated by the Earth's atmosphere and its weather dependent particulate content. Attenuation due to scattering is highest when the wavelength is comparable to the size of the particles in the atmosphere. Because of the much shorter wavelength of laser beams, they are much more severely attenuated than microwave beams, to the extent that power beam interruptions to the terrestrial utility station will occur. But these effects on microwave-based system would never be as severe as with laser transmission.[8]**As** a result of their greater sensitivity to weather-induced attenuation, base load electric utility systems employing laser power transmission will require many spatially diverse receiving sites to deal with weather outages.
Whereas for ground base wireless transmission the calculation performed for Fordyce determined that it would require a transmitting may and a rectenna array, both larger in diameter than the height of an 18 story building, in order to achieve over 90% beam coupling efficiency at ***2.35*** GHz over the distance required. At 2.45 GHz, the S-Band Industrial, Scientific and Medical Band (ISM), the wavelength is -12 cm and the equal diameter apertures at transmitter and rectenna for 90 % beam coupling efficiency are 48.6 m, nearly 160 ft, about 14.5 stories based on 11 ft per story. Regarding the cost the very short range (1- lorn), preliminary demonstrations of W at low power levels (< **kW)** were in general quite costly, however, the cost estimates are coming down for larger power levels and longer ranges [9]. Tens of km WPT systems are in the range of several *$M/* MW-km, whereas similar range wired systems are of order $lO,OOO/WMW-km, at least two orders of magnitude less. Given the economic disparity and beam safety concerns, it is doubtful that short range WPT applications involving beams that are near tangent to the Earth's surface will ever be useful for electric power transmission as compared to wired power delivery. However, this does not preclude research and development tests and demonstrations of WPT from point to point on the Earth's surface, where economic competition is not the prime consideration. Thus it can be concluded that wireless power transmission using ionosphere as media for long distance power transmission is the best option.

VII. CONCLUSION

Now the implementation of wireless power transmission is possible by converting electrical power into microwaves using IMPATT diodes and can be transmitted at a maximum angle of 45 degrees. After receiving this wave the reverse process of reconverting it into electrical power can also be done by using same device i.e. IMPATT diodes and thus electrical power can be further transmitted to load centres using super grid concepts i.e. using super conductors or using normal ACSR conductors. In order to transmit high power from one place to another high frequency band can be used. High frequency wave can be further modulated with wave of frequency ranging in between 10 MHz to 30 MHz so that it can be reflected back from ionosphere. However instead of IMPATT diode magnetron can also be used but that would consume more power, will give less efficiency and will not be economical too. So this wireless technology as proved here is best suited for transmitting high power from one distant place of earth to another with all the features of smart grid and super grid as self healing, consumer participation, resist attack, high quality power, accommodation generation option, enable electricity market, optimize assets, load adjustment, demand response support, greater resilience to loading, decentralization of power generation, price signalling to consumer etc.

REFERENCES

[1] *B.* B. van lpercn and Ija.sscns, "An accurate bridge method for impedance measurements of IMPATT diodes," *A4i(ro}\wc Jourwl, vol. i* 5, no. 1 I, pp. 29- 33, NOV. 1972.

[2] Cannon, B. L., J. F. Hoburg, D. D. Stancil, and S. C. Goldstein, \Magnetic resonant coupling as a potential means for wireless power transfer to multiple small receivers," *IEEE Transactions on Power Electronics*, Vol. 24, No. 7, 1819{1825, 2009.

[3] E. C. Okress, cd., *Rectification,* in *Microwave Power Engineering, vol.* 1. New York: Academic Press, 1968, pp. 273-340.

[4] *G. Czisch, "Low cost but totally renewable electricity supply for a huge supply area" in 2004.* IEE-RE, Universidad Kassel, Germany.

[5] G. Salmer, J. Pribetich, A. Farrayre, and B. Kramer, "Theoretical and experimental study of GaAs IMPATT oscillator efficiency," *Journal of Applied Physics, vol. 44,* no. 1, pp. 314-324, Jan. 1973.

[6] Hashimoto, K. and H. Matsumoto, "Microwave Beam Control for Space Solar Power Satellite (in Japanese)", Proc. of the Institute of Electronics, Information and Communication Engineers, SBC-1-12, 2004.

[7] Iero, D., T. Isernia, and A. F. Morabito, \Optimal constrained field focusing for hyperthermia cancer therapy: A feasibility assessment on realistic phantoms," *Progress In Electromagnetic Research*, Vol. 102, 125{141, 2010.

[8] Karalis, A., J. D. Joannopoulos, and M. Soljacic, \Efficient wireless non-radiative mid-range energy transfer," *Annals of Physics*, Vol. 323, No. 1, 34{48, 2008.

[9] Kurs, A., A. Karalis, R. Moatt, J. D. Joannopoulos, P. Fisher, and M. Soljacic, \Wireless power transfer via strongly coupled magnetic resonances," *Science*, Vol. 317, 83{86, 2007.

[10] Marques, R., F. Martin, and M. Sorolla, *Metamaterial with Negative Parameters: Theory and Microwave Applications*, Wiley-Interscience, New York, 2008.

[11] N. G. Hingorani, " High-voltage dc transmission: a power electronics workhorse," *IEEE Spectrum,* vol. *34,* no. *4,* pp. 63 72, April 1996.

[12] "Reception-conversion subsystem (RXCV) for microwave power transmission system, final report," Raytheon Company, Sudbury, MA, Tech. Report No. ER75- 4386, JPL Contract No. 953968, NASA Contract No. NAS 7 100, Sept. 1975.

[13] R. M. Dickinson,, " Ground based wireless and wired power transmission cost comparison" in 1999. *The Jet Propulsion Laborator*y, *California Institute of Technology.*

[14] R. M. Dickinson, "Performance of a high-power, 2.388 GHz receiving array in wireless power transmission over 1.54 km," in 1976 *IEEE A4TT-S International Microwave Symposium, 1976,* pp. 139-141.

[15] Shinohara, N., T. Mitani, and H. Matsumoto, "Development of Phase and Amplitude Controlled Magnetron", Proc. of Sixth International Vacuum Electronics Conference IVEC2005, 2005,

[16] Takano, T., A. Sugawara, and N. Kamo, "Simplification Techniques of the Constitution of Microwave Transmission Antennas of SPS (in Japanese)", Tech. Rep. of IEICE, SPS2003-09(SPS2004-02), 2004.

[17] W. C. Brown, "Design definition of a microwave power reception and conversion system for use on a high altitude powered platform," Wallops Flight Facility, VA, NASA Report No. CR-1 56866, NASA Contract No. NAS 6-3006, July 1980.

[18] Xu, H., C. Sanabria, A. Chini, S. Keller, U. K. Mishra, and R. A. York, "A C-Band High-Dynamic Range GaN HEMT Low-Noise Amplifier", IEEE Microwave and Wireless Componets Lett., Vol.14, No.6, 2004.

Effects of Unilateral Robotic Limb Loading with Robot Types

Er. Damanbir Singh Virk, Dr V.K Banga, Er Gurjit Singh
Amritsar College of Engineering and Technology, Amritsar, INDIA.

Abstract :This review paper presents the different types of robots used in industries and also the mechanical hazards associated with robot manuplater with special emphasize on overloading. the key parameter in providing the motion to a robot i.e DOF has also been discussed. This research paper has reviewed the hazards corelation with overloading. different categories of loads have also been listed to give a brief preface about overloading.

Keywords: robots, DOF(degree of freedom), manuplater, overloading.

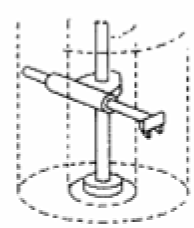
1. Introduction

The industrial scenerio is changing all over the world. Nowdays, in many parts of the world, human workforce has been replaced by robots in developed countries and same change is being seen in developing countries. Robots are identifable as a unique device with computer aided design (CAD) systems and computer aided manufacturing (CAM) systems characterizes the latest trends in the automation of the manufacturing process [1]. Present use of industrial robots in concentrated in rather simple, repetitive tasks which tend not to require high precision. In 1980s relative simple tasks like machine tending, material transfer, painting and welding were economically viable for robotization. But nowadays , industrial robots are becoming increasingly viable in applications which require more precission and sensory sophistication such as assembly task, (etc) [2]. Industrial robots can be said as programmable multifunctional mechanical devices designed to move material, parts tool or specialized devices through variable programmed motions to perform a variety of tasks. They are multifunctional manipulators having a number of degrees of freedom in three-dimentional space which helps them to perfoem different tasks with ease. Although the appearance and capabilities of robots vary vastly, all robots share the features of a mechanical, movable structure under some form of autonomous control. The structure of a robot is usually mostly mechanical and can be called a kinematic chain (its functionality being similar to the skeleton of the human body). The chain is formed of links (its bones), actuators (its muscles) and joints which can allow one or more degrees of freedom. Most contemporary robots use open serial chains in which each link connects the one before to the one after it. These robots are called serial robots and often resemble the human arm. Some robots, such as the Stewart platform, use closed parallel kinematic chains. Robots used as manipulators have an end effector mounted on the last link. This end effector can be anything from a welding device to a mechanical hand used to manipulate the environment. The mechanical structure of a robot must be controlled to perform tasks. The control of a robot involves three distinct phases - perception, processing and action (robotic paradigms). Sensors give information about the environment or the robot itself (e.g. the position of its joints or its end effector). Using strategies from the field of control theory, this information is processed to calculate the appropriate signals to the actuators (motors) which move the mechanical structure[3].

2. TYPES OF ROBOTS:

1 Cylindrical coordinate robot:

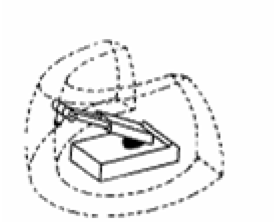
It is used for assembly operations, handling at machine tools, spot welding, and handling at die-casting machines. It's a robot whose axes form a cylindrical coordinate system.



Cylindrical Coordinate Robot

2.2 Spherical coordinate robot:

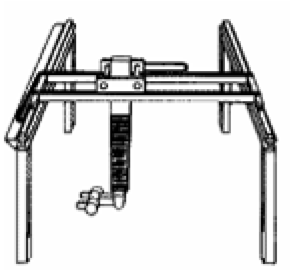
It is used for handling at machine tools, spot welding, die-casting, fettling machines, gas welding and arc welding. It's a robot whose axes form a polar coordinate system.



Sperical Coordinate Robot

2.2.1 Gantry robot:

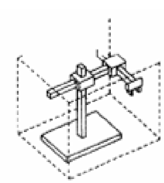
It is used for pick and place work, application of sealant, assembly operations, handling machine tools and arc welding. It's a robot whose arm has three prismatic joints, whose axes are coincident with a Cartesian coordinator.



Gantry Robot

2.3 Rectangular coordinate Robot:

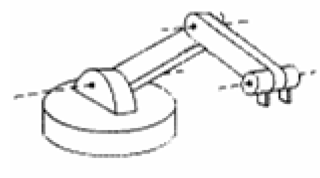
A robot having orthogonal, sliding joints and supported by a nonrotary base as the axis.



Rectangular Coordinate Robot

2.5 Articulated arm robot:

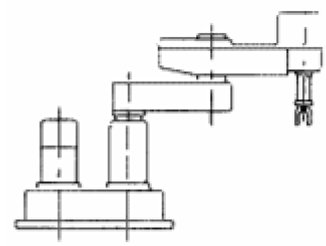
It is used for assembly operations, die-casting, fettling machines, gas welding, arc welding and spray painting. It's a robot whose arm has at least three rotary joints. It is most widely used in the industry.



Articulated Arm Robot

2.6 SCARA robot:

It is used for pick and place work, application of sealant, assembly operations and handling machine tools. It's a robot which has two parallel rotary joints to provide compliance in a plane.



SCARA Robot

3. DEGREES OF FREEDOM

Regardless of the configuration of a robot, movement along each axis will result in either a rotational or a translational movement. The number of axes of movement (degrees of freedom) and their arrangement, along with their sequence of operation and structure, will permit movement of the robot to any point within its envelope. This number typically refers to the number of single-axis rotational joints in the arm, where a higher number indicates an increased flexibility in positioning a tool Robots have three arm movements (up-down, in-out, side-to-side). In addition, they can have as many as three additional wrist movements on the end of the robot's arm: yaw (side to side), pitch (up and down), and rotational (clockwise and counterclockwise).

Robots are generally used to perform unsafe, hazardous, highly repetitive, and unpleasant tasks. They have many different functions such as material handling, assembly, arc welding, resistance welding, machine tool load and unload functions, painting, spraying, etc. Most robots are set up for an operation by the teach-and-repeat technique. In this mode, a trained operator (programmer) typically uses a portable control device (a teach pendant) to teach a robot its task manually. Robot speeds during these programming sessions are slow.

4. HAZARDS ASSOCIATED WITH ROBOTS

The main hazard associated with the application of industrial robot is the working envelope of the robot. The ability of the robot to move in free space which cover a wide area, change configuration and produce unexpected motion immediately can cause hazards to persons operating or standing in the vicinity of the robot. Therefore in any robot installation, hazard analysis should be carried out to identify hazards so that safeguards can be implemented to prevent the occurrence of accidents.

Malfunction and human error can lead to the unexpected movement of the industrial robot which include:

a) Aberrant behaviour of robots caused by control system faults.
b) Jamming of servo-valves.
c) Robot movement cutting its umbilical cord.
d) Splitting of unions on exposed hydraulic hoses.
e) Fault in data transmission causing a larger than anticipated movement of the robot arm.
f) Faults of welding gun and tooling parts.
g) Programming and other operational errors.
h) Precision deficiency, deterioration.
i) Incompatibility of jigs and other tools.

5. BASIC THREE POTENTIAL HAZARDS ASSOCIATED WITH ROBOTIC SYSTEMS:

There are three basic potential hazards associated robotic systems are as follows:

5.1 *Impact*:

This involves such things as being struck by a moving part of the robot, or by parts or tool carried or manipulated by the robot. It can be caused by the unexpected movement of the robot or by the robot ejecting or dropping workpieces or molten metal.

5.2 Trapping :

This can be caused by the movement of the robot in close proximity to fixed objects like machines, equipment, fences, etc. Trapping points can also be caused by the movement of the work carriages, pallets, shuttles or other transfer mechanisms. They can also be presented on the robot itself on the arm or mechanism of the robot.

5.3 Other:

This would include hazards inherent to the application itself like electric shock, arc flash, burns, fume, radiation, toxic substances, noise, etc.

6. ISSUES RESPONSIBLE FOR THE HAZARDS:

These hazards can arise from several sources and should be considered in a typical robot installation which include:

a) *Control Errors:*

These are faults within the control system of the robot like software errors, electrical interference, or faults in the hydraulic, pneumatic, or electrical sub-controls associated with the robot. Electrical interference can come from two sources — line noise and radiated frequency interference. Both types are hazards because they can cause erratic operation of microprocessor controlled robots.

b) *Mechanical Hazards:*

These can be caused by parts or tools carried or manipulated by the robot like sharp edges, heavy weights and exposed electrodes. Mechanical failure may lead to the ejection of workpieces by the robot gripper. Causes of this hazard can be due to overloading, corrosion, fatigue and lack of maintenance.

c) Environmental Hazards:

Application of robots can also cause an environmental hazard in some cases. Examples of this are welding robots which usually produce large amounts of fumes, arc flash and flying particles. Other environmental hazards may include dust, vapour, x-ray, laser, ultraviolet, ionising and non-ionising radiation, flammable and explosive atmospheres.

d) Human Errors:

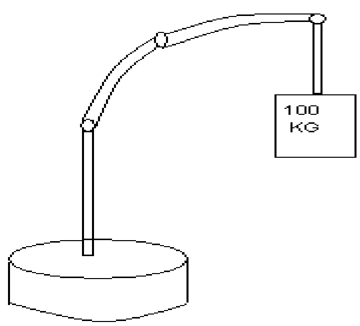
In most robot installations, persons may have to work close to a robot or enter the guarded area of a robot and hence exposed to trapping points and impact. This occurs during programming, teaching, maintenance, or in work handling close to robot or at the loading/unloading station. Lack of familiarity with the equipment is a major cause of human error which can be hazardous.

Ancillary Equipment:

In most robot installations, the robot usually works in conjunction with other equipment like conveyor, machine tool, press, shear, etc. This equipment can also create a hazard if the dangerous parts are within reach of a person and not enclosed by guards.

7. OVERLOADING:

Loads cause stresses, deformations and displacements in structures. Assessment of their effects is carried out by the methods of structural analysis. Excess load or overloading may cause structural failure, and hence such possibility should be either considered in the design or strictly controlled.

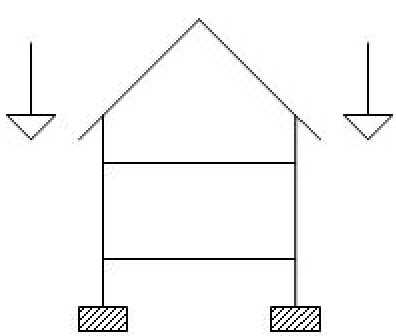


Types of loads

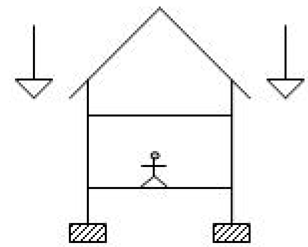
Dead loads are static forces that are relatively constant for an extended time. They can be in tension or compression. The term can refer to a laboratory test method or to the normal usage of a material or structure.

Live loads are usually unstable or moving loads. These dynamic loads may involve considerations such as impact, momentum, vibration, slosh dynamics of fluids, etc.

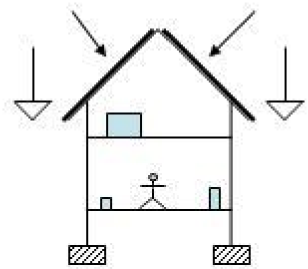
Cyclic loads on a structure can lead to fatigue damage, cumulative damage, or failure. These loads can be repeated loadings on a structure or can be due to vibration.



Dead load



Imposed load



Live snow load

7.1 Dead loads:

The dead load includes loads that are relatively constant over time, including the weight of the structure itself, and immovable fixtures such as walls, plasterboard or carpet. Dead loads are also known as Permanent loads.

The designer can also be relatively sure of the magnitude of dead loads as they are closely linked to density and quantity of the construction materials. These have a low variance, and the designer himself is normally responsible for the specifications of these components.

7.2 Live loads:Live loads, or imposed loads, are temporary, of short duration, or moving. These dynamic loads may involve considerations such as impact, momentum, vibration, slosh dynamics of fluids, fatigue, etc.

Live loads, sometimes also referred to as probabilistic loads include all the forces that are variable within the object's normal operation cycle not including construction or environmental loads.

Roof live loads are produced:

a) During maintenance by workers, equipment and materials, and during the life of the structure by movable objects such as planters and by people.
b) Bridge live loads are produced by vehicles traveling over the deck of the bridge.

8 CONCLUSION:

This paper has given a brief introduction about various types of robots and the hazards associated with them due to overloading. Also different types of overloading in robot manuplaters are explained as with passage of time the robotic arm might develop number of issues related to degree of arm, mechanical britalness and fatigueness other than its robotic arm control is also prone to human errors, who might unmind fully abuse the robotic arm in terms of overloading and handling. Therefore there is need to develop an automated system which alerts the controller and automatically avoid further damage.

REFERENCES

[1] Craig J.J., *Introduction to Robotics: Mechanics and control*, Pearson Education Asia, 2004.
[2] Deb S.R., *Robotics Technology and Flexible Automation*. Tata Mc-Graw Hill, 2002.
[3] Fu K.S., Gonzalez R.C., Lee C.S.G., *Robotics: Control, Sensing, Vision and Intelligence*. Mc-Graw Hill International Editions, 1987.
[4] Galicki M., 2000. Time-optimal controls of kinematically redundant manipulators with geometric constraints. IEEE Transactions on Robotics and Automation 16(1), 89-93.
[5] Gemeinder M. and Gerke M., 2003. GA-based path planning for mobile robot systems employing an active search algorithm. Elsevier Applied Soft Computing 3(2), 149-158.

Analysis and Existing Trends in WiMAX Technology

Er. Gurpal Singh Sidhu, Er Damanbir Singh Virk
Amritsar College of Engineering and Technology, Amritsar, INDIA.

Abstract **: This review paper presents types of Wimax used in industries. Business is increasingly becoming a mobile activity, and as a result the wireless networks and services used to support that development are growing in importance. In both the business-to-business and business-to-consumer (B2B and B2C) environments, the availability of more reliable, higher-capacity wireless data networks is one of the keys to expanding the reach of business into the mobile environment. This transformation is occurring in the context of an overall enterprise shift toward all-IP communications. "IP," or Internet Protocol, describes both the format and the switching technology that drives the core of the Internet. Originally envisioned as a general-purpose data transport, IP has now expanded to support voice and video communications over an integrated IP backbone.**

Keywords: Wimax, Internet protocol, LTE (Long term Evolution, MSC (Mobile Switching Centre).

1. INTRODUCTION

Recognizing these developments, the wireless industry is now aligning itself to take advantage of these trends. The moniker that is used to identify that transition is 4th Generation or 4G, and it is used to describe two primary technologies, WiMAX and Long Term Evolution (LTE)[1]. While there are some differences in the implementations, they share a number of key characteristics:

• Higher Capacity: WiMAX and LTE will use very similar state of- the-art radio technologies to deliver several times the transmission capacity of existing 3G wireless services.

• Reliability: Using advanced signal encoding and smart antenna technologies, WiMAX and LTE will also deliver a major advance in network reliability, even in the most challenging radio environments, such as densely populated urban environments.

All-IP Communications: Whereas in the past, wireless services were divided along voice and data lines, 4G technologies such as WiMAX and LTE are based on the concept of an all-IP network[2]. This means there exists a single IP pipe that is capable of supporting voice, data and video communications. For the end user, this transition will mean a new wireless platform that delivers a far wider range of applications with performance and reliability that mimic the desktop experience. Now, rather than settling for a wireless experience with performance that limits the range and utility of applications, wideband voice, high-quality video and lightning-fast downloads will be available to users regardless of whether they are in the office or on the go. While both WiMAX and LTE promise these capabilities, only WiMAX is available today. With a comprehensive migration plan to evolve from our current 3G infrastructure, Sprint 4G powered by WiMAX is delivering the benefits of 4G to customers in selected markets years ahead of the first LTE rollouts. Because WiMAX networks are already being deployed, users can begin taking advantage of 4G today[3].

BASIC TECHNOLOGY OF WIMAX:

As shown in Figure 1 below, the overall structure of a mobile network involves a number of elements, all of which contribute to the service that is provided.

Mobile Switching Center (MSC):

The MSC is the mobile central office; mobile operators typically maintain one or more per city. With Sprint 4G, that facility is called a WiMAX Service Center or WSC. The MSC complex includes other elements, including Authentication, Authorization and Accounting (AAA) servers; Access Service Node Gateways (ASN-GWs), Media Gateways and Media Gateway Controllers (MGWs/MGCs); Home Subscriber Servers (HSSs); and interfaces to Operations, Administration, Maintenance and Provisioning (OAM&P) systems. The MSCs or WSCs in turn are interconnected with a signaling network to support users who are roaming in other cities.

Cell Towers/Base Stations (BSs):

These are the visible towers that support the radio connection between the mobile network and the users. The BS is comprised of antennas and radio transmission equipment. Different carrier frequencies are typically assigned to each BS, and in busy cells, the coverage area can be divided into three 120° sectors, allowing three different carrier channels to be used, effectively tripling the cell's capacity.

Backhaul Network:

The invisible part of the network is the web of wired and wireless communication links that connect each of the BSs to the MSC or WSC. A single 4G cell site will require tens of megabits of capacity[4]. Providing higher capacity 4G services has required mobile operators to invest heavily in the upgrading of their backhaul facilities.

Mobile Devices:

The last part of the picture is the user's mobile device. Initially that was a simple voice device, but users are increasingly opting for smartphones. The move to 4G will allow a wider range of netbooks, laptops, personal digital assistants (PDAs) and even consumer electronics devices to connect over the network.

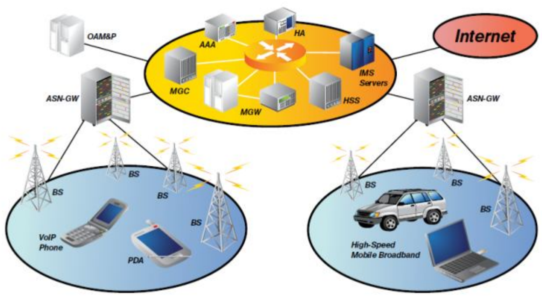


Fig: 1

WiFi AND WiMAX :

While describing the evolution of wireless networks, it is also important to consider another widely successful wireless technology, wireless LAN, or Wi-Fi. Wi-Fi and WiMAX share a common ancestry in that both are products of the Institute of Electrical and Electronics Engineers (IEEE) 802 committees. Originally focused on wired networks like Ethernet (i.e., IEEE 802.3), in the mid-1990s the IEEE 802 committees began expanding their scope to address wireless networks as well. The 802.11 committee was tasked with developing the standards for wireless LANs (WLANs), and the 802.16 committee set to work on standards for wireless metropolitan area networks (MANs) that led to the development of WiMAX.

The IEEE is a worldwide standards body that develops standards using a process governed by a rigorous set of rules to ensure due process, openness, consensus and balance. The result is an open standard based on industry-wide consensus that leads to a larger ecosystem of suppliers, a wider variety of products and, ultimately, to lower costs. So, while Wi-Fi and WiMAX address different areas of the wireless market, they share the same foundation in a set of worldwide standards. Each of the key 802-series standards is supported by a vendor consortium that ensures multivendor interoperability and longterm technological stability. The wireless LAN manufacturers have banded together in the Wi-Fi Alliance, while the wireless MAN manufacturers formed the WiMAX Forum. These vendor organizations play a critical role in the adoption of a new technology. For each important development in the standards, the vendor consortium develops a series of tests to ensure that the capability is implemented consistently in each vendor's product. For WiMAX, those devices are tested and certified in independent WiMAX Forum Designated Certification Laboratories. That certification is essentially an industry guarantee that the device will interoperate with any other certified device[5]. The Wi-Fi Alliance has had tremendous success with their certification program, and the WiMAX Forum is now following in that same track. So while Wi-Fi and WiMAX are often compared, they really address two different sets of requirements. Wi-Fi is usedprimarily in private local networks operating on readily available unlicensed frequency bands; Wi-Fi transmissions have a maximum transmission range of 100m. The use of unlicensed frequency bands means that Wi-Fi transmissions are subject to interference from other users. WiMAX is typically provided by mobile operators over licensed frequency bands, and transmissions have a range of several miles. Given their different design parameters, WiMAX will not replace Wi-Fi, but rather the two will work side-by-side, each targeting the applications for which it was developed. As they are both grounded in the same framework of open international standards, we can anticipate the same type of industry-wide acceptance for WiMAX that we have already seen for Wi-Fi[6].

3. WiMAX, AS 4G TECHNOLOGY

WiMAX is short for Worldwide Interoperability for Microwave Access. It describes a 4G metro-area wireless technology defined in the IEEE 802.16 standards and promoted by the WiMAX Forum. Using our extensive holdings of 2.5 GHz BRS spectrum, Sprint, along with our partner Clearwire Communications, is currently deploying the first nationwide 4G WiMAX network in the U.S. There are a number of factors that argue for WiMAX as the preferred 4G environment:

Timeframe:

While LTE vendors are hoping to initiate trials in the 2010/2011 timeframe, 4G is already available in a number of cities across the country, with plans to cover 120 million people by year-end 2010.

Real-World Experience :

A few small-scale LTE trials have been conducted primarily in the U.S., but there are already 502 WiMAX networks operating in 145 countries around the world.

Open Standards:

WiMAX is based on IEEE developed international standards. Just as we have seen with Wi-Fi, open standards can lead to a larger, more diverse ecosystem, and lower prices to consumers. Where LTE implementations may still incorporate variations based on the mobile operator's business objectives, the WiMAX community is committed to the concept of open standards and multivendor interoperabilitythroughout. Royalty-Free Technology To further ensure lower prices to consumers, the WiMAX Forum supports the Open Patent Alliance (OPA) model offering open, transparent, predictable and nondiscriminatory access to core technologies with the objective of delivering a fair royalty rate to all. WiMAX Forum: The development of those WiMAX devices will be supported by the WiMAX Forum, whose stringent certification process will ensure support for manufacturers and full interoperability for users. Frequency Bands: Sprint 4G is deploying WiMAX in 2.5 GHz bands that we have been using since the late 1990s, while LTE will be deployed in the newly released 700 MHz bands that have yet to be tested for data transport. Global Reach: Mobile operator plans for deploying LTE vary widely, but WiMAX networks are now in commercial use around the world and the WiMAX Forum has already begun work on inter-carrier roaming plans.

SECURITY FEATURES

WiMAX 4G is designed from the outset with state-of-the-art security capabilities consistent with the requirements of the Personal Card Industry (PCI), Health Insurance Portability and Accountability Act (HIPAA) and other industry security mandates. All WiMAX devices and base stations will have X.509 certificates allowing for mutual authentication based on IEEE 802.1x Extensible Authentication Protocol (EAP)-Transport Layer Security (TLS). Once authenticated, over-the-air transmission will utilize encryption based on the Advanced Encryption Standard (AES) using cipher block chaining (CBC) mode and a 128-bit key. While EAP-TLS and AES encryption will protect the over-theair transmission, many customer applications call for end-to-end security. There are several options for how that secure connection could be extended to the customer's data center[7]. First, a customer can implement secure tunnel communications from their firewall, through the Internet and over the Sprint 4G connection using SSL, IPsec, PPTP or L2TP. The Sprint 4G service can also interconnect with Sprint Data LinkSM service that extends a secure VPN tunnel connection from the WiMAX Service Center over an MPLS VPN, IP VPN (IPsec), Frame Relay or traditional dedicated line connection all the way to the customer's data center. Further, the service also supports static IP addressing for firewall traversal and other specialized requirements.

CONCLUSION

This paper has given a brief introduction about the move to WiMAX 4G, which represents a major step forward in wireless communications. With state-of-the art radio technology, Sprint 4G moves wireless performance to a new level while offering speed, reliability and security on par with wired network connections. Extending this type of network service to mobile users is equivalent to moving from dial-up to broadband Internet access. As Sprint 4G is built on an all-IP core network, it brings the mobile network into step with the overall directions in enterprise networking. Most importantly, WiMAX uses IEEE-developed open standards, allowing the entire industry to design, build and integrate WiMAX-compatible products and services. For enterprise users, 4G powered by WiMAX means that you don't have to wait to take advantage of outdated technologies. The revolution has begun.

REFERENCES

[1] S. M. Alamouti, "*A Simple Transmit Diversity Technique for Wireless Communications*," IEEE Journal on Selected Areas in Communications, vol. 16, no. 8, pp. 1451 – 1458, October 1998.

[2] A. Goldsmith, "*Wireless Communications",* Cambridge University Press, 2005.

[3] J. Romero and A. Goldsmith, "*Optimizing Antenna Array Processing in CCI Channels: Is It Better to Cancel or Combine*," Proc. 2006 IEEE Intl. Conf. Commun., June 2006.

[4] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space—Time Wireless Communications,* University Press, 2006

[5] Gemeinder M. and Gerke M., 2003. "*GA-based path planning for mobile robot systems employing an active search algorithm"*. Elsevier Applied Soft Computing 3(2), 149-158.

[6] E.Biglieri,R.Calderbank,T.Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, "*MIMO Wireless Communication"s.* Cambridge University Press, 2006

[7] M. O. Damen, A. Chkeif, and J.-C. Belfiore, "*Lattice Codes Decoder for Space-Time Codes*,"Electronics Letters, vol. 4, pp. 161 – 163, May 2000.

Thin Film as a DMS Material

Harpreet kaur[@] , Amit Sarin[@]
@ Deptt. Of Applied Sciences,
Amritsar College of Engineering and Technology,Amritsar,Punjab,India
Email: harpreet18kaur@yahoo.co.in, amit.sarin@yahoo.com

***Abstract* — Thin films are thin material layers ranging from fractions of a nanometre to several micrometres in thickness. Thin-film coatings play a prominent role on the manufacture of many electric devices. Thin films are deposited onto bulk materials to achieve properties unattainable in substrates alone e.g. Cr coatings used on plastic parts in automobiles impart hardness, metallic luster etc.Mn doped ZnO thin films are prepared by using spin coating technique.These films come under the category of Dilute Magnetic Semiconductors(DMS).This paper deals with fabrication of such a DMS material.**

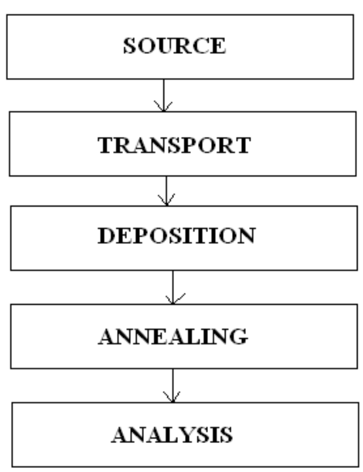
Index Terms – DMS, Thin Film,Spin Coating Technique.

I. Introduction

Thin films are thin material layers ranging from fractions of a nanometre to several micrometres in thickness. Thin-film coatings play a prominent role on the manufacture of many electric devices. Thin films are deposited onto bulk materials to achieve properties unattainable in substrates alone e.g. Cr coatings used on plastic parts in automobiles impart hardness, metallic luster etc.

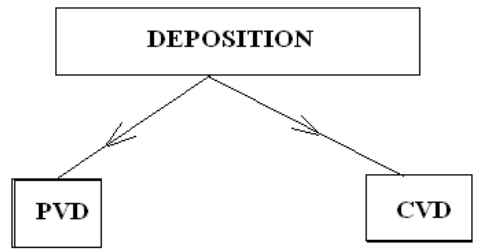
II. PROCESS STEPS

All thin film processes involve four to five sequential steps:



III. DEPOSITION PROCESS

A thin-film deposition process in which a material (metal, alloy, compound, or composite) is either evaporated or sputtered onto a substrate in a vacuum. There are mainly two categories in deposition.



(A) Physical Vapour Deposition:

PVD is a variety of vacuum deposition and is a general term used to describe any of a variety of methods to deposit thin films by the condensation of a vaporized form of the material onto various surfaces (e.g., onto semiconductor wafers). The coating method involves purely physical processes such as high temperature vacuum evaporation or plasma sputter bombardment rather than involving a chemical reaction at the surface to be coated as in chemical vapor deposition. The term physical vapor deposition appears originally in the 1966 book “Vapor deposition” by CF Powell, JH Oxley and JM Blocher Jr[1] but Michael Faraday was using PVD to deposit coatings as far back as 1838.

(B) Chemical Vapour Deposition:

CVD is a chemical process used to produce high-purity, high-performance solid materials. The process is often used in the semiconductor industry to produce thin films. In a typical CVD process, the wafer (substrate) is exposed to one or more volatile precursors, which react and/or decompose on the substrate surface to produce the desired deposit. also. Frequently, volatile by-products are also produced, which are removed by gas flow through the reaction chamber.

IV. Dilute magnetic Semiconductors

Currently, most information is stored in a nonvolatile way in magnetic bits on hard drives that make use of electrons' spin (their orientation up or down), while semiconductor devices like RAM operate by manipulating electron charge. But the ability to manipulate electrons' spin together with controlling their charge flow could create a host of new capabilities for computer technology—like eliminating the need for lengthy boot-up times. This is the topic of an exciting research field known as "spintronics"[2].Manipulating an electron's magnetic state in a device is the key to successful spintronics, and the simplest way to do that is by using a semiconductor material such as zinc oxide(ZnO), which incorporates metal like manganese (Mn). A major obstacle to developing such devices is creating magnetic semiconductor materials that work at room temperature. In their research, their results are crucial for understanding and further development of a new class of semiconductors—dilute magnetic semiconductor.

Indeed, magnetism and semiconducting properties are known to coexist in some ferromagnetic semiconductors materials, such as europium chalcogenides[3,4] and ferrimagnetic and ferromagnetic semiconducting spinel, that have a periodic array of magnetic element (Figure 1A) but these semiconductors are difficult to grow and are incompatible with the semiconductor material such as Si, and GaAs because its crystalline structure is absolutely different and have low curie temperatures (*Tcs*) about 100 K[5,6] . In order to achieve a ferromagnetic semiconductor (Figure 1C), it is necessary to introduce different atoms of impurity, such as Mn, Cr, Co, Ni, Fe and Cu into the structure of non-magnetic semiconductors (Figure 1B).

This category of semiconductor is then called dilute magnetic semiconductor or (DMSs). In Figure 1, the three types of semiconductors are sketched.

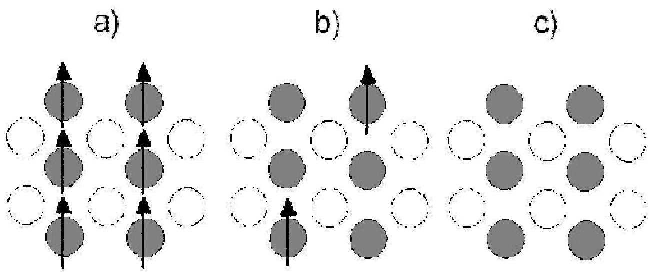


Fig1. Three type of semiconductors: (A) a magnetic semiconductor, in which a periodic array of element is present (B) a dilute magnetic semiconductor and magnetic element. ; (C) a nonmagnetic semiconductor, which no contain magnetic ions

The present paper is focused on the synthesis of transition doped ZnO.Solution processing routes are based on low cost, less toxicity, friendly and easily scalable methods. The results reported thus far, provide a pathway for exploring the effect of different (annealing temperatures) in synthesis preparation for achieving the knowledge about the effect on optical properties of the two i.e. undoped and transition doped ZnO.

ZnO- based DMS:

Zinc oxide (ZnO) is a very promising candidate for future thin-film technology. ZnO materials have received broad attention due to their distinguished performance in electronic, optics and photonics[7,8]. Synthesis of ZnO thin film has been received an active field because of their application as sensor, transducers and catalysts.

Some of the properties of ZnO are as follows:

- It is a direct wide band gap semiconductor (3.37 eV)
- Theoretical studies predicted room temperature ferromagnetism in semiconducting ZnO.
- The thermal equilibrium solubility of magnetic materials in ZnO is larger than 10 mol %.
- Heavy electron doping is readily achieved in ZnO.
- Has potential use in vast applications.
- High Exciton binding energy ~ 60 MeV
- Transparent in visible light spectra.
- It is a good piezoelectric material.
- A good gas sensor.
- It is very resistive to high energy radiation making it suitable for space application.
- It has hexagonal wurtzite structure.
- N-type doping in Fe, Co or Ni-alloyed ZnO was predicted to stabilize high curie temperature ferromagnetism.

ZnO has been extensively researched as a candidate material for developing functional materials for spintronic devices. The research to develop and build devices out of ZnO based DMS has come a fairly long way over the past decade. Many transition element dopants have been used to fabricate DMS out of ZnO, the most popular dopant elements being manganese (Mn) and cobalt (Co).

V. EXPERIMENTAL PLAN

A variety of techniques have been employed to fabricate ZnO thin films such as pulsed laser deposition , RF magnetron sputtering , chemical vapor deposition, spray pyrolysis and sol–gel process [9]. Despite the crystalline quality being inferior to other vacuum deposition techniques, the sol–gel processing still offers the possibility of preparing a small as well as large-area coating of ZnO thin films at low cost for technological applications. In this work, we deal with sol–gel derived Mn doped ZnO thin films. As a consequence, the influence of post-thermal treatments on optical properties will also be investigated.

A Brief outline of the experimental plan is as follows:

- To prepare TM (Mn) doped ZnO films through Sol-Gel process using spin coating technique.
- To prepare thin film samples on different parameters:Doping concentration of Mn ions in ZnO.

(a) Sol Gel Technique

A sol is a dispersion of the solid particles (~ 0.1-1 m) in a liquid where only the Brownian motions suspend the particles. A gel is a state where both liquid and solid are dispersed in each other, which presents a solid network containing liquid components.The sol-gel coating process usually consists of 4 steps:

(1) The desired colloidal particles once dispersed in a liquid to form a sol.

(2) The deposition of sol solution produces the coatings on the substrates by spraying, dipping or spinning.

(3) The particles in sol are polymerized through the removal of the stabilizing components and produce a gel in a state of a continuous network.

(4) The final heat treatments pyrolyze the remaining organic or inorganic components and form an amorphous or crystalline coating.

(b) Spin Coating Technique:

Spin Coating is a procedure used to apply uniform thin films to flat substrates. In short, an excess amount of a solution is placed on the substrate which is then rotated at high speed in order to spread the fluid by centrifugal force. A machine used for spin coating is called 'spin coater, or simply spinner.

Rotation is continued while the fluid spins off the edges of the substrate, until the desired thickness of the film is achieved. The applied solvent is usually volatile, and simultaneously evaporates. So, the higher the angular speed of spinning, the thinner the film. The thickness of the film also depends on the concentration of the solution and the solvent.

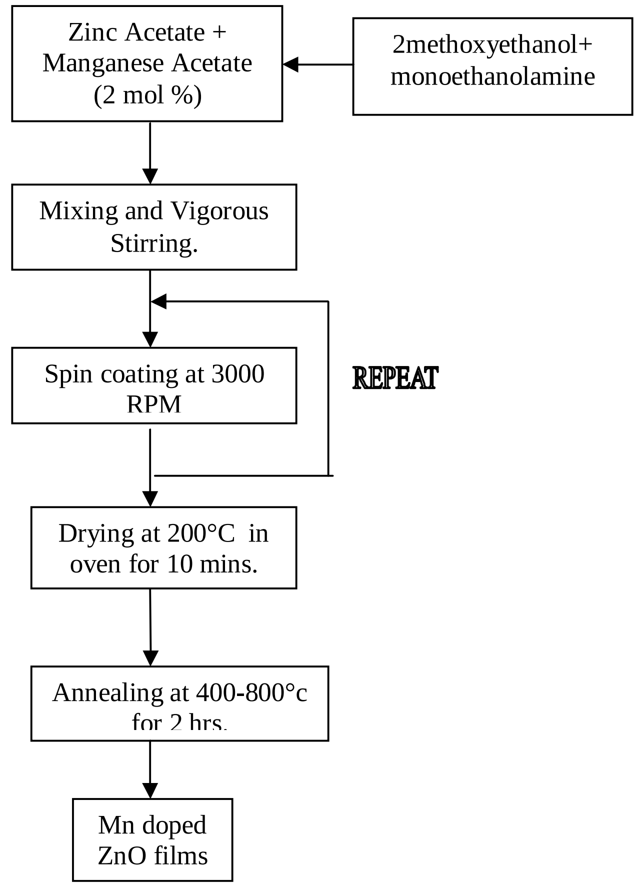
Spin coating is widely used in micro fabrication, where it can be used to create thin films with thickness below 10nm.

VI. SYNTHESIS OF $Zn_{0.96}Mn_{0.04}O$

The Zn1-xMnxO (with x = 0.04) thin films were deposited on quartz substrate. Zinc acetate dihydrate [Zn (CH3COO)2 2H2O] and manganese acetate tetra hydrate [Mn(CH3COO)2 4H2O] were used as precursors; and their powders in the desired molar ratio were dissolved in 20 ml solution of 2-Methoxyethanol (CH3OCH2CH2OH). The molarity of the solution was maintained at 0.2M. The substrate temperature was maintained at the room temperature during the deposition of films. 2-methoxyethanol (CH3OCH2CH2OH) and monoethanolamine are used as the solvent and the stabilizer, respectively. All the mixers were stirred vigorously for 2 hours to form a clear and transparent homogeneous mixture and upon cooling was filtered to remove any foreign particulates and aged for 24 h at room temperature. Mn doped ZnO thin films were prepared by spin coating method. 2-Methoxyethanol was selected not only as a suitable solvent but, specifically, because of its dehydrating property. This de-hydrating capability plays an essential role in the removal of coordinated water from the precursor compound (basic zinc acetate) that promoted the formation of anhydrous oxide structure at room temperature.

A Brief outline of the preparation of thin films is as follows.

The aged solution prepared above was filled with in a syringe and 10 drops of the solution were poured on the quartz substrate fitted in the spin coater., and it was rotated at a speed of 3000 rpm for a duration of 30 s and then was dried at 200°C for 10 minutes. Then again the quartz substrate was fitted in the spin coater and more 10 drops of the solution were poured and again was rotated at speed of 3000 rpm for duration of 30 s and was dried at a temperature of 200°C for 10 minutes. Similarly proceeding in this way 12 coatings of the solution were prepared on the quartz substrate and the thin film so formed was annealed at different temperatures of 400°C, 500°C, 600ºC, 800ºC for 5 hours. So undoped film and doped films at different annealing temperatures were thus prepared. Figure (4) presents the schematic diagram of Mn doped ZnO thin film fabrication using the sol-gel method.



Flowchart showing the formation of Mn doped ZnO films

CONCLUSIONS

Zinc oxide (ZnO) is a very promising candidate for future thin-film technology. Mangenese (Mn) doped ZnO is a DMS material.This DMS sample is synthesized by a simple sol gel process followed by spin coating technique.

REFERENCES

[1] Powell,C.F., Oxley,J.H. and Blocher, J.M., 'Vapor deposition'.
[2] Hyeon, J. L., et. Al. Appl.Phys.Lett. 81(2002) 4020.
[3] Furdyna, J.K., J.Appl.Phys. R29(1988) 64.
[4]Ohno, H., Science 951(1998) 281.
[5]Z.K.Tang,P.Yu,G.K.Wong,M.Kawasaki,A.Ohtomo,H.Koinuma,and Y.Segawa,Solid State Commun. 459(1997) 103.
[6]T.Minami,H.Sato,H.Nanto,andS.Takata,Jpn.J.Appl.Phys.,Part2,L781(1985) 24.
[7]A.Ohtomo,et.al. Appl.Phys.Lett.72 (1998) 2466.
[8]C.H.Bates,W.B.White,and R.Roy,J.Inorg.Nucl.Chem. 397(1966) 28.
[9]T.Fukumura,Zhengwu Jin,A.Ohtomo,H.Koinuma, and M.Kawasaki, Appl.Phys.Lett. 21(1999) 75.
[10]Shubra Singh, P Thiyagarajan, K Mohan Kant D Anita,S Thirupathiah, N Rama Brajesh Tiwari M Kottaisamy and M S Ramachandra Rao J. Phys. D: Appl. Phys. 40 (2007) 6312–6327.

Rice Sample Percentage Purity by Image Processing

Jagdeep Singh Aulakh , Dr. V.K. Banga
Amritsar College of Engineering & Technology, Amritsar

***ABSTRACT*- The purpose of this paper is to find the percentage purity of hulled rice grain sample by image processing technique. Commercially the purity test of rice sample is done according to the size of the grain kernel (full, half or broken). The food grain types and their quality are rapidly assessed through visual inspection by human inspectors. The decision making capabilities of human-inspectors are subjected to external influences such as fatigue, vengeance, bias etc. with the help of image processing we can overcome that. By image processing we can also identify any broken grains mixed . Here we discuss the various procedures used to obtain the percentage quality of rice grains.**

Keywords:- Percentage Purity, Hulled rice grain , pixel Area

I. INTRODUCTION

The quality of the world's most important staple food crop can be determined based on the shape size and texture of the grain. In India the ever increasing population losses in handling and processing and the increased expectation of food products of high quality and safety standards there is need for the growth of accurate fast and objective quality determination of food grains. Now days we are using the chemical methods for the identification of rice grain seed varieties and quality. The chemical method used also destructs the sample used and is also very time consuming method. On the other hand the machine vision or the digital image processing is a non destructive method, it is also very fast and cheap process compared to the chemical method. In the early days of machine vision application to grain quality evaluation, Lai *et* al.(1986) suggested some pattern recognition techniques for identifying and classifying cereal grains.

The same researchers (Zayas *et* al., 1986) also applied the digital image analysis technique to discriminate wheat classes and varieties. Particle size distribution and shape analysis is required in several areas that deal with granular or particulate materials. Bulk properties of such materials are of importance and variations in the size and shape of particles can result in significant change in their properties and/or value. Image processing is a common tool for applications in particle characterization, metallurgy, agriculture, biotechnology, etc. Digital imaging systems have found increasing use in such analyses as they are economical, fast and accurate. One of the areas of digital imaging applications is in testing the quality of food materials. Image analysis systems for food materials are not only efficient but also non-destructive in sample handling (Sun, 2000). Brosnan and Sun (2002) have reviewed the use of computer vision systems in the agricultural and food industry. They have discussed areas like assessment of fruits, vegetables and nuts, as also the successful use of computer vision systems in the analysis of grain characteristics and the evaluation of foods such as meats, cheese and pizza. Digital image analysis has also been used to evaluate the effect of moisture content on cereal grains by studying its effect on the physical appearance and kernel morphology (Tahir et al., 2007). Recently, Moreda et al. (2009) have reviewed the use of nondestructive machine vision technologies for fruit and vegetable size determination.

II. IMAGE CAPTURE

Flat Bed Scanning (FBS) this process uses the desktop scanner. In this the rice grain is placed on the glass plate of the scanner and covered with a black sheet of paper.

Fig. 1 Captured Image

Fig.2 Captured Image 2

Digital camera of high pixel resolution rate can also be used. To collect image data the camera should be placed at a location situated with a plane normal to the object's path. The black background was used. The environment was controlled to improve the data collection with simple plain background. The images acquired were 319 x 300 pixels in size. Images were captured and stored in JPG format automatically. Through data cable these images has been transferred and then stored in disk.

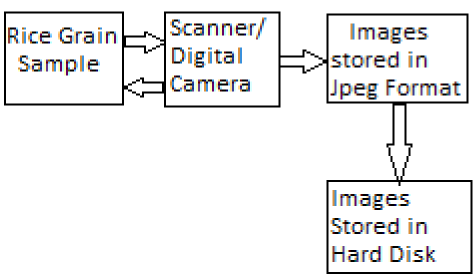


Fig.3 Basic Building Block for Image Capture

III. METHODOLOGY

a)In this first of all we set the level for the background and then subtract the image(fig.2) from this background. By practicing this we get the more uniform background.

b)Then we adjust the strechlimit of the obtained image from the last step. By this we get more contrast between the grains and the background, as we set the ratio of 0 and 1 for both the background and grains kernels.

c) Here we convert this image into the binary image (Fig.4) for performing other morphological operations.

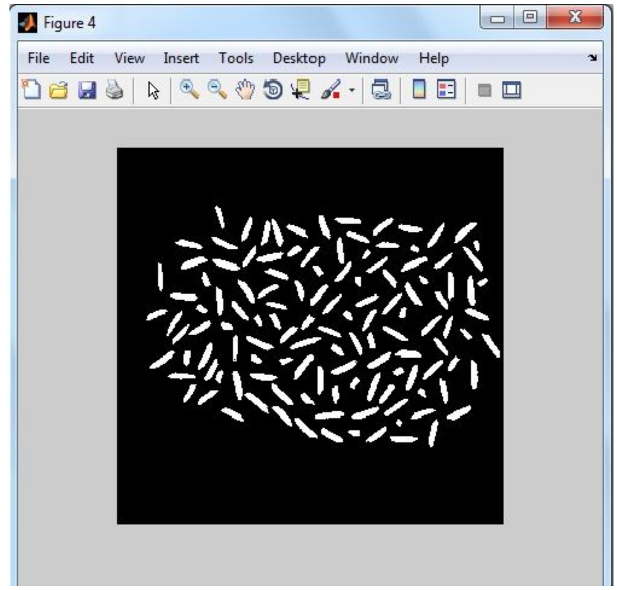


Fig.4 Binary Image

IV. PIXEL AREA OF GRAINS

Here we find the connected components of our final binary image. By this we get the following information Connectivity, Imagesize, Numobjects, PixelIdxlist about our final binary image.from this we also get the exact number of grain kernels present in our image.

After this we find the pixel area of the each grain present in our binary image.once we have the pixel area of each grain we can also map the pixel matrix of the whole image with the grain number for the grain area and zero for the rest of background. With the help of Matlab software we can show the number of rice grains in the sample image.

```
T =

    Connectivity: 8
       ImageSize: [300 319]
      NumObjects: 125
    PixelIdxList: {1x125 cell}
```

Fig.5 shows the output in command window

In the fig.5 we have number of grains and image size in pixels.

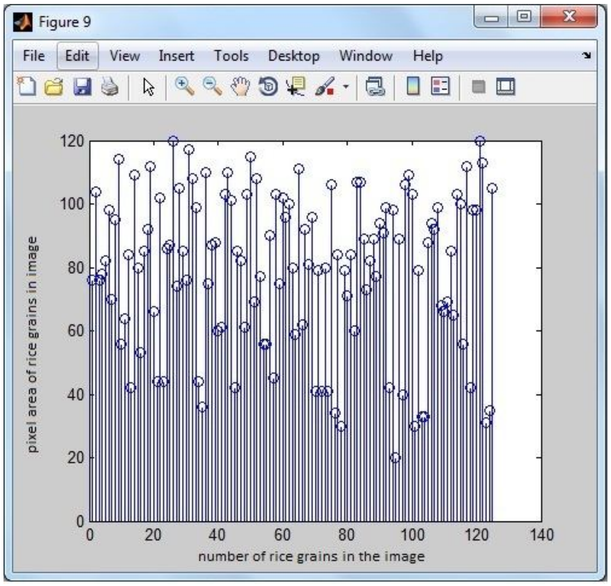


Fig.5 shows the output in stem graph.

In the above stem graph we have the plotting of pixel area of each rice grain corresponding to the number of rice grains present in the sample image.

V. REJECTING THE BROKEN GRAINS

We know from the data we have, that the broken or the half grain kernels occupies the lesser pixel area as compare to the healthy grains. So we set a threshold value of the pixel area for the average healthy grain kernel and the values lower than that of threshold will be discarded. Here we obtain a new binary image without the broken or the half grain kernel.

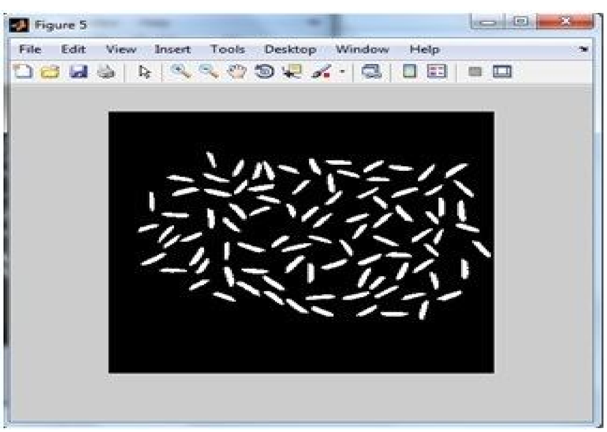


Fig.6 Resultant image with no broken grains.

By using this new binary image (fig.6) we can again find the connected components and get the information about Connectivity, Imagesize, Numobjects, PixelIdxlist . and with the help of the regionprops we can find the pixel area of each grain of this new binary image now we can also find the number of rice grains present in the resultant image.

```
N =

    Connectivity: 8
       ImageSize: [300 319]
      NumObjects: 92
    PixelIdxList: {1x92 cell}
```

Fig.7 Matlab output of resultant image i.e.(Fig 6)

Here also we have the number of rice grain kernels and the pixel size of image.

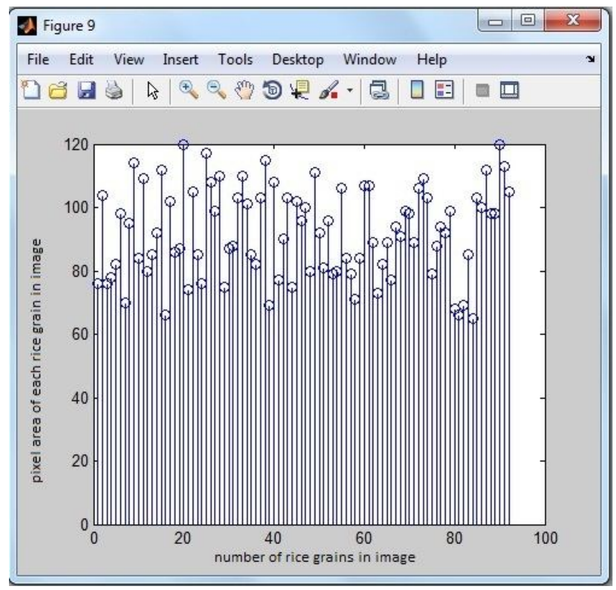


Fig.8 shows the output in stem graph.

In the above stem graph we have the plotting of pixel area of each rice grain after rejecting the broken rice grains corresponding to the number of rice grains present in the sample image.

VI. PERCENTAGE PURITY OF GIVEN SAMPLE

On studying both the outputs of Matlab command window and the stem graphs we can clearly see that the second output (fig.6,7&8) has more healthy grains and also the number of grains are lesser because half or broken grains are discarded. Now if we divide the number of grains in the second output by the number of grains in first output (fig.5&6) and multiply it by 100. We can achieve the percentage purity of the given sample.

VII. CONCLUSION

Here we conclude that purity percentage of rice samples can effectively be done by using the image processing techniques. With our coding in Matlab software we can calculate that how pure is our sample. The setup used is also very common and easily available. This is also more accurate than the human visual inspection. All this leads to better quality in food processing by image processing.

VIII. FUTURE WORK

For future work we have to find some alternative method where we not only compare the pixel area but also compare the length of each rice grain for more accurate results.

REFERENCES

[1] J. Paliwal, N.S. Visen, D.S. Jayas, N.D.G. White," Cereal Grain and Dockage Identification using Machine Vision" 2003

[2] B. S. Anami, D. G. Savakar,"Improved Method for Identification and Classification of Foreign Bodies Mixed Food Grains Image Samples" 2005

[3] N.S. Visen, J. Paliwal, D.S. Jayas and N.D.G. White Image "Analysis of bulk grain samples using neural networks" 2004

[4] Qing Yao, Jianhua Chen, Zexin Guan "Inspection of rice appearance quality using machine vision," Global Congress on Intelligent Systems. 2009.

[5] Siriluk Sansomboonsuk and Nitin Afzulpurkar , "The Appropriate Algorithms of Image analysis for Rice Kernel Quality Evalution ," 20th Conference of Mechanical Engineering Network of Thailand18-20 October 2006, Nakhon Ratchasima, Thailand

[6] Agustin, O.C., Byung-Joo Oh, " Automatic Milled Rice Quality Analysis", Future Generation Communication and Networking", 2008.

[7] G. van Dalen "Determination of the size distribution and percentage of broken kernels of rice using flatbed scanning and image analysis",Unilever Research Laboratory, Foods Research Centre, Olivier van Noortlaan 120, 3133 AT Vlaardingen, The Netherlands,

[8] S. Anami1, D. G. Savakar, "Improved Method for Identification and Classification of Foreign Bodies Mixed Food Grains Image Samples" 1K. L. E. Institute of Technology. Hubli- 580030, .India anami_basu@hotmail.com, ICGST-AIML Journal, ISSN: 1687-4846, Volume 9, Issue I, February 2009

[9] Aggarwal, Amit K. and Mohan, Ratan) "Aspect Ratio Analysis Using Image Processing for Rice Grain Quality", International Journal of Food Engineering: Vol. 6: Iss. 5, Article 8, 2010 .

[10]Sanjivani Shantaiya , Mrs.Uzma Ansari," Identification Of Food Grains And Its Quality Using Pattern Classification" M.tech (CSE) –IV Sem, RITEE, CSVTU, Raipur sanjivaninice@gmail.com Reader (CSE), RITEE, CSVTU, Raipur arshi.uzma@gmail.com, Special Issue of IJCCT Vol. 2 Issue 2, 3, 4; 2010 for International Conference [ICCT-2010], 3rd-5th December 2010

Jagdeep Singh Aulakh is working as a lecturer in Department of Electronics & Communication Engineering in Amritsar College of Engineering and Technology, Amritsar, Punjab, India. He did his B.Tech (Electronics and Communication Engineering) from Punjab Technical University, Jalandhar. India. and pursuing M.Tech (Electronics and Communication Engineering) from Amritsar College of Engineering and Technology, Amritsar, Punjab, India.

Vijay Kumar Banga is working as professor and Head of the Department of Electronics & Communication Engineering, Amritsar College of Engineering and Technology, Amritsar, Punjab, India. He obtained his B. E (Electronics and Instrumentation Engineering) from Punjabi University, Patiala, Punjab, India, M.Tech (Electronics and Instrumentation) from Panjab University , Chandigarh, India and Ph.D. in Electronics (Artificial Intelligence) form Thapar University, Patiala., India. Presently, he has 13 years of research and UG & PG teaching experience. He has 35 research papers to his credit in various international journals and conferences. He is member of Board of Studies in Electronics and Communication Engineering of Punjab Technical University, Jalandhar, Punjab, India. His areas of research and interest include Robotics, Artificial Intelligence, automated control systems and Genetic algorithms.

Survey of Anomaly Management System In VANET

Jagjit Singh, Neha Sharma
Amritsar College of Engineering and Technology, Amritsar
Email id: jagjitsingh_sandhu@yahoo.com, er.neha.ruchi@gmail.com

***Abstract*:. In this paper we have conducted a survey of all the possible vulnerbalties and attacks which might occur in VANET when the network becomes reality and various type anomaly management systems and their architecture. VANET is basically a technology that uses moving vehicles as nodes to create a mobile network. VANET turns every participating vehicle into wireless sensor nodes and which allowing vehicles to communicate with each other, create a network with wide range. As vehicle fall out of the signal range they drop out from the network. We mainly understand VANET by taking it subset of MANET.**

I. INTRODUCTION

The use of self-driving vehicles is becoming more and more significant in last few years. The fact that the vehicles are self-driving (whether autonomous or not), can lead to greater difficulties in identifying failure and anomalous states, since the operator cannot rely on its own body perceptions to identify failures. Moreover, as the autonomy of self-driving vehicles increases, it becomes more difficult for operators to monitor them closely, and this further increases the difficulty of identifying anomalous states, in a timely manner. People drive their vehicle to work, to go market, to visit friends and to many other places. Students take the bus to college. The economy depends largely on the goods that are delivered by trucks. This mobility is usually taken for granted by most people and they hardly realize that transportation forms the basis of our civilization. As the world grow and the population increases, more traffic is generated which has many adverse effects. Not having a proper transportation system costs people their time, safety and money. The need for a more efficient, balanced and safer transportation system is obvious. This need can be best met by the implementation of autonomous vehicle network or adhoc VANET. In the future automated vehicles will help to avoid accident and reduce congestion. The future vehicles will capable of determining the best route and warn each other about the conditions ahead.

II. ANOMALIES IN VANET

The various anomalies associated with VANET are tunnel attack, jamming, wormhole, forgery, denial of service attack, masquerading etc.

2.1 TUNNEL ATTACK: Since GPS signals dissappear in tunnels, hacker may destroy this temporary loss of positioning information to inject false data once the vehicle leaves the tunnel and before it receives an original position update as figure below illustrates. The physical tunnel in this example can also be replaced by an area jammed by the attacker, which results in the same effects.

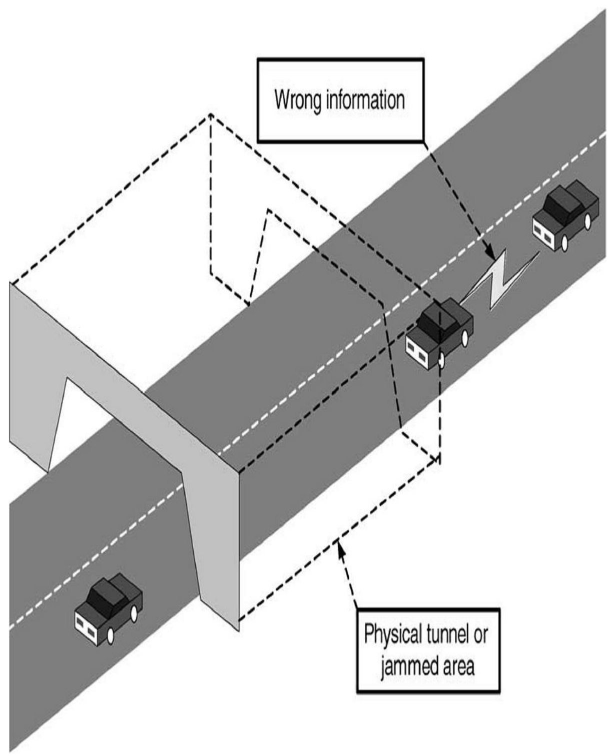


Fig 1: Tunnel attack

2.2 JAMMING: The jammer intentionally generates interfering transmissions that prevent communication within their reception range. As the network coverage area can be well-defined, at least locally, jamming is allow effort destroy opportunity.

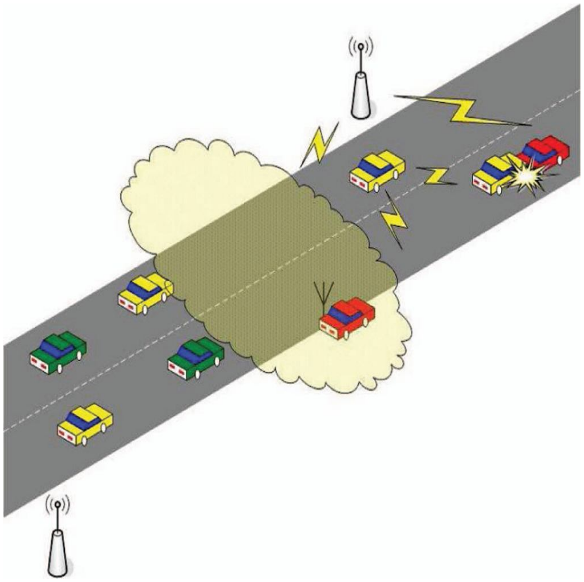


Fig 2: Jamming

2.3 FORGERY: The timely and correctness receipt of application data is a major problem. The figure illustrates the rapid "contamination" of large portions of the vehicular network coverage area with wrong information where a single hacker forges and transmit false hazard warning which are taken up by all vehicles in both traffic schmes.[2]

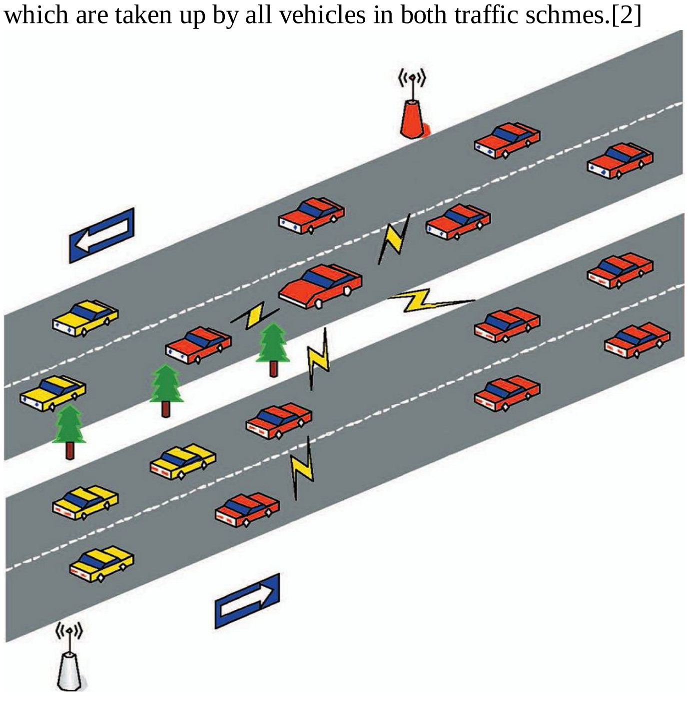


Fig 3: Forgery

2.4 DENIAL OF SERVICE ATTACK: Hackers may seek to start excessive actual requests in order to exhaust the resources of the AP. A general possible solution would be to limit the number of actual requests which can be processed in a unit of time period. This method can assurance that the server is not defeated by DoS. But this could also delay a request. In wireless environment, typically the attacker attacks the communication medium to cause the channel jam or to create some problems for the nodes from accessing the network. The basic purpose is to prevent the actual nodes from accessing the network services and from using the network resources. The attack may result in overtiredness of the nodes and networks resources. In VANET DOS shall not be allowed to happen, where seamless life critical information must reach its intended destination securely and time. [3]

2.5 MASQUERADING: The hacker actively pretend to be another vehicle by using erroneous identities and can be motivated by malicious or rational objectives.

III. IDS ANAMOLY MANAGEMENT SYSTEM AND ITS ARCHITECTURE

An IDS is a device or software application that monitor network or system activities for malicious activities or policy violations and produce report to a management station. An IDS evaluates a suspected intrusion once it has taken place it also watches for attacks that originate from within system. Considering the malicious road side attacker that is sending incorrect emergency braking warnings, then IDS detection process could be as follows .A first event of a wrong action within the active safety system would be discovered, if an emergency braking event is received from a previously unknown node. One would normally expect such an event to come from a previously known node. A second event might come from another vehicle that previously passed this area and also received the same kind of message. After passing the respective area, this car's intrusion detection system understands, that for itself, there was no emergency braking event and transfers this analysis to another follow up vehicles.

Combining above two events results in a strong evidence that the received warning message must be a fake warning. Thus, the intrusion detection system tells the active safety system to denies the warning and communicates the detection results to other nodes, especially follow up nodes. In order to actualise such an intrusion detection system in VANETs, we apply of a modular cross layer intrusion detection system. On every node, different modules are in charge of assembling audit data on different layers. A local decision module receives continuously audit data summaries from the other modules and analyses them with the aid of additional information, available from other non-network devices, such as radar, Global positioning system and sensors.

Expect from the central decision module, the different modules are: a monitoring module for the network and routing layer, a context information module, an application evaluation module, an action module and a module for communication with intrusion detection systems on other nodes.[4]

The monitoring module for network activity and routing is responsible for collecting data on communication within the node's communication range. This includes monitoring the neighboring nodes' forwarding attitude, as well as the generation of a list of current and previous neighbors, including their positions and movements, i.e. the creation of a network topology development report. The current neighboring nodes' position data can be verified by active probing messages, or Global Positioning System and sensor data from the context information module, in order to help to identify a malcious node behavior. On the application layer, the received warning messages are first evaluated by the application evaluation module, which uses knowledge from applications, in combination with sensor data provided by the context information module. This verification could enable, for instance, the detection of false icy road warnings, which would be quite difficult when the outside temperature is above 11 degree.

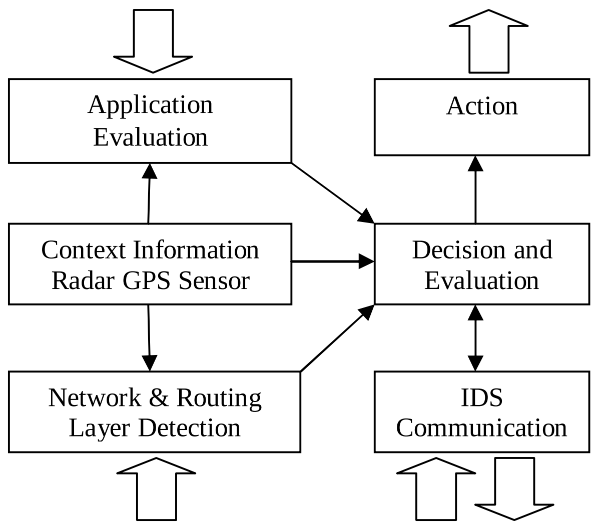


Fig 4: Cross Layer Architecture

The intrusion detection communication module is used to share the evaluated audit data with other nodes and forward other nodes' study results to the local decision module. Analyze results are either exchanged directly, via multiple hops, or via a temporarily accessible road network infrastructure. Especially in case of distributed node density, a potentially accessible road network infrastructure will help to communicate intrusion alerts to follow up vechiles. On the other hand, they help to detection requires several kind of trust between the participating nodes. Hence, we will use dynamic trust establishment between communicating nodes and establish first trust relations during a direct communication phase, to keep them for later usage.[5]

IV. CONCLUSION

Safety is a primary concern to many road users. Securing VANET communication is a crucial and serious issue, since failure to do so will delay the deployment of technology on the road. The safety requirements can be powerfully supported by many safety applications, such as traffic report and accidents notifications. In this paper we study and survey various types of anomalies present in vehicular adhoc network and intrusion detection system to detect such nodes.

REFERENCES

[1] Y.-C. Hu, A. Perrig and D.B. Johnson, Packet leashes: A defense against wormhole attacks in wireless networks, in: Proceedings of IEEE Infocom'03, 2003.

[2] Maxim Raya and Jean-Pierre Hubaux "Securing vehicular ad hoc networks", Journal of Computer security, IOS Press Amsterdam, The Netherlands, Volume 15, Issue 1(January 2007), pages 39-68

[3] J. Blum, A. Eskandarian, "The Threat of Intelligent Collisions", IT Professional, IEEE Computer Society, 2004.

[4] Y. Zhang and W. Lee, "Intrusion detection in wireless ad-hoc networks,"in Proceedings of the 6th annual international conference on Mobile computing and networking. ACM Press, 2000, pp. 275–283.

[5] Y. Zhang, W. Lee, and Y.-A. Huang, "Intrusion detection techniques formobile wireless networks," Wirel. Netw., vol. 9, no. 5, pp. 545–556, 2003 .

Multiple Load Balancing in AD Hoc Network

Minakshi[1], Tanu Preet Singh[2]
[1]Lect. Department of Computer & Science, Sai Institute of Engineering & Technology Amritsar
[2]Associate Professor Department of Computer Science & Technology Amritsar, 143001
Er.mina08@yahoo.com , tanupreet.singh@gmail.com

Abstract: **An ad-hoc network consists of a set of mobile nodes which are connected with each other by using radio waves. Load balancing is the process of improving the performance of a parallel. This network does not have any infrastructure or central administration, hence it is called infrastructure less network. As the nodes are mobile, it is very difficult to find the path between two end points. This paper presents a solution for finding path between nodes in mobile ad hoc network. The multipath routing protocol with load balancing provides a solution for the congestion network and increases it's capacity. To consider that the use of multiple paths simultaneously for transmission data allows to improve the network performance. A number of additional modifications are incorporated to the protocol in order to allow its smooth operation. Some of the parameters used to evaluate its performance are packet delays and throughput. The results of this algorithm shows better throughput as compared to existing algorithms. In this paper we present the performance analysis of various load balancing algorithms based on different parameters, considering two typical load balancing approaches static and dynamic. The analysis indicates that static and dynamic both types of algorithm can have advancements as well as weaknesses over each other. Deciding type of algorithm to be implemented will be based on type of parallel applications to solve.**

.Keywords—Load balancing (LB), Ad hoc network; multipath routing protocol; load balancing; AODV; DSR.

I. INTRODUCTION

A Mobile Ad hoc Network (MANET) consists in a collection of wireless mobile nodes, which form a temporary network without relying on any existing infrastructure or centralized administration [1]. The advantages of ad hoc networks are the convenience (no central administration), mobility, productivity, deployment and expandability. As the nodes in the network are mobile, the topology of network changes unpredictably. Hence it is difficult to generate path between two nodes. This paper deals with the development of on-demand ad-hoc network routing which can achieve load balancing for packet switched network. The algorithm is adaptive, distributed and is inspired by swarm intelligence. Ant algorithms are the class of optimizing algorithms under swarm intelligence (SI)[2][3]. Routing in ant algorithm [4][5] is through interaction of network exploration agents called ants. According to this algorithm, a group of mobile agents builds path between pairs of nodes by exchanging information and updating routing tables. MANET networks have several usages. First these networks were devised to be used in military applications. MANET networks are mostly used in survey, helping and saving operations, tracing and operations, scientific conferences.

The problem of mobile ad-hoc network (MANET) can be summarized in the answer of this question: how to find the route between the communicating end-points. One of the main reasons is that routing in MANETs is a particularly challenging task due to the fact that the topology of the network changes constantly and paths which were initially efficient can quickly become inefficient or even infeasible. Moreover, control information in the network is very restricted. This is because the bandwidth of the wireless medium is very limited, and the medium is shared. It is therefore important to design algorithms that are adaptive, robust and self-healing. Moreover, they should work in a localized way, due to the lack of central control or infrastructure in the network [6,8].

This paper is organized as follows: we start with a review of the routing on the ad hoc network. Then, we present the multipath routing and the load balancing on this network as well as some strategy proposed to put the emphasis on two multipath protocols MSR (Multipath Source Routing) and AOMDV *(*Ad Hoc On demand Multipath Distance Vector) [9]. Afterwards, we propose a solution called LB-AOMDV *(*Load Balancing-AOMDV*)* that support at once multipath routing and load balancing.

II. OVERVIEW OF PROTOCOLS

A simple **DSR** is a simple source routing protocol for MANETs, in which route caching is heavily used. If the route to the destination is not known, a route discovery process is initiated in order find a valid route. Route discovery is based in flooding the network with route request (RREQ) packets. Every mobile host that receives a RREQ packet checks the contents of its route cache, and if it is the destination or it has a route to the destination it replies to the RREQ with a route reply (RREP) packet that is routed back to the original source. In case none of the above holds, the host that received the RREQ re-broadcasts it to its neighborhood. In this way the RREQ message is propagated till the destination. Note that both RREQ and RREP packets are also source routed. The RREQ message maintains the path traversed across the network allowing thus the RREP message to route itself back to the source by traversing the recorded path backwards. The route carried back by the RREP packet is cached at the source for future use. If any link on a source route is broken, the source host is notified with a special route error (RERR) packet from intermediate nodes. When the source gets this packet removes any route using this link from its cache. More details and enhancement to this basic DSR operation can be found in [9].

SCTP was recently adopted by IETF, and is a reliable transport protocol that operates on top of a connectionless packet based network such as IP. One of the most important new ideas that SCTP introduced is that of multi-homing. A single SCTP association (session), is able to use alternatively anyone of the available IP-addresses without disrupting an ongoing session. However, this feature is currently used by SCTP only as a backup mechanism that helps recovering from link failures.

SCTP maintains the status of each remote IP address by sending Heartbeat messages and it is thus able to detect a specific link failure and switch to another IP address. Another novel feature is that SCTP decouples reliable delivery from message ordering by introducing the idea of *streams*. The stream is an abstraction that allows applications to preserve in order delivery within a stream but unordered delivery across streams. This feature avoids HOL blocking at the receiver in case multiple independent data streams exist in the same SCTP session. Congestion control was defined similar to TCP, primarily for achieving TCP friendliness [10].

In this paper, we propose two methods to Improve the Ad-Hoc On-Demand Distance-Vector (**AODV**) protocol. The main goal in the design of the protocol was to reduce the routing overhead, buffer overflow, end-to-end delay and increase the performance. A multi-path routing protocol is proposed which is based on AODV and Ant Colony Optimization (ACO). This protocol is refereed to Multi-Route AODV Ant routing (MRAA). Also we propose a load balancing method that uses all discovered paths simultaneously for transmitting data. In this method, data packets are balanced over discovered paths and energy consumption is distributed across many nodes through network.

III. ROUTING ON AD HOC NETWORK

The routing is a method which attends to forward the information to destination along the network. It consists to determine an optimal forwarding for packets along the network according to certain criteria (hop number, e.g). The problem consist to find the investment with minimum cost of nominal capacity and reserve that provide the routing of nominal traffic and guarantee its reliability in case of any failure of link or node.

In this method the performance [11] [12] of the processors is determined at the beginning of execution. Then depending upon their performance the work load is distributed in the start by the master processor. The slave processors calculate their allocated work and submit their result to the master. A task is always executed on the processor to which it is assigned that is static load balancing methods are non-preemptive. The goal of static load balancing method is to reduce the overall execution time of a concurrent program while minimizing the communication delays. A general disadvantage of all static schemes is that the final selection of a host for process allocation is made when the process is created and cannot be changed during process execution to make changes in the system load.

4. AOMDV Protocol : Ad hoc on demand multipath distance vector

To reduce interruption of communications in ad hoc network, the discover procedure of routes must be efficient specially with the continuous mobility of the nodes and also the frequent change of network topology, many routing protocols are proposed such as AOMDV: the multipath routing protocol [13] that extends the single path AODV protocol to compute multiple path routing.

A. Routing Definition

The main idea in AOMDV is to compute multiple paths during route discovery procedure for contending link failure. In fact, the main goal to concept this protocol is to search multiple routes during the same route discovery procedure, but only the best path based on some metric (number of hop) is chosen and is used for data transmission between source and destination. The other paths are used only when the primary path fails. This protocol is intended for ad hoc network where the mobility of nodes is very important and consequently the route breaks frequently. AOMDV use the information available in AODV, but to compute multiple paths it adds additional number of control packet “overhead”. AOMDV is based on two essential mechanisms:

• A route update to establish and maintain multiple loop-free paths at each node.
• A distributed protocol to find link-disjoint paths.

B. Multipath route construction without loop-free

AOMDV is based on the advertised hopcount [14]. The advertised hopcount of a node *i* for a destination *d* represents the maximum hopcount of the multiple paths for *d* available at *i*. The protocol only accepts alternate routes with hopcount lower than the advertised hopcount, alternate routes with higher or the same hopcount are discarded. This condition is necessary to guarantee loop-freedom. Figure shows the structure of the routing table entries for AODV and AOMDV.

• In AOMDV, advertised_hopcount replaces hopcountin AODV.
• A *route_list* replaces *nexthop* and essenntially defines multiple next hops with respective hopcounts.

C. Computing Multiple Loop free Paths

AOMDV allow building multiple link disjoint paths. It ensures multiple paths without common link between routes from source to destination. Additional modifications are made in the route discovery process to allow formation of node-disjoint paths from intermediate nodes to the source and destination. AOMDV add a new field called "first hop" for every *RREQ* packet. This field indicate the first hop (neighbor to node source) to set. In addition, each node maintains a list, firsthop_list, for each RREQ to keep track of the list of neighbors of the source through which a copy of the RREQ has been received. At the intermediate nodes, unlike in AODV, duplicate copies of RREQ are not immediately discarded. Each copy is examined to see if it provides a new node-disjoint path to the source. This is ascertained by examining the first hop field in the RREQ copy and the firsthop_list in the node for the RREQ. If it does provide a new path, the AOMDV route update rule is invoked to check if a reverse path can be setIn the round robin [13] processes are divided evenly between all processors. Each new process is assigned to new processor in round robin order. The process allocation order is maintained on each processor locally independent of allocations from remote processors. With equal workload round robin algorithm is expected to work well. Round Robin and Randomized schemes [12] work well with number of processes larger than number of processors. Advantage of Round Robin algorithm is that it does not require inter-process communication. Round Robin and Randomized algorithm both can attain the best performance among all load balancing algorithms for particular special purpose applications. In general Round Robin and Randomized are not expected to achieve good performance in general case.

D. AOMDV problems

In such protocols a link failure in the primary path, through which data transmission is actually taking place, causes the source to switch to an alternate path instead of initiating another route discovery. A new route discovery occurs only when all precompiled paths break. The problem with these Multipath protocols [15] is that though during the route discovery process multiple paths are discovered, only the best path based on some metric is chosen and is used for data transmission. The other paths are used only when the primary path fails.

Actually, the compute and the maintenance of multipath between source and destination require a very important occupation of routing table, achieve tremendously memory resource at every node and increase the heading packet size. These constitute a handicap, in view that we have only one path to transmit.

V. IMPROVEMENT TO MULTIPLE PATH PROTOCOLS

In this part, we propose an extension to AOMDV protocol in order to support certain mechanism and technique to improve its performance. AOMDV can allow finding many routes between source and destination during the same route discovery procedure but only one path is used to transmit data. When the source receives one or many *RREP* packets from many disjoint paths, it decides:

- If one *RREP* is received, therefore only one route layout from source to destination is used to send data packets.
- If many *RREP* are received, the source chooses the best route based on the short number of "hop count". The other routes remain waiting the *RERR* packet that indicates the failure of the principal route; in this case the best path from alternate paths is used to transmit data. We provide some modifications to this routing decision in order to AOMDV protocol uses many routes between source and destination and load balancing in the network.

VI. A NEW PROPOSED METRIC

The AOMDV protocol selects the route with the lower hop count to forward data. However, the less congestion routes can provide short end to end delay than routes providing lower hopcount. To choose the less congestion routes, we need a new metric which allow source node to select the less congestion routes. For this reason, we propose a new metric which achieve load balancing between the selected routes to take into account the number of active paths through every nodes according to the following The division with *np* hops, forming the route *p*, ensures that the metric takes into account the hopcount number to estimate the traffic load.

```
If (no route to destination)
{
Initiate route discovery as in AOMDV;
}
If (single known route)
{
Forward data packet to specified route;
}
Else
{
Forward data packet to best route;
}
If (no route to destination)
{
Initiate route discovery as in AOMDV;
}
If (single known route)
{
Forward data packet to specified route;
}
Else
*/if N routes are known from source to destination*/
{
Distribute forwarding data packet to less congestion routes;
}
```

The Route maintenance is similar to AOMDV. In such protocols, link failures in the primary path, through which data transmission is actually taking place, cause the source to witch to an alternate path instead of initiating another route discovery. A new route discovery occurs only when all pre computed paths break.

To build the LB-AOMDV protocol, we redefine the structure of RREP packet by adding a new field called buffer_size which take into account the traffic load on the route. This traffic load is expressed as the sum of buffer_size of intermediate nodes for each route between source and destination. When an intermediate node receives a RREP packet, it increments the new field with the size of its buffer.

On the other hand, when the source receives RREP packet, it divides the value of the buffer_size field by the hopcount of each route between source and destination in order to have the congestion level. The algorithm to compute the congestion level of each route between source and destination is as follows:

This algorithm is executed between source and destination to select a list of less congestion routes. The new structure of routing table entries for LB-AOMDV is shown in Figure. We still add another additional field buffer_size in the route_list. Each node sorts the route_list field by the descending value of buffer size. Each node sends data packets by using the route with the minimal buffer size. The LB-AOMDV protocol establishes three paths between source and destination nodes. The packets sent by source node are scheduled according to Round-Robin (RR) algorithm [16].

VII. PERFORMANCE EVALUATION

We use NS2 to simulate our LB-AOMDV protocol. For the initial simulations and the validation of the system the following parameters have been chosen:

All nodes have the same transmission range of 200 meters. The mobility model selected is the *random waypoint* model. In this mobility model, a node moves in the direction of the destination with a speed uniformly chosen between the minimal speed and maximal speed.

A. Parameter to evaluate With the aim to evaluate the LB-AOMDV performance, we compare the simulations of the single path routing protocol such as AODV, DSR, and the multipath routing source destination sequence number hop count timeout **buffer_size** destination

if (node A receives RREP)
{
if (node A not the source)
{
Buffer_size += buffer_size of node A ;
}
if (node A is the source)
{
*/compute the congestion level of route i from source to destination */
Congestion_ level(i)= buffer_size(i) / hopcount(i) ;
}
}

B. Simulation results

1) Average end-to-end delay versus the network load From figure 2, we note the increase of the average end to end delay according to the network load for all the routing protocols. The LB-AOMDV protocol is the most efficient because, under heavy load (40 connections) its average end to end delay is about 26% less than AODV protocol, 21% less than DSR protocol and 4% less than MSR protocol.

We explain the effect of the end-to-end delay decrease of the LB-AOMDV protocol by the use of the *less congestion route selection* mechanism which distribute traffic load fairly across routes selected between source and destination.

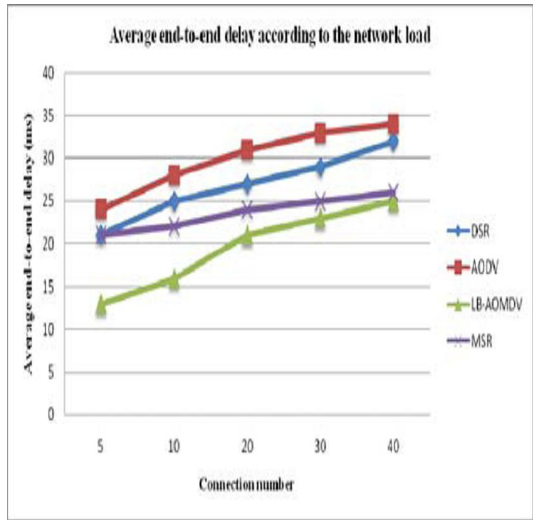


Fig 2: Average end-to-end delay versus the network load

2) *Average buffer size versus the network load*

Figure 3 shows that the average buffer sizes increase according to the network load for all the routing protocols. According to this figure, we note that the multipath routing protocols have less loaded buffers then single path routing protocols. We notice that our protocol reduces the congestion level of the network and increases its capacity.

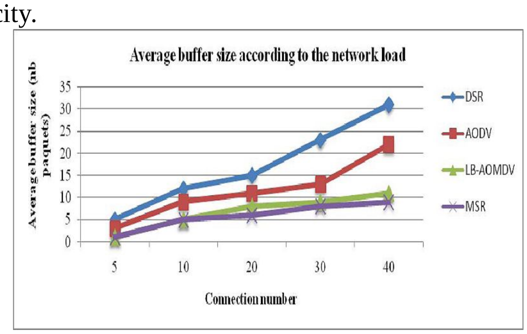


Fig 3. *Average buffer size versus the network load*

3) Traffic Overhead versus the network load The observation of figure 4 shows that our protocol generates the highest traffic overhead. When the number of connections is equal to 5, the traffic overhead produced by all protocols is low. This traffic increases significantly when the network load increases (till 40 connections). The traffic overhead (TOH) generated by DSR protocol under heavy load (40 connections) is about 60% less than that generated by LB-AOMDV protocol. TOH generated by DSR is 50% less than that generated by MSR and AODV protocols. We can

explain these results by the use of high number of control packets to search and maintain routes belonging to multipath routing protocols.

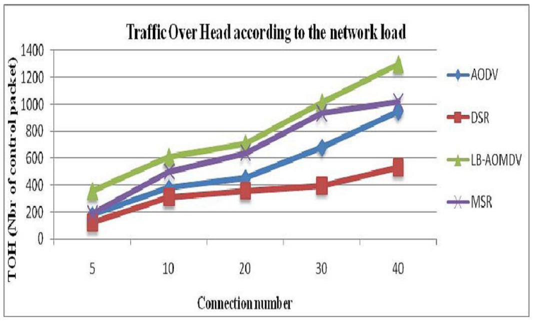


Fig 6. *TOH versus the network load*

CONCLUSION

In this work, we have studied multipath routing protocols in this paper. We have focused on load balancing mechanism to fairly distribute the traffic on different active routes selected between source and destination nodes. To select the less congested routes, we have proposed a new multipath routing protocol called LB-AOMDV with a new metric which is the buffer size. Among the performance evaluation of different routing protocols simulated: DSR, AODV and MSR, we conclude that our protocol: LB-AOMDV improves the network performance in terms of: capacity and congestion level compared to MSR and the single path routing protocols under heavy loaded network.

In the future work, we would like to include another metric to our protocol which is QoS.

REFERENCES

[1] Abolhasan, M., Wysocki, T., and Dutkiewicz, E., "A review of routing protocols for mobile ad hoc etworks," Ad Hoc Networks 2, pp. 1-22 (2004).

[2] E. Bonabeau, M. Dorigo, and G. Théraulaz, Swarm intelligence: from natural to artificial systems, Oxford University Press, 1999.

[3] T. White, "Swarm intelligence and problem solving in telecommunications", Canadian Artificial Intelligence Magazine, spring, 1997.
International Journal of Next-Generation Networks (IJNGN),Vol.1, No.1, December 2009.

[4] G. Di Caro and M. Dorigo, "Mobile agents for adaptive routing", Proc. 31st Hawaii International Conference on System Sciences, IEEE Computer Society Press, Los Alamitos, CA, pp. 74-83, 1998.

[5] Schoonderwoerd R, Holland O, Bruten J, Rothkrantz L. "Ant-Based load Balancing in telecommunications networks, Adaptive Behavior Hewlett-Packard Laboratories, Bristol-
England, pp 162-207, 1996.

[6] G. Di Caro and M. Dorigo, *AntNet: distributed stigmergetic control for communications networks*, Journal of Artificial Intelligence Research, 9 (1998), 317–365.

[7] G. Di Caro, F. Ducatelle, and L. M. Gambardella, *AntHocNet: an adaptive nature-inspired algorithm for routing in mobile ad hoc networks*, Tech. Report IDSIA-27-04-2004, Dalle Molle Institute for Artificial Intelligence (IDSIA), Manno-Lugano, Switzerland, September 2004.

[8] M. Dorigo and G. Di Caro, *The ant colony optimization metaheuristic*,in New Ideas in Optimization, D. Corne, M. Dorigo, and F. Glover, eds., McGraw-Hill, London, UK, 1999, 11–32. [8] Macker. J and Corson. S, Mobile ad hoc networks (MANET), 1997, http://www.ietf.org/html.charters/manet/charter.html.

[9] YuHua Yuan, Hui Min Chen, and Min Jia "An optimized Ad hoc On-Demand Multipath Distance Vector (AOMDV) Routing Protocol," Communications, 2005 Asia-Pacific Conference on Volume, Issue, 03-05 Oct. 2005 Page(s):569 – 573.

[10] R.R. Stewart et al., Stream control transmission protocol, RFC 2960, October 2000.

[11] Derek L. Eager, Edward D. Lazowska , John Zahorjan, "Adaptive load sharing in homogeneous distributed systems", IEEE Transactions on Software Engineering, v.12 n.5, p.662-675, May 1986.

[12] R. Motwani and P. Raghavan, "Randomized algorithms", ACM.

[13] Erik Andersson, Magnus Andersson, Mattias Flodin, Peter Gardfjäll, Alexander Hellström, "Understanding AOMDV routing in practice", May 15, 2003

[14] Marina M. K., Das S. R., "Ad hoc On- demand Multipath Distance Vector Routing" , Computer Science Department, Stony Brook University,2003

[15] Mueller S., P. Tsang, and D Ghosal, "Multipath Routing in Mobile Ad Hoc Networks: Issues and Challenges" , IEEE conference on Computer Communications (INFOCOM 2003)..

[16] Paul Southerington, "The Smoothed Round-Robin Scheduler", Member, IEEE, ECE742, 28 APRIL 2005

Reactive and Proactive Routing Comparison in AD-HOC Networks Using NS-2 Simulator

Neha[1], Tanu Preet Singh[2]
[1]Lecturer, Department of Computer Science & Engineering, SIET, Amritsar
[2]Associate Professor, Department of Computer Science & Engineering, ACET, Amritsar
neha_rkb_87@yahoo.co.in, tanupreet.singh@gmail.com

Abstract:- **Mobile Ad hoc Networks are the special type of wireless networks, where mobile nodes are connected through wireless interfaces forming a temporary network. They don't need fixed infrastructure. Due to higher mobility in nodes and dynamic infrastructure of MANETS, Routing is important issue in ad hoc networks. MANETS is classified in two types of reactive and proactive routing protocols. Many routing protocols of these two types have been proposed so far. Amongst the most popular ones are Ad hoc On-demand Distance Vector (AODV), Destination-Sequenced Distance-Vector Routing protocol (DSDV), Dynamic Source Routing Protocol (DSR), and Optimum Link State Routing (OLSR). Despite the popularity of those protocols, research efforts have not focused much in evaluating their performance when applied to variable bit rate (VBR). In this paper we present our observations regarding the performance comparison of AODV, DSR, DSDV, OLSR routing protocols for VBR in wireless or mobile ad hoc networks (MANETs). It performs simulations using NS-2 simulator. The comparison studies have shown that reactive protocols perform better than proactive protocols. Further DSR has performed well for the performance parameters namely delivery ratio and routing overload while AODV performed better in terms of average delay.**

Keywords- Reactive and Proactive routing protocols, MANET, Network Simulator (NS-2) and VBR Traffic.

I. INTRODUCTION

A mobile ad-hoc network(MANET) is a self-configuring network of mobile routers (and associated hosts) connected by wireless links." [1]. Some of the main features of MANET are listed below [2]:

a) MANET can be formed without any preexisting infrastructure.
b) It follows dynamic topology where nodes may join and leave the network at any time and the multi-hop routing may keep changing as nodes join and depart from the network.
c) It does have very limited physical security, and thus increasing security is a major concern.
d) Every node in the MANET can assist in routing of packets in the network.
e) Limited Bandwidth & Limited Power.

1.1 Routing Protocols

Ad-Hoc network is called as Mobile Ad-Hoc Network (MANET) because of mobility of nodes in network. They are IBSS (Independent Basic Service Set), because they does not need AP(Access Point) for communication in nodes. MANETs is a self-configuring network and form an uninformed topology. These nodes behave like routers in network to route the packet. MANETs are used in those areas where wire and wireless infrastructures are unreachable. Due to rapid change of topology in MANETs, MANETs routing protocols are required. The routing protocol is required whenever the source needs to communicates with destination. Routing protocols are classified as Proactive (Table Driven Routing Protocol), Reactive (On Demand Routing Protocol) and Hybrid (having the advantages of both proactive and Reactive routing protocols) routing protocols.
MANETs routing protocols are classified as:-

A. Reactive protocols
B. Proactive protocols

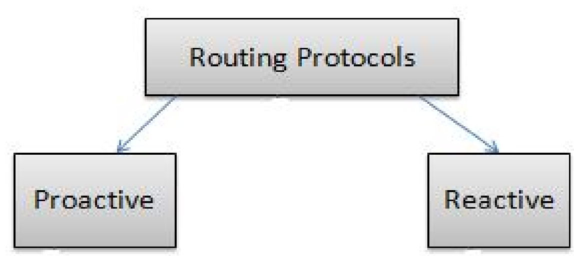


Fig 1. Types of Routing Protocols

This paper presents a performance comparison of two types of reactive and proactive routing protocols in MANET based on results analysis obtained by running simulations with different scenarios in Network Simulator version 2 (NS-2) [3]. A description of considered Reactive routing protocols is given in Section 2. A description of considered Proactive routing protocols is given in Section 3.A characteristics comparison of all considered routing protocols is given in Section 4. Simulation based analysis are described in Section 5 and finally Section 6 gives the conclusion and future scope of this paper.

II. REACTIVE ROUTING PROTOCOLS

Every node in this routing protocol maintains information of only active paths to the destination nodes. A route search is needed for every new destination therefore the communication overhead is reduced at the expense of delay to search the route. Rapidly changing wireless network topology may break active route and cause subsequent route search [4].

Examples of reactive protocols are:

a) Ad hoc On-demand Distance Vector Routing (AODV).
b) Dynamic Source Routing (DSR).
c) Location Aided Routing (LAR).
d) Temporally Ordered Routing Algorithm (TORA).

2.1 AD HOC ON DEMAND DISTANCE VECTOR (AODV)

The Ad hoc On Demand Distance Vector (AODV)[5] routing algorithm is a routing protocol designed for ad hoc mobile networks. AODV is capable of both unicast and multicast routing. It is an on demand algorithm, meaning that it builds routes between nodes only as desired by source nodes. It maintains these routes as long as they are needed by the sources. Additionally, AODV forms trees which connect multicast group members. The trees are composed of the group members and the nodes needed to connect the members. AODV uses sequence numbers to ensure the freshness of routes. It is loop-free, self-starting, and scales to large numbers of mobile nodes. The AODV protocol uses route request (RREQ) messages flooded through the network in order to discover the paths required by a source node. An intermediate node that receives a RREQ replies to it using a route reply message only if it has a route to the destination whose corresponding destination sequence number is greater or equal to the one contained in the RREQ. The RREQ also contains the most recent sequence number for the destination of which the source node is aware. A node receiving the RREQ may send a route reply (RREP) if it is either the destination or if it has a route to the destination with corresponding sequence number greater than or equal to that contained in the RREQ. If this is the case, it unicasts a RREP back to the source. Otherwise, it rebroadcasts the RREQ. Nodes keep track of the RREQ's source IP address and broadcast ID. If they receive a RREQ which they have already processed, they discard the RREQ and do not forward it. As the RREP propagates back to the source nodes set up forward pointers to the destination. Once the source node receives the RREP, it may begin to forward data packets to the destination. If the source later receives a RREP containing a greater sequence number or contains the same sequence number with a smaller hop count, it may update its routing information for that destination and begin using the better route. As long as the route remains active, it will continue to be maintained. A route is considered active as long as there are data packets periodically traveling from the source to the destination along that path. Once the source stops sending data packets, the links will time out and eventually be deleted from the intermediate node routing tables. If a link break occurs while the route is active, the node upstream of the break propagates a route error (RERR) message to the source node to inform it of the now unreachable destination.

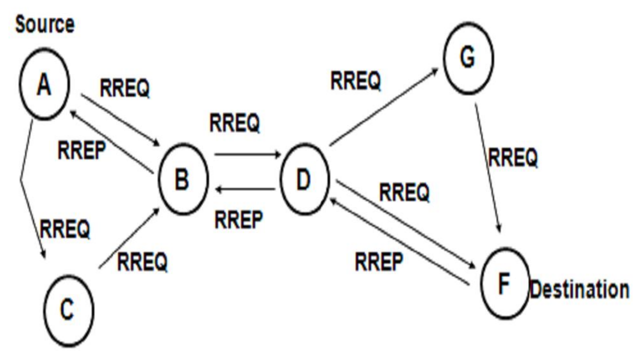


Fig 2. Routing in AODV

2.2 DYNAMIC SOURCE ROUTING (DSR)

Dynamic Source Routing(DSR)[6] is a routing protocol for wireless mesh networks and is based on a method known as *source routing*. It is similar to AODV in that it forms a route on-demand when a transmitting computer requests one. Except that each intermediate node that broadcasts a route request packet adds its own address identifier to a list carried in the packet. The destination node generates a route reply message that includes the list of addresses received in the route request and transmits it back along this path to the source. Route maintenance in DSR is accomplished through the confirmations that nodes generate when they can verify that the next node successfully received a packet. These confirmations can be of link-layer acknowledgements and passive acknowledgements or network-layer acknowledgements specified by the DSR protocol. However, it uses source routing instead of relying on the routing table at each intermediate device. When a node is not able to verify the successful reception of a packet it tries to retransmit it. When a finite number of retransmissions fail, the node generates a route error message that specifies the problematic link, transmitting it to the source node. When a node requires a route to a destination, which it doesn't have in its route cache, it broadcasts a *Route Request (RREQ)* message, which is flooded throughout the network. The first RREQ message is a broadcast query on neighbors without flooding. Each RREQ packet is uniquely identified by the *initiator's address* and the *request id.* A node processes a route request packet only if it has not already seen the packet and its address is not present in the route record of the packet. This minimizes the number of route requests propagated in the network. RREQ is replied by the destination node or an intermediate node, which knows the route, using the *Route Reply (RREP)* message. The return route for the RREP message may be one of the routes that exist in the route cache (if it exists) or a list reversal of the nodes in the RREQ packet if symmetrical routing is supported. In other cases the node may initiate it owns route discovery mechanism and piggyback the RREP packet onto it. Thus the route may be considered unidirectional or bidirectional. DSR doesn't enforce any use of periodic messages from the mobile hosts for maintenance of routes. Instead it uses two types of packets for route maintenance: Route Error (RERR) packets and ACKs. Whenever a node encounters fatal transmission errors so that the route becomes invalid, the source receives a RERR message. ACK packets are used to verify the correct operation of the route links. This also serves as a passive acknowledgement for the mobile node. DSR enables multiple routes to be learnt for a particular destination.DSR does not require any periodic update messages, thus avoiding wastage of bandwidth.

III. PROACTIVE ROUTING PROTOCOLS

In proactive routing scheme every node continuously maintains complete routing information of the network. This is achieved by flooding network periodically with network status information to find out any possible change in network topology. Current routing protocol like Link State Routing (LSR) protocol (open shortest path first) and the Distance Vector Routing Protocol (Bellman-Ford algorithm) are not suitable to be used in mobile environment. Destination Sequenced Distance Vector Routing protocol (DSDV) and Wireless routing protocols were proposed to eliminate counting to infinity and looping problems of the distributed Bellman-Ford algorithm.
Examples of Proactive Routing Protocols are: [7]

a) Destination Sequenced Distance Vector Routing (DSDV).
b) Optimized Link State Routing Protocol (OLSR).
c) Global State Routing (GSR).
d) Hierarchical State Routing (HSR).

3.1 DESTINATION - SEQUENCED DISTANCE VECTOR ROUTING (DSDV)

Destination-Sequenced Distance-Vector Routing (DSDV)[8][9] is a table-driven routing scheme for ad hoc mobile networks based on the Bellman-Ford algorithm. The improvement made to the Bellman-Ford algorithm includes freedom from loops in routing tables by using sequence numbers. It was developed by C. Perkins and P. Bhagwat in 1994. The DSDV protocol can be used in mobile ad hoc networking environments by assuming that each participating node acts as a router. Each node must maintain a table that consists of all the possible destinations. In this routing protocol, an entry of the table contains the address identifier of a destination, the shortest known distance metric to that destination measured in hop counts and the address identifier of the node that is the first hop on the shortest path to the destination. Each mobile node in the system maintains a routing table in which all the possible destinations and the number of hops to them in the network are recorded. A sequence number is also associated with each route/path to the destination. The route labeled with the highest sequence number is always used. This also helps in identifying the stale routes from the new ones, thereby avoiding the formation of loops. Also, to minimize the traffic generated, there are two types of packets in the system. One is known as "full dump", which is a packet that carries all the information about a change. However, at the time of occasional movement, another type of packet called "incremental" will be used, which will carry just the changes, thereby, increasing the overall efficiency of the system. DSDV requires a regular update of its routing tables, which uses up battery power and a small amount of bandwidth even when the network is idle. Whenever the topology of the network changes, a new sequence number is necessary before the network re-converges; thus, DSDV is not suitable for highly dynamic networks.

3.2OPTIMIZED LINK STATE ROUTING (OLSR)

Optimized Link State Routing (OLSR)[10][11] protocol is a proactive routing protocol where the routes are always immediately available when needed. OLSR is an optimization version of a pure link state protocol in which the topological changes cause the flooding of the topological information to all available hosts in the network. OLSR may optimize the reactivity to topological changes by reducing the maximum time interval for periodic control message transmission. Furthermore, as OLSR continuously maintains routes to all destinations in the network, the protocol is beneficial for traffic patterns where a large subset of nodes are communicating with another large subset of nodes, and where the [source, destination] pairs are changing over time. OLSR protocol is well suited for the application which does not allow the long delays in the transmission of the data packets. The best working environment for OLSR protocol is a dense network, where the most communication is concentrated between a large numbers of nodes. OLSR reduce the control overhead forcing the MPR(multipoint relaying) to propagate the updates of the link state, also the efficiency is gained compared to classical link state protocol when the selected MPR set is as small as possible. But the drawback of this is that it must maintain the routing table for all the possible routes, so there is no difference in small networks, but when the number of the mobile hosts increase, then the overhead from the control messages is also increasing. This constrains the scalability of the OLSR protocol. The OLSR protocol work most efficiently in the dense networks.

IV. COMPARISON OFAODV,DSR,DSDV,OLSR

A comparison of the characteristics of the above proactive and reactive ad hoc routing protocols DSDV, DSR, AODV and OLSR is given in following Table 1.

Protocol Property	DSDV	DSR	AODV	OLSR
Multicast Routes	No	Yes	No	Yes
Distributed	Yes	Yes	Yes	Yes
Unidirectional Link Support	No	Yes	No	Yes
Multicast	No	No	Yes	Yes
Periodic Broadcast	Yes	No	Yes	Yes
QoS Support	No	No	No	Yes
Routes Maintained In	Route Table	Route Cache	Route table	Route table
Reactive	No	Yes	Yes	No

Table 1. Comparison of protocols

V. SIMULATION BASED ANALYSIS

5.1 Simulation Environment

We have used network simulator version 2 (NS-2)[12] for simulation, most widely used network simulator. We simulated network for simulation time of 1000 sec and area of 1000 m *1000 m.

5.2 Simulation Results

In this section we analyze the ad- hoc networks proactive and reactive routing protocols performance results with variable bit rate (VBR). Most of the previous work is limited on performing simulations for ad –hoc networks with a constant bit rate(CBR)[13][14][15][16]. Our work differ in that we use VBR. We have used Average Delay, Delivery Ratio and Normalized Routing Overload as performance parameters while varying network parameters such as Pause Time and Number of Nodes.

5.2.1 Effect of Varying Pause Time

Pause time can be defined as time for which nodes waits on a destination before moving to other destination. We used this as a parameter as it is measure of mobility of nodes. Low pause time means node will wait for less time thus giving rise to high mobility scenario. Figure 3 (3a, 3b, 3c) shows various performance parameters v/s pause time when other parameters were constant. From figure we can observe that normalized overload for DSDV and OLSR is almost constant. This is because of their proactive nature due to which they offer constant routing overhead in all cases. While for reactive protocols considered here as we increased pause time routing overload has decreased .This is because as routing pause time increases mobility decreases and thus link breakage become rare which in turn will decrease number of route request from sources and hence decreasing overhead. Also DSR outperformed AODV as it maintains multiple routes to a destination. In case of failure in one route other route will be used rather than initiating route request. Also from figure we can see that average delay for proactive protocol was better at high mobility as they use route already in the table, and no time is required to find route as opposite to reactive protocols as they will wait for route formation. But at lower mobility, we can observe that reactive protocols performed better in terms of average delay among which AODV outperformed DSR. This is because DSR may not use optimum path always unlike AODV. While delivery ratio for DSR and AODV was near to 100% with DSR performing better because of multiple path information in its route cache (AODV always stores best path). Also proactive protocols performed poor in case of high mobility.

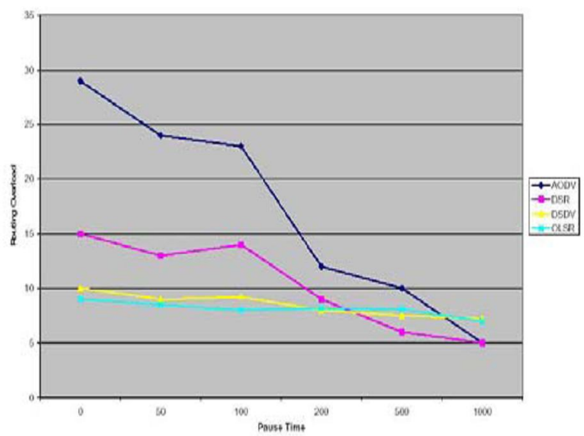


Fig. 3(a) Routing Overload v/s Pause Time (ms.)

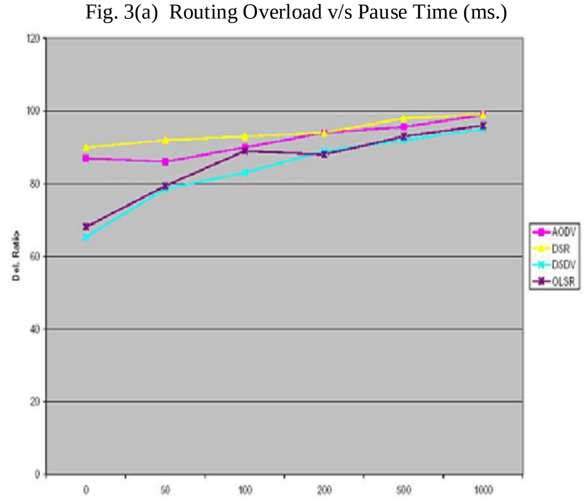


Fig. 3(b) Delivery Ratio v/s Pause Time (ms.)

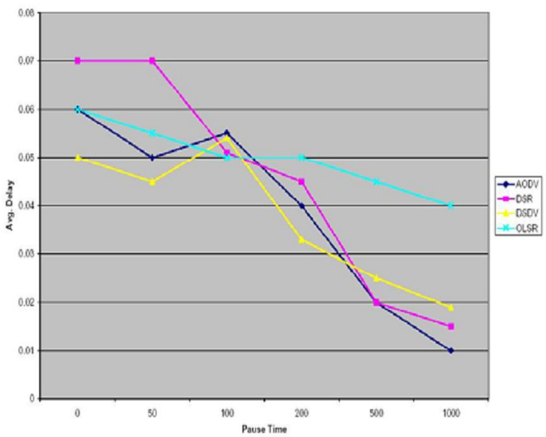


Fig. 3(c) Average Delay (ms) v/s Pause Time (ms)

Fig 3. Various Performance parameters versus Pause Time

5.2.2 Effect of Varying Number of Nodes

Number of nodes may be another varying parameter as it plays important role in performance. Figure 4 (4a,4b,4c) shows various performance parameters versus no. of nodes. From figure we can observe that routing overload for all protocol increased as no. of nodes increased but among them AODV performed poorer as this might be due to flooding of routing packets. We can observe that overhead for DSDV and OLSR also increased as increase in number of packet have increased the size of their routing table and also number of broadcast. While in case of less number of nodes all protocols performed poorer in terms of delivery ratio as nodes breakage may be more and no route may be available, again DSR outperformed all with respect to Delivery Ratio. In case of average delay, AODV was better than DSR but proactive protocols performed well due to their proactive nature.

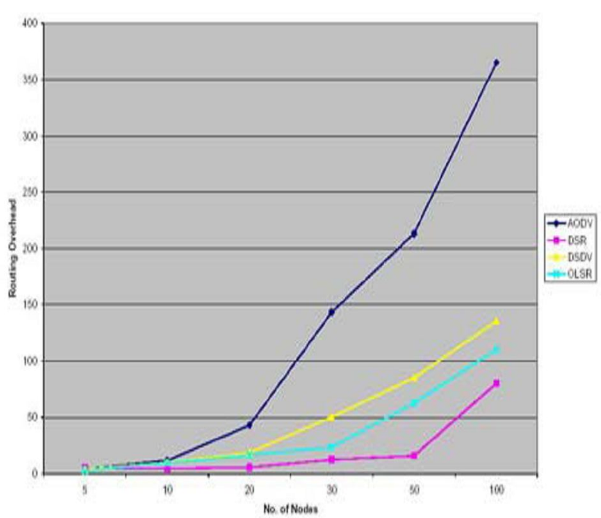


Fig. 4(a) Routing Overhead v/s No. of Nodes

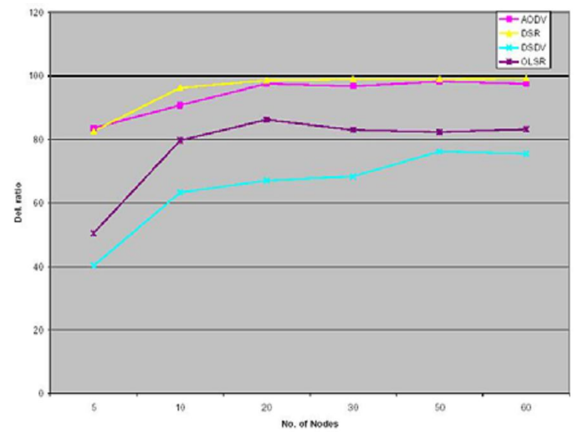


Fig. 4(b) Delivery ratio v/s No. of Nodes

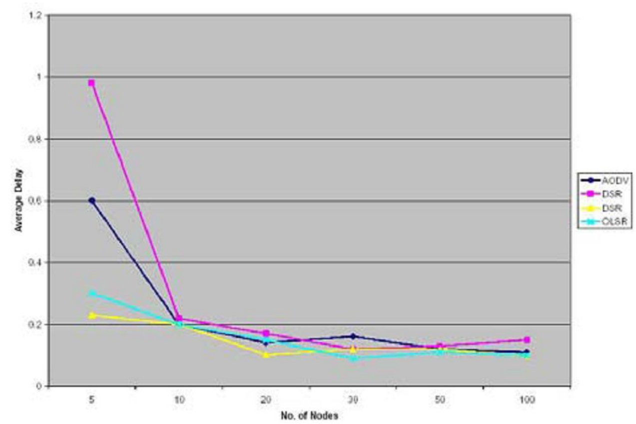


Fig. 4(c) Average Delay (ms) v/s No. of nodes
Figure 4. Various Performance Parameter V/s No. of Node

VI. CONCLUSION AND FUTURE SCOPE

Mobile Ad-Hoc Networks has the ability to deploy a network where a traditional network infrastructure environment cannot possibly be deployed. With the importance of MANET comparative to its vast potential it has still many challenges left in order to overcome. performance comparison of routing protocol in MANET is one of the important aspects. This paper presents a performance comparison of proactive and reactive routing protocols for mobile ad-hoc wireless networks. AODV and DSR are reactive protocol while DSDV and OLSR are proactive protocols. Both reactive protocols performed well in high mobility scenarios than proactive protocol. High mobility result in highly dynamic topology i.e. frequent route failures and changes. Both proactive protocols fail to respond fast enough to changing topology. Routing overhead in Proactive protocols remain almost constant and OLSR being winner irrespective of mobility while in AODV it increases with increase in mobility. Both AODV and DSR use reactive approach to route discovery, but with different mechanism. DSR uses source routing and route cache. On other hand AODV uses routing tables, one route per destination, sequence number to maintain route. The general observation from simulation is that DSR has performed well compared to all other protocols in terms of Delivery ratio while AODV outperformed in terms of Average delay. DSR however generates lower overhead than AODV while OLSR and DSDV generate almost constant overhead due proactive nature. Future work can be extended to various other types of protocols. The analysis of such protocols on the performance parameters like standard deviation, energy consumption can be also done in the future.

REFERENCES

[1] Wikipedia, *"The free encyclopedia-, Mobile ad-hoc Network"*, http://en.wikipedia.org/wiki/Mobile_ad-hoc_ network, *Oct*-2004.

[2] Charles E.Perkins and Elizabeth M. Royer, "*Ad hoc on demand distance vector (AODV) routing (Internet-Draft)*", Aug-1998.

[3] Network Simulator version 2 (NS-2) webpage Available at: http://www.isi.edu/nsnam/ns, accessed at May 1st, 2010.

[4] Mario Joa-Ng, "*A Peer-to-Peer Zone-Based Two-Level Link State Routing for Mobile Ad Hoc Networks*", IEEE Journal on selected areas in communications, Vol. 17, No. 8, Aug-1999.

[5] C. E. Perkins, E. Belding-Royer, and S. Das, "Ad Hoc On-demand Distance Vector (AODV) Routing," Experimental RFC 3561, Internet Engineering Task Force (IETF), July 2003.

[6] D. Johnson, Y. Hu, and D. Maltz, "The Dynamic Source Routing Protocol (DSR) for Mobile Ad Hoc Networks for IPv4," Experimental RFC 4728, Internet Engineering Task Force (IETF), February 2007.

[7] Padmini Misra, "*Routing Protocols for ad hoc mobile wireless Networks"*, http://www.cse.ohio-state.edu /~jain/cis788-99/ftp/adhoc_routing/#TDRP, Nov-1999.

[8] C. E. Perkins and P. Begat, "Highly Dynamic Destination- Sequenced Distance-Vector Routing (DSDV) for Mobile Computers," *Proceedings of the ACM SIGCOMM'94 Conference on Communications Architectures, Protocols, and Applications*, London, United Kingdom, August 1994, pp. 234-244.

[9]C.Perkins ,Praving Bhagwat, "Highly dynamic destination sequenced distance vector routing (DSDV) for Mobile computers".

[10] "Optimized Link State Routing Protocol", Internet draft, WWW.irtf.org.

[11] http://en.wikipedia.org/wiki/OLSR.

[12] J. Broch et al., "A Performance Comparison of Multihop Wireless Ad Hoc Network Routing Protocols", Proc. IEEE/ACM MOBICOM '98,Oct.1998, pp. 85–97.

[13] Jiwei Chen, Yeng-Zhong Lee, Daniela Maniezzo, Mario Gerla, "Performance Comparison of AODV and OFLSR in Wireless Mesh Networks",

[14] Yangcheng Huang, Saleem Bhatti, Søren-Aksel Sørensen, " Adaptive MANET Routing for Low Overhead",http://www.cs.standrews.ac.uk/~saleem/papers /2007/adamus2007/adamus2007-hbs2007.pdf.

[15] D. B. Johnson, D. A. Maltz, and J. Broch. DSR: "The Dynamic Source Routing Protocol for Multi - Hop Wireless Ad Hoc Networks". In Ad Hoc Networking, edited by Charles E. Perkins, Chapter 5, pp. 139-172, Addison-Wesley, 2001.

[16] Dr. Yogesh Chaba and Naresh Kumar Medishetti "Routing Protocols in mobile Ad - hoc Networks- A simualtion Study", Journal of Computer Science, JCS Vol.1 No.1,pp 83-88, August 2005.

Dataware House and its other aspects

Neha choudhary
M.Tech, CSE,Sri sai college. pathankot
nehachoudharydaviet@gmail.com

***Abstract*—Data warehousing is a commercially significant database technique that draws data from widely disparate sources to support business analytical needs. It requires architectural approaches, such as multidimensional data and summary views, that are totally unlike traditional transaction processing systems and it presents its own set of problems but it offers new analytical approaches based on multidimensionality and aggregation. This paper will concentrate on datawarehousing models, its advantages and disadvantages, design methodology of dataware house and data mart.**

***Keywords*— Dataware house, data mart,database,models.**

I. Introduction

A data warehouse is a relational database that is designed for query and analysis rather than for transaction processing.It usually contains historical data derived from transaction data,but it can include data from other sources. It separates analysis workload from transaction workload and enables an organization to consolidate data from several sources.

In addition to a relational database, a data warehouse environment includes an extraction, transportation, transformation, and loading (ETL) solution, an online analytical processing (OLAP) engine, client analysis tools, and other applications that manage the process of gathering data and delivering it to business users .Data warehousing is one of the hottest topics in the computing industry today. For business executives, it promises significant competitive advantage for their companies, while information systems managers see it as the way to overcome the traditional roadblocks to providing business information for managers and other end users.

In computing, a data warehouse (DW) is a database used for reporting and analysis. The data stored in the warehouse is uploaded from the operational systems. The data may pass through an operational data store for additional operations before it is used in the DW for reporting. A data warehouse maintains its functions in three layers: staging, integration, and access. Staging is used to store raw data for use by developers. The integration layer is used to integrate data and to have a level of abstraction from users. The access layer is for getting data out for users. Data warehouses can be subdivided into data marts. Data marts store subsets of data from a warehouse. This definition of the data warehouse focuses on data storage. The main source of the data is cleaned, transformed, catalogued and made available for use by managers and other business professionals for data mining, online analytical processing, market research and decision support. However, the means to retrieve and analyse data,to extract, transform and load data, and to manage the data dictionary are also considered essential components of a data warehousing system. Many references to data warehousing use this broader context. Thus, an expanded definition for data warehousing includes business intelligence tools, tools to extract, transform and load data into the repository, and tools to manage and retrieve metadata.

Besides being a storehouse for a large amount of data, they must possess systems in place that make it easy to access the data and use it in day to day operations. A data warehouse is sometimes said to be a major role player in a decision support system (DSS). DSS is a technique that organizations use to come up with facts, trends, or relationships that can help them make effective decisions or create effective strategies to accomplish their organizational goals.

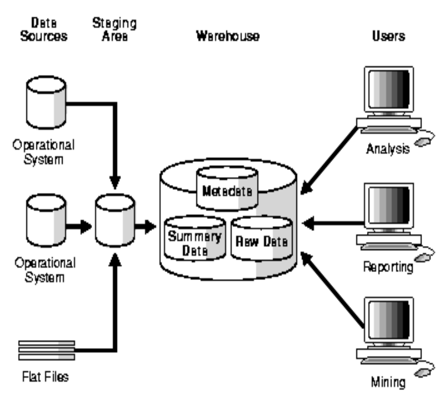


Fig. 1 Architecture of dataware house

II. Data Warehousing Models

There are many different data warehouse models. Online Transaction Processing, which is a data warehouse model, is built for speed and easy use. Another type of data warehouse model is called Online Analytical processing, which is more difficult to use and adds an extra step of analysis within the data. Usually it requires more steps that slows the process down and requires much more data in order to analyse certain queries.

In addition to this model, one of the more common data warehouse models include a data warehouse that is subject oriented, time variant, not volatile, and integrated. Subject oriented means that data is linked together and is organized by relationships. Time variant means that any data that is changed in the data warehouse can be tracked. Usually all changes to data are stamped with a time-date and with a before and after value, so that the changes throughout a period of time can be shown. Non volatile means that the data is never deleted or erased. This is a great way to protect the most crucial data. Because this data is retained, users can continue to use it in a later analysis. Finally, the data is integrated, which means that a data warehouse uses data that is organizational wide instead of from just one department. Besides the term data warehouse, a frequently used term is data mart. Data marts are smaller and less integrated than data housings. They might be just a database on human resources records or sales data on just one division.

III. BENEFITS AND LIMITATIONS OF DATAWARE HOUSE

Followings are the benefits and limitations of dataware house:

A. Benefits

A data warehouse maintains a copy of information from the source transaction systems. This architectural complexity provides the opportunity to:

1) Maintain data history, even if the source transaction systems do not.

2) Integrate data from multiple source systems, enabling a central view across the enterprise. This benefit is always valuable, but particularly so when the organization has grown by merger.

3) Improve data quality, by providing consistent codes and descriptions, flagging or even fixing bad data.

4) Present the organization's information consistently.

5) Provide a single common data model for all data of interest regardless of the data's source.

6) Restructure the data so that it makes sense to the business users.

7) Restructure the data so that it delivers excellent query performance, even for complex analytic queries, without impacting the operational systems.

8) Add value to operational business applications, notably customer relationship management (CRM) systems.

B. Limitations

1) The major limitations associated with data warehousing are related to user expectations, lack of data and poor data quality.

2) Building a data warehouse creates some unrealistic expectations that need to be managed.

3) A data warehouse doesn't meet all decision support needs. If needed data is not currently collected, transaction.

4) Long initial implementation time and associated high cost.

5) Limited flexibility of use and types of users - requires multiple separate data marts for multiple uses and types of users.

6) Difficult to accommodate changes in data types and ranges, data source schema, indexes and queries.

IV. BOTTOM-UP AND TOP-DOWN DESIGN METHODOLOGY

A. Bottom-up design

Ralph Kimball, a well-known author on data warehousing,is a proponent of an approach to data warehouse design which he describes as bottom-up. In the bottom-up approach data marts are first created to provide reporting and analytical capabilities for specific business processes. Though it is important to note that in Kimball methodology, the bottom-up process is the result of an initial business oriented Top-down analysis of the relevant business processes to be modelled. Data marts contain, primarily, dimensions and facts. Facts can contain either atomic data and, if necessary, summarized data. The single data mart often models a specific business area such as "Sales" or "Production." These data marts can eventually be integrated to create a comprehensive data warehouse. The integration of data marts is managed through the implementation of what Kimball calls "a data warehouse bus architecture. The data warehouse bus architecture is primarily an implementation of "the bus", a collection of conformed dimensions and conformed facts, which are dimensions that are shared (in a specific way) between facts in two or more data marts.

The integration of the data marts in the data warehouse is centred on the conformed dimensions (residing in "the bus") that define the possible integration "points" between data marts. The actual integration of two or more data marts is then done by a process known as "Drill across". A drill-across works by grouping (summarizing) the data along the keys of the (shared) conformed dimensions of each fact participating in the "drill across" followed by a join on the keys of these grouped (summarized) facts. Maintaining tight management over the data warehouse bus architecture is fundamental to maintaining the integrity of the data warehouse. The most important management task is making sure dimensions among data marts are consistent. In Kimball's words, this means that the dimensions "conform".

Some consider it an advantage of the Kimball method, that the data warehouse ends up being "segmented" into a number of logically self contained (up to and including The Bus) and consistent data marts, rather than a big and often complex centralized model. Business value can be returned as quickly as the first data marts can be created, and the method gives itself well to an exploratory and iterative approach to building data warehouses. For example, the data warehousing effort might start in the "Sales" department, by building a Sales-data mart. Upon completion of the Sales-data mart, The business might then decide to expand the warehousing activities into the, say, "Production department" resulting in a Production data mart. The requirement for the Sales data mart and the Production data mart to be integrated, is that they share the same "Bus", that will be, that the data warehousing team has made the effort to identify and implement the conformed dimensions in the bus, and that the individual data marts links that information from the bus. Note that this does not require 100% awareness from the onset of the data warehousing effort, no master plan is required upfront. The Sales-data mart is good as it is (assuming that the bus is complete) and the production data mart can be constructed virtually independent of the sales data mart (but not independent of the Bus).

If integration via the bus is achieved, the data warehouse, through its two data marts, will not only be able to deliver the specific information that the individual data marts are designed to do, in this example either "Sales" or "Production" information, but can deliver integrated Sales-Production information, which, often, is of critical business value. An integration (possibly) achieved in a flexible and iterative fashion.

C. Top-down design

Bill Inmon, one of the first authors on the subject of data warehousing, has defined a data warehouse as a centralized repository for the entire enterprise.Inmon is one of the leading proponents of the top-down approach to data warehouse design, in which the data warehouse is designed using a normalized enterprise data model. "Atomic" data, that is, data at the lowest level of detail, are stored in the data warehouse. Dimensional data marts containing data needed for specific business processes or specific departments are created from the data warehouse. In the Inmon vision the data warehouse is at the centre of the "Corporate Information Factory" (CIF), which provides a logical framework for delivering business intelligence (BI) and business management capabilities.Inmon states that the data warehouse is:

1) Subject-oriented.
2) Non-volatile.
3) Integrated.
4) Time-variant.

.

The top-down design methodology generates highly consistent dimensional views of data across data marts since all data marts are loaded from the centralized repository. Top-down design has also proven to be robust against business changes. Generating new dimensional data marts against the data stored in the data warehouse is a relatively simple task. The main disadvantage to the top-down methodology is that it represents a very large project with a very broad scope. The up-front cost for implementing a data warehouse using the top-down methodology is significant, and the duration of time from the start of project to the point that end users experience initial benefits can be substantial. In addition, the top-down methodology can be inflexible and unresponsive to changing departmental needs during the implementation phases.

V. DATA WAREHOUSE VERSUS OPERATIONAL SYSTEM

Operational systems are optimized for preservation of data integrity and speed of recording of business transactions through use of database normalization and an entity-relationship model. Operational system designers generally follow the Codd rules of database normalization in order to ensure data integrity. Codd defined five increasingly stringent rules of normalization. Fully normalized database designs (that is, those satisfying all five Codd rules) often result in information from a business transaction being stored in dozens to hundreds of tables. Relational databases are efficient at managing the relationships between these tables. The databases have very fast insert/update performance because only a small amount of data in those tables is affected each time a transaction is processed. Finally, in order to improve performance, older data are usually periodically purged from operational systems.

Data warehouses are optimized for speed of data analysis. Frequently data in data warehouses are denormalized via a dimension-based model. Also, to speed data retrieval, data warehouse data are often stored multiple times—in their most granular form and in summarized forms called aggregates. Data warehouse data are gathered from the operational systems and held in the data warehouse even after the data has been purged from the operational systems.

VI. ADVANTAGES AND DISADVANTAGES OF DATA WAREHOUSE

The number one reason for implementing a data warehouse is so that employees or end users can access the data warehouse and use the data for reports, analysis, and decision making. Using the data in a warehouse can help locate trends, focus on relationships, and help users understand more about the environment that a business operates in. Data warehouses also increase the data's consistency and allow it to be checked repeatedly to determine how relevant it is. Because most data warehouses are integrated, users can pull data from many different areas of their business, for instance human resources, finance, IT, accounting, etc.

While there are plenty of reasons to have a data warehouse, it should be noted that there are a few negatives associated with this. This includes the fact that it is time consuming to create and to keep operating.

Users might also have a problem with current systems being incompatible with the data. It is also important to consider future equipment and software upgrades. These may also need to be compatible with the data.

Finally, security might be a huge concern, especially if the data is accessible over an open network such as the Internet. Users will not want their competitor to view, hack, or destroy their data.

VII. DATA MART

A data mart is the access layer of the data warehouse environment that is used to get data out to the users. The data mart is a subset of the data warehouse which is usually oriented to a specific business line or team. There can be multiple data marts inside a single corporation; each one relevant to one or more business units for which it was designed. Data marts may or may not be dependent or related to other data marts in a single corporation. If the data marts are designed using conformed facts and dimensions, then they will be related. In some deployments, each department or business unit is considered the owner of its data mart including all the hardware, software and data. This enables each department to use, manipulate and develop their data any way they see fit; without altering information inside other data marts or the data warehouse. In other deployments where conformed dimensions are used, this business unit ownership will not hold true for shared dimensions like customer, product, etc.

The related term describes the situation that occurs when one or more business analysts develop a system of linked spreadsheets to perform a business analysis, then grow it to a size and degree of complexity that makes it nearly impossible to maintain. data mart is a simple form of a data warehouse that is focused on a single subject (or functional area), such as Sales, Finance, or Marketing. Data marts are often built and controlled by a single department within an organization. Given their single-subject focus, data marts usually draw data from only a few sources. The sources could be internal operational systems, a central data warehouse, or external data A data warehouse, unlike a data mart, deals with multiple subject areas and is typically implemented and controlled by a central organizational unit such as the corporate Information Technology (IT) group. Often, it is called a central or enterprise data warehouse. Typically, a data warehouse assembles data from multiple source systems. Nothing in these basic definitions limits the size of a data mart or the complexity of the decision-support data that it contains. Nevertheless, data marts are typically smaller and less complex than data warehouses; hence, they are typically easier to build and maintain.

VIII. DEPENDENT AND INDEPENDENT DATA MARTS

There are two basic types of data marts: dependent and independent. The categorization is based primarily on the data source that feeds the data mart. Dependent data marts draw data from a central data warehouse that has already been created. Independent data marts, in contrast, are standalone systems built by drawing data directly from operational or external sources of data, or both. The main difference between independent and dependent data marts is how you populate the data mart; that is, how you get data out of the sources and into the data mart. This step, called the Extraction-Transformation-and Loading (ETL) process, involves moving data from operational systems, filtering it, and loading it into the data mart. With dependent data marts, this process is somewhat simplified because formatted and summarized (clean) data has already been loaded into the central data warehouse. The ETL process for dependent data marts is mostly a process of identifying the right subset of data relevant to the chosen data mart subject and moving a copy of it, perhaps in a summarized form.

With independent data marts, however, you must deal with all aspects of the ETL process, much as you do with a central data warehouse. The number of sources is likely to be fewer and the amount of data associated with the data mart is less than the warehouse, given your focus on a single subject. The motivations behind the creation of these two types of data marts are also typically different. Dependent data marts are usually built to achieve improved performance and availability, better control, and lower telecommunication costs resulting from local access of data relevant to a specific department. The creation of independent data marts is often driven by the need to have a solution within a shorter time.

IX. STEPS IN IMPLEMENTING A DATAMART

Simply stated, the major steps in implementing a data mart are to design the schema, construct the physical storage, populate the data mart with data from source systems, access it to make informed decisions, and manage it over time:

A. Designing

The design step is first in the data mart process. This step covers all of the tasks from initiating the request for a data mart through gathering information about the requirements, and developing the logical and physical design of the data mart. The design step involves the following tasks:

1) Gathering the business and technical requirements.
2) Identifying data sources.
3) Selecting the appropriate subset of data.
4) Designing the logical and physical structure of the data mart.

B. Constructing

This step includes creating the physical database and the logical structures associated with the data mart to provide fast and efficient access to the data. This step involves the following tasks:

1) Creating the physical database and storage structures, such as table spaces, associated with the data mart.

2) Creating the schema objects, such as tables and indexes defined in the design step.

3) Determining how best to set up the tables and the access structures.

C. Populating

The populating step covers all of the tasks related to getting the data from the source, cleaning it up, modifying it to the right format and level of detail, and moving it into the data mart. More formally stated, the populating step involves the following tasks:

1) Mapping data sources to target data structures.

2) Extracting data

3) Cleansing and transforming the data.

4) Loading data into the data mart.

5) Creating and storing metadata.

D. Accessing

The accessing step involves putting the data to use: querying the data, creating reports, charts, and graphs, and publishing these. Typically, the end user uses a graphical front-end tool to submit queries to the database and display the results of the queries. The accessing step requires that you perform the following tasks:

1) Set up an intermediate layer for the front-end tool to use. This layer, the meta layer, translates database structures and object names into business terms, so that the end user can interact with the data mart using terms that relate to the business function.

2) Maintain and manage these business interfaces.

3) Set up and manage database structures, like summarized tables, that help queries submitted through the front-end tool execute quickly and efficiently.

E. Managing

This step involves managing the data mart over its lifetime. In this step, you perform management tasks such as the following:

1) Providing secure access to the data.

2) Managing the growth of the data.

3) Optimizing the system for better performance.

4) Ensuring the availability of data even with system failures.

CONCLUSION

Data warehousing provides unique analytical benefits to the enterprise. However, there appears to be more practical experience than theoretical. data warehousing includes business intelligence tools, tools to extract, transform and load data into the repository, and tools to manage and retrieve metadata .Data in the data warehouse is subject oriented, time variant, non volatile and integrated.Data warehouse follow top down and bottom up design methodology. Operational systems follow normalization technique whereas data warehouse follow denormalization technique.

The data mart is a subset of the data warehouse and it is the access layer of the data warehouse environment that is used to get data out to the users. Data mart is a simple form of a data warehouse that is focused on a single subject (or functional area), such as sales, finance, or marketing. Data marts are typically smaller and less complex than data warehouses. Hence, they are typically easier to build and maintain.

REFERENCES

[1]http://docs.oracle.com/cd/B10501_01/server.920/a96520/concept.html

[2] Data Warehouse: From Architecture to Implementation By Barry Devlin

[3] http://en.wikipedia.org/wiki/Data_warehouse

[4] Inmon, Bill (1992). Building the Data Warehouse. Wiley

[5] Hillard, Robert (2010). Information-Driven Business. Wiley

[6] http://www.tech-faq.com/data-warehouse.html

[7] http://en.wikipedia.org/wiki/Data_mart

[8]http://docs.oracle.com/javase/1.4.2/docs/guide/jdbc/getstart/datasource.html

[9] http://www.rpi.edu/datawarehouse/dw-about-cmp.html

[10]http://www.nasi.com/NGS-IQ_datawarehouse-manager.php

[11] http://en.wikipedia.org/wiki/Operational_data_store

[12] Golfarelli/Rizzi, Data Warehouse Design, p 291

[13] http://varietysoftworks.com/jbaldwin/Education/CS615-DataWarehouse.html

[14] The Data Warehouse Toolkit by Ralph Kimball (John Wiley and Sons, 1996)

[15] Building the Data Warehouse by William Inmon (John Wiley and Sons, 1996)

[16] http://www.whamtech.com/adv_disadv_dw.html

[17]http://dssresources.com/faq/index.php?action=artikel&id=47

[18]Kimball, The Data Warehouse Lifecycle Toolkit, p 332

"WIMAX" The Future of Wireless

Pawanpreet kaur[1], Shaweta Mahajan[2], Narinder Sharma[3] &Junesh Kumar[4]
[1]Lecturer in Lord Krishna Polytechnic College, Subhanpura.
[2]Lecturer in Govt Polytechnic College, Bhikhiwind.
[3]Assistant Professor in ACET, Amritsar.
[4]Lecturer in Lord Krishna Polytechnic College, Subhanpura.
Email: Pawan.preet11@yahoo.co.in, mahajan.shaweta@yahoo.in, narinder.acet@gmail.com & junesh_kumar@hotmail.com

***ABSTRACT:-*WiFi have emerged as promising broadband access solutions for the latest generation of wireless LANs, The new era of communication, currently employed in some parts of the world, is Worldwide Interoperability for Microwave Access (WIMAX). It is the latest technology which is approved by IEEE 802.16 group, which is a standard for point-to-multipoint wireless networking. Wimax vision is to deliver "last mile" broadband connectivity to home or business locations, also its data rates are comparable with Cable and Digital Subscriber line (DSL) rates. It has the capability which connects to the ISP (Internet Service Provider) even when you are roaming outside home or office. The Wimax technology is becoming the way to avert the impending crisis of rural connectivity i.e. it will be accessible till the last mile. This paper explains about the purpose of Wimax, over the limitation of WIFI**

Keywords:- WIFI, Wimax, DSL, ISP LOS, NLOS, OFDM.

I. INTRODUCTION

Wireless access techniques are continuously expanding their transmission bandwidth, coverage, and Quality of Service (QoS) support in recent years. With the huge market success of Wireless Local Area Networks (WLANs) (governed by the IEEE 802.11 standard),the new-generation Wireless technique, WiMAX (i.e., IEEE 802.16), has now been standardized and deployed. Traditionally, WiFi hotspots are connected to the Internet through a wired connection (e.g., Ethernet), and therefore have high deployment costs, particularly in remote rural or suburban areas with low population densities. Therefore, it is necessary to develop new schemes capable of providing sufficient bandwidth to meet the enormous access requirements of WiFi nodes while simultaneously reducing the backhaul cost. It has been suggested that the evolving family of WiMAX-based Wireless Metropolitan Area Network (WMAN) technologies represent a promising solution for providing WLAN hotspots with backhaul support. Using a WiMAX-based backbone network to connect WiFi hotspots to the Internet not only avoids the requirement for a costly wired infrastructure, but also makes possible the provision of mobile hotspot services to realize Intelligent Transportation System (ITS) applications. 802.16 network a QoS control protocol was presented to support an integrated QoS for converged networks comprising WiMAX and WiFi systems. However, in the mechanisms required to satisfy the QoS requirements (e.g., bandwidth assignment, scheduling, admission control, and so forth) were not considered, while implementing the proposed QoS provisioning mechanism required a major rework of the WiMAX and WiFi control protocols. Accordingly, the current study proposes an efficient and unified connection-oriented architecture for integrating WiMAX and WiFi technologies in broadband wireless networks. In the proposed approach, a new wireless Access Point (AP) device, designated as WiMAX/WiFi AP (W2-AP), is developed to manage the WiMAX/WiFi interface.

II. WIMAX BACKGROUND

The IEEE 802.16 group was formed in 1998 to develop an air-interface standard for wireless broadband. The group's initial focus was the development of a LOS-based point-to-multipoint wireless broadband system for operation in the 10GHz–66GHz millimeter wave band. The resulting standard—the original 802.16 standard, completed in December 2001—was based on a single-carrier physical (PHY) layer with a burst time division multiplexed (TDM) MAC layer. Many of the concepts related to the MAC layer were adapted for wireless from the popular cable modem DOCSIS (data over cable service interface specification) standard The IEEE 802.16 group subsequently produced 802.16a, an amendment to the standard, to include NLOS applications in the 2GHz–11GHz band, using an orthogonal frequency division multiplexing (OFDM)-based physical layer. Additions to the MAC layer, such as support for orthogonal frequency division multiple access (OFDMA), were also included. Further revisions resulted in a new standard in 2004, called IEEE 802.16-2004, which replaced all prior versions and formed the basis for the first WiMAX solution. These early WiMAX solutions based on IEEE 802.16-2004 targeted fixed applications, and we will refer to these as fixed WiMAX . In December 2005, the IEEE group completed and approved IFEEE 802.16e-2005, an amendmentto the IEEE 802.16-2004 standard that added mobility support. The IEEE 802.16e-2005 forms the basis for the WiMAX solution for nomadic and mobile applications and is often referred to as mobile WiMAX The basic characteristics of the various IEEE 802.16 standards are summarized in table. Note that these standards offer a variety of fundamentally different design options. For example, there are multiple physical-layer choices: a single-carrier-based physical layer called WirelessMANSCa, an OFDM-based physical layer called WirelessMAN-OFDM, and an OFDMA- based physical layer called Wireless-OFDMA. Similarly, there are multiple choices for MAC architecture, duplexing, frequency band of operation, etc. These standards were E developed to suit a variety of applications and deployment scenarios, and hence offer a plethora of design choices for system developers. In fact, one could say that IEEE 802.16 is a collection of standards, not one single interoperable standard

Table1: Comparison of WIMAX Standards

	802.16	*802.16-2004*	*802.16e-2005*
Status	Completed December 2001	Completed June 2004	Completed December 2005
Frequency band	10GHz–66GHz	2GHz–11GHz	2GHz–11GHz for fixed; 2GHz–6GHz for mobile applications
Application mobile	Fixed LOS	Fixed NLOS	Fixed and NLOSApplication
MAC architecture	Point-tomultipoint, mesh	Point-to-multipoint, mesh	Point-to-multipoint, mesh
Transmission scheme	Single carrier only	Single carrier, 256 OFDM or 2,048 OFDM	Single carrier, 256 OFDM or scalable OFDM with 128, 512, 1,024, or 2,048 subcarriers
Modulation	QPSK, 16 QAM, 64 QAM	QPSK, 16 QAM, 64 QAM	QPSK, 16 QAM, 64 QAM
Gross data rate	32Mbps – 134.4Mbps	1Mbps–75Mbps	1Mbps–75Mbps
Multiplexing	Burst TDM/TDMA	Burst TDM/TDMA/ OFDMA	Burst TDM/TDMA/ OFDMA
Duplexing	TDD and FDD	TDD and FDD	TDD and FDD
Channel bandwidths	20MHz, 25MHz, 28MHz	1.75MHz, 3.5MHz, 7MHz 14MHz, 1.25MHz, 5MHz, 10MHz, 15MHz, 8.75MHz	1.75MHz, 3.5MHz, 7MHz, 14MHz, 1.25MHz, 5MHz, 10MHz, 15MHz, 8.75MHz
Air-interface designation	WirelessMANSC	WirelessMAN-SCa WirelessMAN-OFDM WirelessMAN-OFDMA WirelessHUMAN	WirelessMAN-SCa WirelessMAN-OFDM WirelessMAN-OFDMA WirelessHUMAN
WiMAX implementation	None	256 - OFDM as Fixed WiMAX	Scalable OFDMA as Mobile WiMAX

III. WIRELESS ACCESS

Wireless access networks are basically divided into three categories according to the ITU (International Telecommunication Union).

- *Fixed wireless access (FWA)* Wireless access application in which the location of the end-user termination and the network access point to be connected to the end-user are fixed.
- *Mobile wireless access (MWA)* Wireless access application in which the location of the end-user termination is mobile.
- *Nomadic wireless access (NWA)* Wireless access application in which the location of the end-user termination may be in different places but it must be stationary while in use.

Mobile WiMAX is based on the IEEE 802.16d standard (a.k.a. fixed WiMAX), but is designed to maintain connectivity on the go, tracking a receiver at speeds of up to 37 miles per hour (60 km/h). IEEE 802.16d is a fixed wireless technology, offering wireless Internet connectivity to fixed users at ranges of up to 31 miles (50 km) from the transmitting base. IEEE 802.16d, and the previous releases, are not designed to be used while the receiver is in motion.There are several releases of WiMAX of which the most current release is referred to as Release-e. Mobile WiMAX, based on IEEE 802.16e-2005, will initially operate in the 2.3 GHz, 2.5 GHz, 3.3 GHz and 3.4-3.8 GHz spectrum bands providing data rates of up to 46 Mbps/sector in the DL and 7 Mbps/sector in the UL. Support for additional bands will be added on the basis of market demand and new spectrum allocations (i.e., 5.1 and 5.8

GHz). The technology and exact specifications of Mobile WiMAX will change as it undergoes refinements throughout its preliminary stages. All mobile WiMAX products will support handoffs and power-saving mechanisms.Fixed WiMax, IEEE 802.16d, has been commercially deployed by several carriers throughout the world and the preferred frequency band of operation has been 2.3 GHz. Recently, different bands of operation have also been explored, such as the 700 MHz band, the unlicensed bands at 5.1/5.8 GHz as well as some ATS bands (i.e., the L-Band and 2.1 GHz). Sprint/Nextel will be using the previous MMDS/ITFS band to deploy a nationwide mobile WiMAX network for its 4G initiative.

IV. TECHNOLOGICAL FEATURES

- Multiple handoff mechanisms, ranging from hard handoffs (with break-before make links) to soft handoffs (with make-before-break links).

- Power-saving mechanisms for mobile devices.

- Advanced QoS and low latency for improved support of real-time applications.

- Advanced Authorization, Authentication, and Accounting (AAA) functionality.

- Orthogonal Frequency Division Multiple Access (OFDMA), a multiplexing technique that is being proposed for 4G systems due to its robustness to multi path environments and its flexibility in managing spectrum resources, and improved indoor coverage. The Third Generation Partnership Project (3GPP) has incorporated OFDMA in its LTE (Long Term Evolution) specification and the Third Generation Partnership Project Two (3GPP2) is moving in the same direction.

- Time Division Duplex (TDD) and Frequency Division Duplex (FDD), which dominates in 3G networks. Whereas FDD keeps the uplink and the downlink channels separate in frequency, TDD is a less complex, more efficient mechanism that uses a single frequency channel, with uplink and downlink traffic separated by a guard time. In addition, for IP-based services the use of a single channel for the uplink and the downlink makes it substantially less complex and more cost-effective to implement MIMO and beam-forming in WiMAX networks than in CDMA-based networks. MIMO and beam-forming are expected to bring a substantial improvement in throughput in TDD-based WiMAX networks.

- Adaptive Base band Modulation Schemes are used to increase spectral efficiency: QPSK, 8-PSK and M-ary QAM (16QAM and 64 QAM).

- High Data Rates: the inclusion of MIMO antenna techniques along with flexible sub-channelization schemes, advance coding and modulation will enable the support of peak DL data rates of up to 63 Mbps/sector and peak UL data rates of 28 Mbps/sector in a 10 MHz channel.

- WiMAX has a scalable physical-layer architecture that allows for the data rate to scale easily with available channel bandwidth. This scalability is supported in the OFDMA mode, where the FFT (fast fourier transform) size may be scaled based on the available channel bandwidth. For example, a WiMAX system may use 128-512-, or 1,048-bit FFTs based on whether the channel bandwidth is 1.25MHz, 5MHz, or 10MHz, respectively. This scaling may be done dynamically to support user roaming across different networks that may have different bandwidth allocations.

- WiMAX supports a number of modulation and forward error correction (FEC) coding schemes and allows the scheme to be changed on a per user and per frame basis, based on channel conditions. AMC is an effective mechanism tomaximize throughput in a time-varying channel. The adaptation algorithm typically calls for the use of the highest modulation and coding scheme that can be supported by the signal-to-noise and interference ratio at the receiver such that each user is provided with the highest possible data rate that can be supported in their respective links.

- For connections that require enhanced reliability, WiMAX supports automatic retransmission requests (ARQ) at the link layer. ARQ-enabled connections require each transmitted packet to be acknowledged by the receiver; unacknowledged packets areassumed to be lost and are retransmitted. WiMAX also optionally supports hybrid-ARQ, which is an effective hybrid between FEC and ARQ.

- The WiMAX solution has a number of hooks built into the physical-layer design, which allows for the use of multiple-antenna techniques, such as beamforming, space-time coding, and spatial multiplexing. These schemes can be used to improve the overall system capacity and spectral efficiency by deploying multiple antennas at the transmitter and/or the receiver. Chapter 5 presents detailed overview of the various multiple antenna techniques.

- WiMAX supports strong encryption, using Advanced Encryption Standard (AES), and has a robust privacy and key-management protocol. The system also offers a very flexible authentication architecture based on Extensible Authentication Protocol (EAP), which allows for a variety of user credentials, including username/password, digital certificates, and smart cards.

V. WIMAX ARCITECTURE

The IEEE 802.16e-2005 standard provides the air interface for WiMAX but does not define the full end-to-end WiMAX network. The WiMAX Forum's Network Working Group, is responsible for developing the end-to-end network requirements, architecture, and protocols for WiMAX, using IEEE 802.16e-2005 as the air interface.The WiMAX NWG has developed a network reference model to serve as an architectureframework for WiMAX deployments and to ensure interoperability among various WiMAX equipment and operators. The network reference model envisions a unified network architecture for supporting fixed, nomadic, and mobile deployments and is based on an IP service model. Figure below shows a simplified illustration of an IP-based WiMAX network architecture. The overall network may be logically divided into three parts: (1) mobile stations used by the end user to access the network, (2) the access service network (ASN), which comprises one or more base stations and one or more ASN gateways that form the radio access network at the edge, and (3) the connectivity service network (CSN), which provides IP connectivity and all the IP core network functions. The network reference model developed by the WiMAX Forum NWG defines a number of functional entities and interfaces between those entities. (The interfaces are referred to as reference points.) Figure1 shows some of the more important functional entities

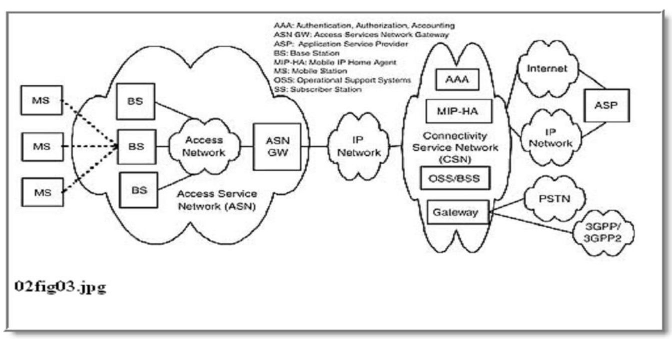


Fig 1: IP-Based WiMAX Network Architecture

BASE STATION (BS): The BS is responsible for providing the air interface to the MS. Additiona functions that may be part of the BS are micromobility management functions, such as handoff triggering and tunnel establishment, radio resource management, QoS policy enforcement, traffic classification, DHCP (Dynamic Host Control Protocol) proxy, key management, session management, and multicast group management

ACCESS SERVICE NETWORK GATEWAY (ASN-GW): The ASN gateway typically acts as a layer 2 traffic aggregation point within an ASN. Additional functions that may be part of the ASN gateway include intra-ASN location management and paging, radio resource management and admission control, caching of subscriber profiles and encryption keys, AAA client functionality, establishment and management of mobility tunnel with base stations, QoS and policy enforcement, foreign agent functionality for mobile IP, and routing to the selected CSN.

CONNECTIVITY SERVICE NETWORK (CSN): The CSN provides connectivity to the Internet, ASP, other public networks, and corporate networks. The CSN is owned by the NSP and includes AAA servers that support authentication for the devices, users, and specific services. The CSN also provides per user policy management of QoS and security. The CSN is also responsible for IP address management, support for roaming between different NSPs, location management between ASNs, and mobility and roaming between ASNs. Further, CSN can also provide gateways and interworking with other networks, such as PSTN (public switched telephone network), 3GPP, and 3GPP2.

VI. FLEXIBILITY

Some of the key features that make WiMAX a flexible technology can be summarized as follows:

- *Network Deployment*: WiMAX can be deployed both in green-field deployments, where network operators rely exclusively on WiMAX for the edge infrastructure, and in overlay or complementary networks, where operators embed WiMAX within their networks to increase capacity and throughput as necessary to deliver true wireless broadband service.

- *Global roaming*: Among WiMAX service providers will allow subscribers to access different networks using the same device and a single, familiar interface.

- *Spectrum*: Mobile WiMAX can be deployed in several licensed bands (2.3 GHz, 2.5 GHz, 3.3 GHz, 3.4-3.8 GHz) with channel sizes ranging from 3.5 MHz to 10MHz, unlicensed bands (i.e.: 5.3/5.8 GHz)and partially licensed bands (3.65 GHz).

- *ORTHOGONAL FREQUENCY DIVISION MULTIPLEXING (OFDM*) Mobile WiMAX uses OFDM as a multiple-access technique, whereby different users can be allocated different subsets of the OFDM tones. OFDMA facilitates the exploitation of frequency diversity and multiuser diversity to significantly improve the system capacity. All profiles currently defined by the WiMax Forum specify the 256-carrier OFDM air interface.

- It allows digital signal to be transmitted simultaneously on multiple F carrier waves. Adaptable to NLOS schemes.

- Resistant to multi-path effects.
- Spectrally efficient technique to transmit wireless digital data.
- Able to deliver higher bandwidth efficiency.

There are some obstacles in using OFDM in transmission system in contrast to its advantages. A major obstacle is that the OFDM signal exhibits a very high Peak to Average Power Ratio (PAPR).

VII. WIMAX MOBILITY ISSUES

- Device availability is a major issue
- In some markets spectrum availability is limited
  a) Bands < 3 GHz is better suited for mobile access
  b) Licenses for fixed WiMAX may not allow service provider to offer mobile services
- Current demand for WiMax is mostly for fixed services.
  a) Underserved Regions, Developing Market.
- Demand for wireless data is growing, but still it is limited
  a) Mobile operators may see need for a data-only technology when demand is higher
  b) Demand may drive additional spectrum allocations for wireless mobile data service
- WiMax is not going to supplant other wireless technologies
  a) It will not replace Wi-Fi in the LAN
  b) Cellular technologies may still be needed for voice and data in the WAN

CONCLUSIONS

- It is expected that WiMax becomes the dominant standard for Wireless MAN in the world market, at least, in fixed broadband networks.
- WiMax products will have to be delivered to the market needs and those for the end-users will have to be extremely easy to install.
- Focus is too often on technologies.
  a) Subscribers pay for services, not technologies.
  b) Technologies enable services, but should not be a burden on users.
  c) Broadband capabilities are important, but bandwidth is not the only meter to assess service

REFERENCES

[1] Introduction to Wi-Fi technology, Retrieved on September 24, 2006 from www.wi-fitechnology.com

[2] Broadcom, 802.11n: Next-Generation Wireless LAN Technology,White Paper dated April 2006

[3] IEEE Std 802.16-2004, "IEEE Standard for Local and Metropolitan Area Networks – Part 16: Air Interface for Fixed Broadband Wireless AccessSystems", Oct. 2004.

[4] C. Eklund, et al, "IEEE Standard 802.16: A Technical Overview of the Wireless MAN Air Interface for Broadband Wireless Access", *IEEE Commun*. *Mag*., vol. 40, no. 6, pp. 98-107, June 2002

[5] J. Chen; W. Jiao and Q. Guo, (2005). An integrated QoS control architecture for IEEE 802.16 broadband wireless access systems. In Global Telecommunications Conference, 2005 (IEEE GLOBECOM), Vol. 6. (pp.3335-3340), St. Louis, MO.

[6] X. Hei, Y. Liu and K. Ross, "IPTV over P2P Streaming Networks: the Mesh-pull Approach," IEEE Communications Magazine, vol. 46, no. 2, pp. 86-92, Feb. 2008

Fuzzy Logic Approach used for handover scheme

[1]Ruhani Jain <ruhani_jain@yahoo.com>
[2]Amarpreet Singh amarmandeep@yahoo.com >

***Abstract* - In indoor wireless communication networks, mobile computing devices need to be handed off to different base stations based on certain criteria. Though many handover schemes have been proposed yet it has been a big ask to deliver the desired quality of service. In this paper, an efficient handover scheme has been proposed which makes use of fuzzy logic theory. The proposed system uses four parameters to make the handover decision. These parameters are signal strength from the home base station (SSH), signal strength from the neighboring base station (SSN) , distance between the mobile user (MU) and the home base station and the available network bandwidth (BW) in the neighboring cells. Higher bandwidth helps in achieving high data rates. The simulation results reveal that the proposed system executes an efficient handoff taking care of the probability of the false handover.**

1. INTRODUCTION

In outdoor wireless communication networks such as mobile WiMAX (802.16e), mobile stations move all the time, thus they need to be handed off to different base stations based on certain criteria. In this paper, a fuzzy logic based scheme for fast selection of best base station and of handover technique at the handover time is presented in order to minimize the delay during handover for sensitive multimedia traffic. The scheme considers several parameters such as receiver power levels, handover type, traffic type, base station load and mobile station speed for making the handover decision by the mobile station.A user initiates a call in one cell and is quite likely to move to another cell while the user in unaware of the fact that the serving base station has changed. This process of changing the base station in order to maintain the required signal level is termed as handover i.e. the old serving base station handovers the call to a new base station. The handover enables a user to continue his call in a new cell without any interruption. Handover not only involves looking for a new base station in a cell but also allocating voice and control signals to the channel which is to be allocated the new entrant in the cell. It should provide the desired quality of service (QoS) to the MU. Handover can be classified into 2 main types:

1. Hard Handover: In hard handover, A MU is connected to only one base station throughout the entire call and is mostly used GSM and GPRS systems. Hard handover can be seamless or non-seamless.

1\. 2. Soft Handover: A soft handoff allows two or moreconnections to neighboring base stations to be monitored i.e. the network is overlapped.

Handoff may also be classified as Vertical handoff and horizontal handoff [2].

The signal strength from MU is continuously monitored through a reverse channel by the cell site. As MU moves away from the base station, there is a gradual decrease in the signal strength from the base station. The cell sitebegins to look for newer base stations to handover the call when the signal strength decreases continuously. When the signal strength reaches a threshold level, the call is transferred to a new base which presents the maximum signal strength to the MU.

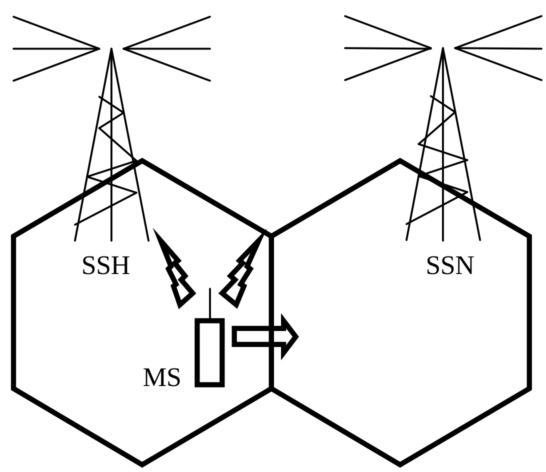


Fig. 1. A cellular mobile system

The traditional schemes generally makes use of the received signal strength from the home base station (SSH) and that of the neighboring base station (SSN) while ignoring the other parameters such as the available network bandwidth, distance which changes continuously according to the velocity of the MU and play an important role in the handover decision making.

In this paper, we have proposed a handover scheme using fuzzy logic approach considering four parameters stated below:

1. SSH – denotes the signal strength from the home base station. It indicates the availability of the network.

SSN – denotes the signal strength from the neighboring base station. It indicates the availability in the neighboring network

1. Distance (D) – denotes how far the mobile unit is from the home base station.
2. Network Bandwidth (BW) – denotes the availability of channels the neighboring cells.

2. RELATED WORK

In the past, several schemes have been suggested in order to provide a seamless handoff.

Leonard Barolli, FatosXhafa, Arjan Koyama, Akio Koyama, Makota Takizawa [1] proposed an intelligent handoff system using fuzzy logic and random walk model considering parameters signal strength from present and neighboring base station and the distance between mobile station and base station. It removes the ping-pong effect during handover.

PresilaIsrat, NamviChakma and M.M.A Hashem[2] proposed adaptive handoff management protocol using fuzzy logic approach considering parameters received signal strength, distance and speed. They provide simulation for both inter and intra-system handoff.

*Qing He[*3] proposed vertical handoff decision algorithm between WWAN and WLAN using fuzzy logic approach considering parameters received signal strength, available network bandwidth, monetary cost and user preferences. It reduces redundant handoffs and balanced network resources.

3. HANDOFF DECISION PROBLEM

To enhance the system capacity and efficient utilization of the channel bandwidth, the cells have been divided into smaller micro cells. In such a case, the handoff occurrence is inversely proportional to the cell size. Such conditions demand the triggering of handoff only if necessary i.e. the time at which the handoff occurs should be accurate. Due to the fading effects in cellular environment, it is hard to design an appropriate algorithm which executes the handoff at that instant.

To eliminate the unwanted handoff's, a value greater than the threshold value is set. If the threshold level at the cell boundary is -90dB, we set up a value higher than -90dB say -90dB+α such that the handoff is triggered at this threshold. The handover decision depends upon the value of α. Selecting an appropriate value of α reduces the unnecessary handoff's.

The above proposition does not give good results all the time so we have to consider some other parameters to make the accurate handover decision. Since the available channels are limited and all the users cannot be the accommodated at same time, so the base station looks for those neighbor which have sufficient bandwidth and best signal strength.

4. FUZZY LOGIC APPROACH

1. Fuzzification: is the process where the crisp values are converted into fuzzy. The uncertainty in crisp values form the fuzzy values. This conversion is represented by the Member functions. Hence fuzzification process gives the membership values for the given crisp quantities.Fuzzy interference system: It is the heart of the fuzzy logic system. It is also called as fuzzy rule based systems. The FIS formulates suitable rules and makes the decision based upon the set of rules.

2. De-fuzzification: De-fuzzification performs the inverse function of the fuzzifier. It involves fuzzy to crisp conversions so that the crisp quantities may be utilized for further processing. It is also called as "rounding off" method. Several methods are available for defuzzifying like centroid method, weighted average method, center of sums etc.

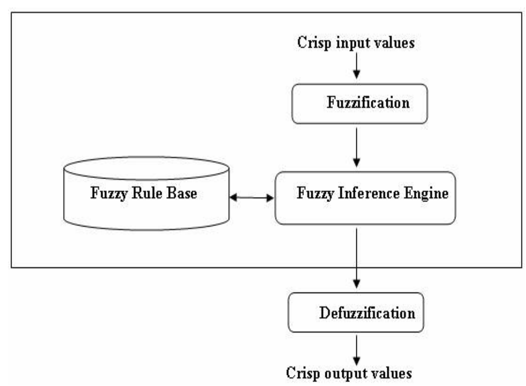


Fig. 2: Fuzzification

In the proposed fuzzy logic system, the input to the FIS are the signal strength from the home base station (SSH) and the neighbor base station (SSN) , the distance of mobile unit from the home base station (D) and the available network bandwidth in the neighbor cells. The output is the handover decision (HO).

In this paper, the handover threshold is set as 0.937 i.e. as soon as the handover value nears this value, the handover process begins.

1. $HO_{TH} = 0.937$

The FIS allocates the membership values to the input and the output parameters. The following fuzzy sets are formed:

μ (SSH) = {Weak (W), Not So Weak (NSW), Good (G) and Excellent (E)}

μ (SSN) = {Weak (W), Not So Weak (NSW), Good (G) and Excellent (E)}

μ (D) = {Near (N), Not So Near (NSN), Not So Far (NSF), Far (F)}

μ (ABW) = {Very Low (VL), Low (L), High (H), Very High (VH)}

μ (HO) = {Low (L), Medium (M), High (H), Very High (VH)}

The FIS houses a system called Fuzzy Rule Base (FRB) which consists of several rules. These rules are formed by several if-then statements which are responsible for the decision making and producing the requisite output. Such rule based systems are called as universal approximators.

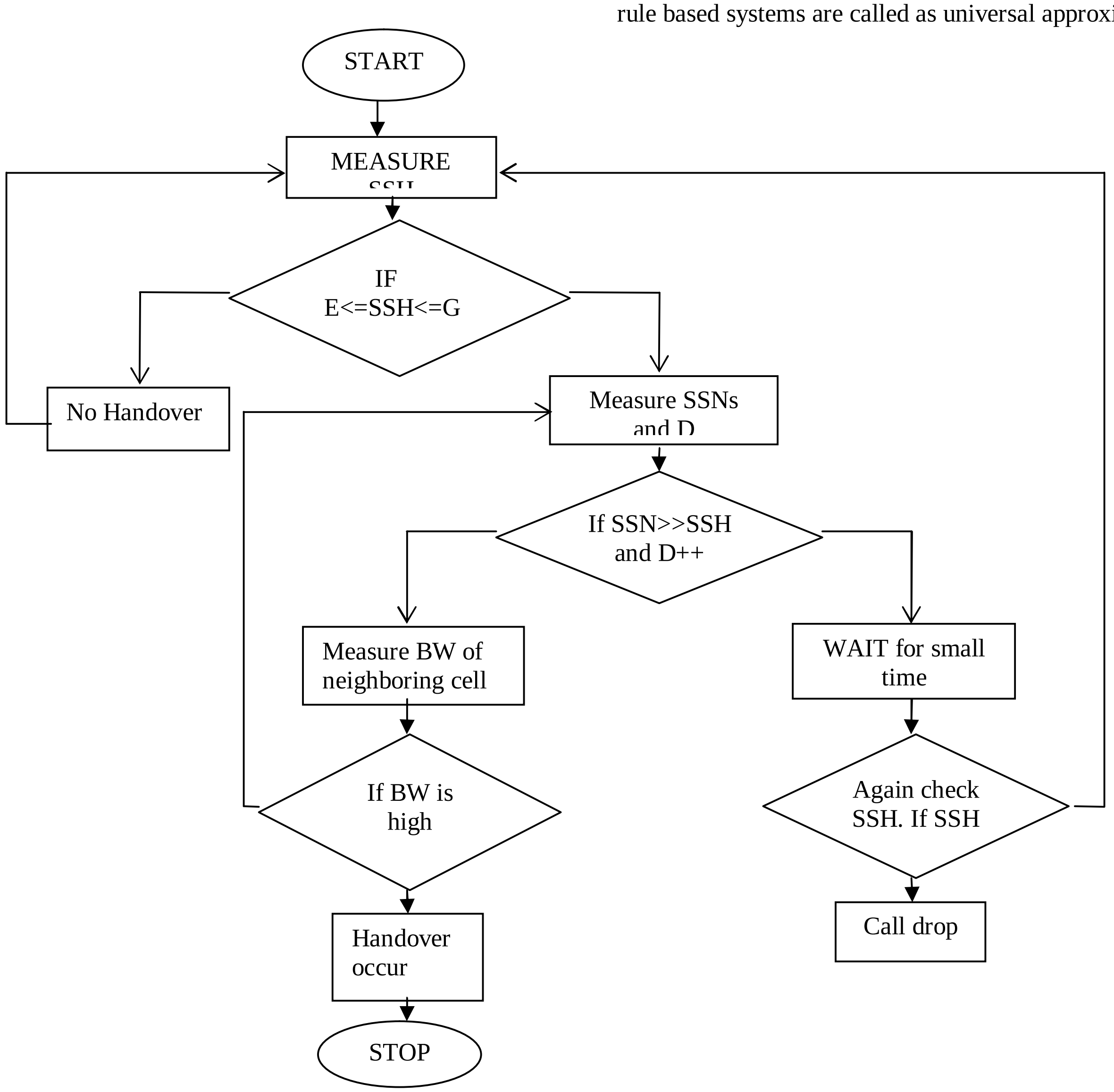


Figure 3. Flowchart for handover decision

Figure 3 represents the flow chart which is the basis of the simulation in order to produce an efficient handover received signal strength in the home cell with a minimal distance from the base station eliminates any need for handover. As the signal strength reduces to a value below a prescribed level, then the system looks for the signal strength of neighboring base stations and the distance between the MU and the home station. If the MU is nearing the cell boundary and the distance from the home cell is increasing rapidly, it looks for the new base stations in the new cell and measures their receiver signal strength and the station begins which has good signal strength as well as high available bandwidth. If the MU is near the base station home cell but the signal strength is quite low i.e. the MU is in hole, the base station waits for some time. If the user remains in a hole for a longer period, the call drop may occur.

Table 1 represents the set of rules which have been formulated to calculate the handover decision.
On the basis of these rules we have calculated the handoff values in various situations considering the other parameters also. We have taken lesser number of rules in order to minimize latency and take the handover decision at accurate time.

5. SIMULATION RESULTS

a. Relation between SSH, D and HO: Figure 4 represents the relation between the parameters. As the distance from the BS increases and the SSH decreases, we see a gradual increase in the value of handover.

b. Relation between SSH, SSN and HO: Figure 5 represents the relation between the parameters. As the SSN>>>SSH, the probability of handover is more.

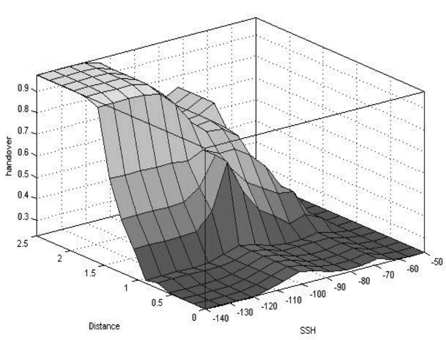


Fig. 4. Relation between SSH, D and HO

decision. The base station continuously monitors the signal strength through a reverse channel. A high bandwidth available in the new cells. The handover criteria is based on the fact that the new base station should not only have a good signal strength but also sufficient bandwidth to accommodate high data rates. If the signal strength of neighboring base station is excellent but the available network bandwidth is very low, then a search for a new base decision. The base station continuously monitors the signal strength through a reverse channel. A high

SSH	Distance	Bandwidth	SSN	HO	SSH	Distance	Bandwidth	SSN	HO
E	-	-	-	L	NSW	F	H	G	H
G	N	-	W	L	NSW	F	H	E	VH
G	N	-	NSW	L	NSW	F	VH	E	VH
G	NSN	-	W	L	W	N	-	-	L
G	NSN	-	NSW	L	W	NSN	-	W	L
G	NSF	-	W	L	W	NSN	-	NSW	L
G	NSF	-	NSW	M	W	NSF	VL	W	L
G	NSF	-	G	M	W	NSF	VL	NSW	L
NSW	N	-	W	M	W	NSF	VL	G	L
NSW	NSN	-	W	M	W	NSF	L	G	M
NSW	NSF	VL	W	L	W	NSF	H	G	H
NSW	NSF	VL	NSW	L	W	NSF	VH	G	H
NSW	NSF	L	NSW	M	W	F	VL	G	L
NSW	NSF	H	NSW	M	W	F	L	G	M
NSW	NSF	VL	G	M	W	F	H	G	VH
NSW	NSF	L	G	M	W	F	VH	G	VH
NSW	NSF	H	G	H	W	F	VL	E	M
NSW	F	VL	NSW	L	W	F	L	E	H
NSW	F	L	NSW	M	W	F	H	E	VH
NSW	F	H	NSW	M	W	F	VH	E	VH

Table 1. Fuzzy rules set

c. Relation between SSN, BW and HO: Figure 6 represents the relation between the parameters. As the SSN>>>SSH, the probability of handover is more.

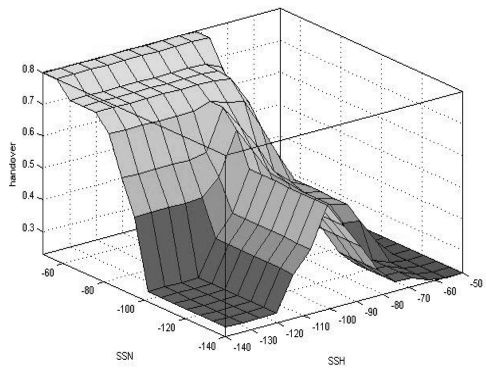


Fig. 5. Relation between SSH, SSN and HO

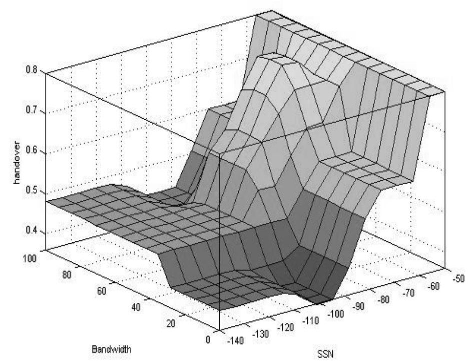


Fig. 6.Relation between SSH, BW and HO

Table 2 shows the minimum and maximum values of various parameters to make the accurate handover decision and Table 3 shows the results of the simulation. The handover threshold has been set to 0.937.

	Min	Max
SSH	-140dBm	-60dBm
D	0km	2.5km
BW	0kbps	100kbps
SSN	-140dBm	-60dBm

Table 2. Min and max values of parameters

	Min	Max
SSH	-140dBm	-60dBm
D	0km	2.5km
BW	0kbps	100kbps
SSN	-140dBm	-60dBm

Table 3. Simulation results

CONCLUSION

The proposed fuzzy logic system that we have used to demonstrate the handover decision making , emphasis on the available bandwidth from the neighboring cells. The simulation results show that a better handover decision can be made by selecting the required threshold limits. The simulation results vary depending upon the varying user requirements. However, in real situations the fuzzy sets can be modified to achieve better effects

REFERENCES

1. Leonard Barolli, FatosXhafa, Arjan Koyama, Akio Koyama, MakotaTakizawa, “An Intelligent Handoff System for Wireless Cellular Networks Using Fuzzy Logic and Random Walk Model”, Proc. 2008 IEEE International Conference on Complex, Intelligent and Software Intensive Systems, pp. 5-11, 2008.
2. PresilaIsrat, NamviChakma and M.M.A Hashem, “A Fuzzy Logic-Based Adaptive Handoff Management Protocol for Next-Generation Wireless Systems”, Proc. 11th International Conference on Computer and Information Technology ICCIT 2008 , pp.288-293, Dec. 2008.
3. Qing He, “A Fuzzy Logic Based Vertical Handoff Decision Algorithm between WWAN and WLAN”, Proc. International Conference on Networking and Digital Society, pp. 561-564, IEEE 2010.

Insurance Aspects & Maximum Probable Loss Considerations for Construction of Transmission & Distribution Lines

Sanjeev Soni
Associate Professor (EEE), Amritsar College of Engineering & Technology, Amritsar
Email: ssoni_79@yahoo.com

Abstract: **For more than a century we are using electricity. Transmission & Distribution (T&D) lines used to distribute electrical energy to the places often far away from where it has been generated. Today the demand of electrical energy is still on the increase causing constant modifications, extensions and development of the current networks. Thus the financial risk involved might be transferred to insurers. This paper laid down the different insurance aspects concerning construction of T&D Lines and also in the event of maximum probable loss caused due to the climatic conditions such as windstorms, ice, avalanche etc.**

Index Terms: T&D, Maximum Probable Loss (MPL), Third Party Liability (TPL), Increased Cost of Work (ICOW) Probable Maximum Loss (PML) etc.

I. INSURANCE ASPECTS

The Insurance aspect of T&D lines includes mainly the following point:

A) MATERIALS DAMAGE:

i) Natural Perils: Transmission & Distribution (T&D) lines are characterized by a widespread physical presence. Due to their high susceptibility to natural hazards, especially windstorm and ice, these lines harbor a massive loss accumulation potential. This is why insurers are carefully insuring such risks. The natural hazards accumulation risk is normally controlled by the application of reasonable sub-limits. Although T&D lines are designed to perform during extreme weather conditions, even worse weather conditions could affect the network resulting in big damage.

Another hazard is flood exposure of overhead power transmission line foundations during construction. Excavation works for foundations of power transmission lines can suffer severe damage caused by erosion of stagnant water. Pouring of concrete for foundations should occur within a few days following completion of the excavation. The number of excavations open at any one time should be limited to the minimum required for performance of works according to the local conditions and works programme. Unless compliance is evident a special condition or preferably special exclusion should be added. Bush fire can also represent a large exposure for the T&D lines. This exposure can be reduced by the control of the vegetation and/or adequate routing. Lighting is also a possible peril, nevertheless, some equipment is fitted to deal with such exposure, during the maintenance this peril and damaged equipment is excluded.

ii) Serial Losses: A transmission line basically consists of masts or poles and wires. Consequently the same kind of losses could be triggered by the same failure of design, workmanship or material. This could be limited by adding serial loss clause.

iii) Theft and Burglary: Because of the recent increase in the price of copper material, the hazard to theft of copper cables has increased. This can be limited by storing the wires in guarded and fenced storage areas.

iv) Access Roads: T&D lines are found almost everywhere on earth from mountainous areas to the desert from cities to areas with almost no inhabitants. Especially for remote areas access to the construction site can be very difficult or even be inaccessible for some time during the year because of snow. In case of damage this could prolong the repair time, as well as increase the repair cost, after demobilization of the working construction site and preassembling areas, in particular during the maintenance period.

v) Inland Transit: The size of transported equipment might cause difficulties because of limited access roads. Inland transit cover should therefore be carefully assessed.

vi) Special Equipment: Special equipment such as cranes or helicopters might be needed for the construction of a T&D lines. Following a loss this might trigger additional costs because such equipment is not readily available.

Fig.1: Special Equipments used in Transmission Line

vii) Reliability of Electrical Power System: The power system should be designed in such a way that no damage occurs in case of a reasonably foreseeable contingency. This means that design criteria take into account natural events of a certain return period (e.g. 50 years). Consequently one could consider a clause that excludes losses below the design criteria for natural perils.

B) THIRD PARTY LIABILITY:

Third party liability exposure depends very much on the route of the T&D line. It can be very low in remote areas such as mountains but significantly increase in areas of large populations such as cities. TPL exposure must be carefully assessed for each individually project.

TPL exposure will be in direct relation to Urban density. The lower the voltage, the higher the penetration of Urban areas. The most important exposure phases are related to the connection works to the grid and town substations and testing & commissioning period.

Damage to crops, forest and cultures as a consequence of fire could happen in some country areas, and can be controlled by adequate endorsement. Scaffolding and other method of protection shall be provided to minimize and mitigate surroundings exposure. Employer's liability during construction and erection period should be considered due to the nature of the works [4].

C) DELAY IN START UP:

The sum insured under DSU generally amounts to the difference between expected revenues and the variable costs, i.e. costs not incurred if the project is inoperative. This is also called the annual gross profit. The insurance can only be triggered by a material damage loss covered under section I of the CAR/EAR policy. Indemnity under the policy should be on actual loss sustained basis. For a detailed description of the DSU we refer to the lMlA paper WGP 63 (09).

The main issue in order to establish the scope of this cover is the difficulty in determining the real impact on profit due to interrupted line as usually lines are connected in a grid with alternative routing. Developing countries have a higher exposure due to the lack of development of their grids, therefore the alternative routes cannot be considered in case of a loss. The delay in restoring and repairing works can be longer depending on the difficulty of access.

i) Electrical Price Volatility and leeway Clause: The price for electrical energy is very volatile. This is why it is rather difficult to know what the actual loss sustained could be in case of a delayed start up of a plant a few years after policy signing. The effect of the price volatility of electrical energy can be limited by a leeway clause. This means that a change on the price up to a fixed percentage of the sum insured for DSU could be included. However this should not change that a claim shall always be indemnified on the basis of actual loss sustained. In addition the maximum indemnity should not exceed the sum insured multiplied for the leeway stated in the policy. The premium should be adjusted at the end of the policy period accordingly.

ii) Consequential Damages: Our world has become extremely dependent on electrical energy. As a consequence even a small damage to a power line can affect many industrial plants that depend on electricity supplied by the line resulting in large consequential business interruption losses. Needless to say that the consequential damage to an industrial plant that depends for its operation on the T&D line can be very costly. The corresponding exposure should therefore be assessment carefully.

iii) Contingent Business Interruption: DSU can be extended to include contingent events like Denial of Access, Public Authorities, Suppliers and Customers. With regard to these exposures we refer to the lMlA paper WGP 55 (08).

iv) Increased Costs of Working (ICOW): One big advantage of electrical power is that it can be easily transported via large distances. The failure of an overhead line does not necessarily mean complete loss of power to a region because there might be possible alternatives that could be used to reroute electrical energy. The costs due to rerouting could be covered by the increased cost of working section.

v) Risk Management Service: The implementation of adequate risk management measures plays an important role when insuring T&D lines.

vi) Accumulation: Accumulation of T&D lines is mainly an issue of operational covers. The reason for this is that T&D lines under construction will exit the construction policy as soon as the risk is handed over to the principal or put into operation [6].

II. MPL CONSIDERATIONS

The Maximum Probable Loss is an estimate of the maximum loss which could be sustained by the insurers as a result of any one occurrence considered by the underwriter to be within the realms of probability. This ignores such coincidence and catastrophes as may be possibilities but which remain highly improbable. As MPL assessment depends on several peculiar and geographical instances all the relevant considerations have been dealt with according to criteria governing loss events frequency and severity regardless of specific situations which would certainly affect the MPL. These factors can be briefly grouped into four categories as follows:

- Policy Wordings Scope (perils covered and extensions);
- External Hazards
- Natural Hazards
- Project Intrinsic Hazards

i) Policy Wordings Scope: The wording directly applies to material damage MPL assessment to estimate the ultimate exposure under the policy resulting from the combination of technical aspects and insurance related issues. The most common extensions which have to be included in the MPL assessment (full limits summed up to base MPL material damages) are:

- Expediting expenses

- Removal of debris
- Escalation/ Indexation Clause
- Existing Properties
- Increased cost of working
- Experts fees
- Third party liability

ii) External Hazards: External hazards category refers to perils rising from entities/individuals either part of the project environment or involved in its development.

Given the strategic importance of T&D lines (power supply) and the actual impossibility to have continuous watchman service and/or fenced areas all along the construction site, it makes exposure to third party individuals' actions somewhat critical.

Terror attacks (easy target and large consequential damages arising from delay in startup) and consequences of strikes riots or civil commotions can affect this type of projects with higher frequency than other engineering projects with well-defined and watched locations.

Fig.2: Potential Hazards due to Terror Attacks

Valuable goods stored at the constructions site or partially assembled along the line (copper cables and ceramics insulators) are also exposed to theft notwithstanding the relative difficulties rising from transportation of bulky items.

Finally, other nearby man-made hazards (e.g. upstream dams, airport and motor ways) should always be adequately taken into considerations they could result in an increased exposure to flood and aircrafts/vehicles impact.

iii) Natural Hazards: Natural hazards as already mentioned are the most likely MPL scenario for most of the T&D lines projects. The length of these lines (up to hundreds kilometers) and the relative low value per kilometers suggest an event – like earthquake flood or wind storm – affecting large sections of a line to be the actual MPL scenario.

Material damages severity due to natural hazards mainly depends on the maximum wideness of the T&D line section (it could also be up to the full line for specific hazards like earthquake) which can be actually affected by a single event. Partial hand over of T&D lines other than for large projects including different sections part of a network is usually not possible as a line from A to B can be operated only when it is completely finished and connected to its own substations.

As a consequence completed sections not in operation remain exposed to natural hazards for the whole construction policy period until the completion of the whole project.

Fig.3: Natural Hazards (Windstorm, Earthquake, Ice Avalanche, Erosion of foundation following flooding & danger from landslide or mudflow.)

iv) Project Intrinsic Hazards: Single elements of a T&D line do not present any key exposure as foundations towers frames, cables and insulators are well known and engineered items. On other hand, T&D line can run through remote and impressive are where site accessibility, erection of high rise tower and soil condition can be critical.

Use of special equipment like mobile cranes or helicopter result in additional exposure during erection phases as these equipments are critical to operate subject to changes in local condition which could affect their effectiveness (e.g. bad weather, wind, crane stabilizers not appropriately position on stable soil) resulting in material damages like partial tower collapse [3].

III. MPL ASSESSMENT PROCESS

A general guidance to the MPL assessment process including detailed information on each stage is provided by means of the following flowchart. Project information can consistently vary case by case and sufficient information depends on specific features although they should at least a grant to identify the construction process and a breakdown of cost for various items. Once the most probable hazards have been identified an easy cross check of probable maximum loss and maximum possible loss helps to better define the cost of reinstating the lost or damaged portion of works under PML scenario.

Finally, additional costs because of policy extension have to add up to the amount of physical damages to get the MPL which is usually more than the original cost of damaged item.

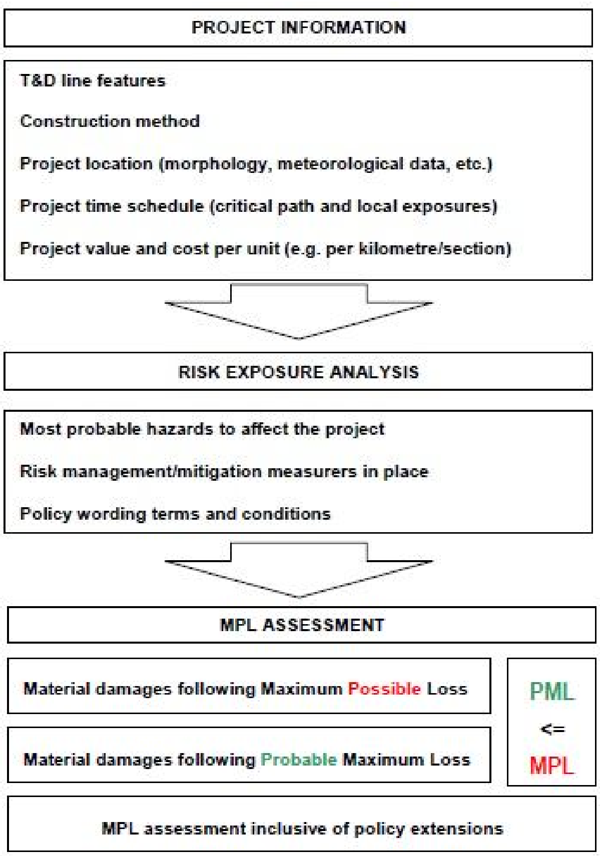


Fig. 4: MPL Assessment Process

IV. MPL AND LOSS SCENARIOS

Loss scenarios which can affect T&D overhead lines projects are various and impact (frequency and severity) various significantly depending on local conditions.

The table shows the most common scenarios grouped by hazard categories described in previous paragraph and this can be considered as a rationale irrespective of specific project feature and local external condition which could be prevailing on other hazards.

The scenarios are solid base to identify major exposures but the cannot be accounted as completely exhaustive of possible loss circumstances [7].

Hazards	MPL Scenarios	Frequency H	Frequency M	Frequency L	Severity H	Severity M	Severity L
Natural hazards							
Earthquake	Total loss of most of the towers (incl. foundations) and cables for projects within 50km from the epicentre.				X		
Ice and snow accumulation	Large sections of the line (e.g. highest elevation amsl) can be affected by the same event resulting in towers and cable collapse.				X		
Wind storms	Large sections of the line can be affected (mainly towers) for a continuous period also during cable laying.				2)	X	
Flood	Total loss of foundations and earthworks for large sections (e.g. in a valley) plus possible towers collapse.		1)			X	3)
Landslides and avalanches	Limited sections affected by total losses resulting in consistently increased costs of /time for reconstruction.						X
Lightning	Lightning storms usually affect limited areas with possible damages to a limited number of substation(s) and/or single towers.						X
Subsidence	Local effects depending on subsoil conditions affecting towers' foundations (also as consequence of EQ).						X
External Hazards							
Aircraft impact	Minor local damages depending on nearby airfields and missing signals during cables laying across valleys.			X			X
Bush fires	From limited to large sections affected by bush fires ignited by external causes resulting in damages to towers (jeopardized stability) and cables (efficiency).			X		X	
Terrorism & SRCC	Local attacks (e.g. explosive devices) to destroy substations or single towers are most likely limited to restricted areas/sections.		X				X
Theft	Theft of minor to moderate quantities of valuable goods (bulky items) stored at the construction site or partially assembled along the line.	X				X	
Nearby man-made hazards	Railway lines, motorways, other overhead lines, power plants, dams (basins/tailing facilities) etc. which bring additional exposure due to related activities (e.g. fuel , pressure vessels explosions or flood waves).			X	4)	X	
Project intrinsic hazards							
Fire	The most exposed items are substations and storage areas and PML usually refers to the largest fire unit.		X			X	
Faulty design and workmanship	Serial items (e.g. towers, cables or insulators) can suffer losses triggered by the same fault (faulty workmanship) although well-consolidated technologies and material allow considering the exposure as moderate.			X		5)	X
Construction operations	Lifting, erection and cable laying operations are intrinsically risky given special equipments and high rise structures although related to limited sections (e.g. 1 tower total loss because of crane's jib failure).		X				X

1) Frequency is not included for natural hazards as it depends on the location of the project.
2) Wind storm severity have to be increased to high in case of projects located in areas subject to heavy snow falls/freezing rain or hurricanes/typhoons. A layer of ice 1 cm thick means an additional weight of almost 100 kg per 100m and it increases the diameter and with it, the area of the cable exposed to wind forces.
3) Flood severity can be considered low where morphology allows to clearly separating different flood areas/chat basins.
4) Nearby man-made hazards have to increased to high in case of upstream dams or other installation which can trigger events which affects large areas.
5) Faulty design severity has to be increased to medium in case serial losses are not included in the policy wording.
6) *Natural Hazards:*

a) Earthquake: High exposure during the execution of foundation, substation and pylon erection. Direct and indirect damages and collapse of pylons could happen.
b) Windstorm: Producing losses over the whole project but mainly affecting pylon.
c) Ice and snow: Combined effect of ice and wind could change the aerodynamic conditions of the cables as a consequence of the higher charges produced by the deposit of ice and snow on them. Higher exposure at the end of the construction period and during testing.

Snow avalanches in hilly areas producing collapse of pylon.

d) Wild fire: Producing damages in the substation and stored equipments.

e) Lightening: Basically affecting substation and eventually one pylon.

f) Flood: Direct and indirect damages due to landslides affecting excavation works, foundations, access roads and producing pylon collapse.

g) Thermal Phenomena/solar wind: Induction current problems due to solar irradiation.

h) Soil Conditions: Affecting to excavation works and foundations.

7) Manmade Hazards:

- Terrorism: Higher exposure on substation due to accumulation. Some damages have also been produced to pylons.
- SRCC: Damage to substation that is the most exposed part of the project due to accumulation.
- Theft: Stored equipments (mainly cables) in case of absence of security measures.
- Operational Errors: Basically affecting to substation during testing and commissioning period.
- Aircraft collision: Producing loss or damage in pylons, cables and isolators during maintenance period as a consequence of maintenance works.

REFERENCES

[1] Blanco I, Brault Eric. *"Construction of Transmission & Distribution Lines"*, in the proceedings of International Conference of Electrical Power Systems, pp. 33-38, 2002.

[2] Jenne M. and Colautti M *"Potential Hazards in the Errection of Transmission Lines"* in Power Engineering Society Summer Meeting, Vol. 3, pp. 1645-56, 2000.

[3] Borges, C.L.T.; Falcao, D.M. *"Impact of Climatic Conditions on the Reliability, Losses and Cost Profile of a Transmission Line"*, in the proceedings of IEEE Power Tech Conference, Bologna, Vol. 2, 2003.

[4] Donnelly M. K., Dagle J. E., Trudnowski D. J., and Rogers G. J. "*Impacts of Third Party Utility on Transmission System Stability"*. IEEE Transactions on Power Systems, Vol. 11 (2), pp.741–746, 1996.

[5] Funabashi T., Koyanagi K., and Yokoyama R. "*A Review of Cost Estimation of Transmission & Distribution Linse"* in the proceedings of IEEE Power Tech Conference, Berlin, Vol. 2, pp. 1– 6, 2003.

[6] Joglekar J. Jayesh, Nerkar and Y.P. *"Erection & Insurance Cost in the Design of Transmission Lines"* in the proceedings of IEEE Transaction on Power Systems, Vol. 10 (3), pp. 41–46, 1998.

[7] Walters D. *"Maximum Probable Loss Parameters in the designing of Transmission Lines"* in the proceedings of IEEE Transaction on Power Systems, Vol. 10 (3), pp. 94–98, 1998.

Respiratory Sound Analysis using MATLAB

SonaliRajput[1] ,Satishpawar[2], Itenderpalsingh[3]
Mtech Ece Amritsar college of Engineering and Technology,Amritsar
Email: satishpwr75@gmail.com, itenderpalsingh@gmail.com

Abstract — **Bronchitis, pneumonia and many other pul-monary diseases cause respiratory disorders. The respira-tory sound signal can be processed by using several tech-niques for diagnostic information. Computerized analysis of respiratory sound can facilitates the detection of changes in respiratory sound & storing the record of the meas-urement made from patient. This may help for diagnosis of respiratory disorders and treatment for a patient suffering from various respiratory diseases. Most of the respiratory sound energy is concentrate below 200 Hz. Respiratory sound frequencies overlaps heart sound frequencies. Sup-pression of heart sound was done by using frequency do-main and adaptive signal filtering method. The respiratory sound signals are classified based on peak en-ergy of signal in 40Hz – 160Hz.**

Index Terms— Respiratory sound disorder, Heart sound signal, Respiratory sound signal

1. INTRODUCTION

Bronchitis, pneumonia and many other pulmonary diseases causes respiratory disorders. Characterization of the respiratory sound signal may help for diagnosis and monitor-ing of respiratory disorder. Diagnosis of respiratory disorder by auscultation is subjective process depend on the individu-als hearing, experience and ability to differentiate between the sound pattern [1], [2], [3]. The limitation of it can be overcome by using digital signal processing techniques [3], [4]. The quantification and analysis of noise free respiratory sound signal may be possible for better diagnosis. An inter-ference of cardiac sound signal into respiratory sound signal is called artifacts. Artifact suppression is needed for auto-matic diagnosis of respiratory disorder. Recording of res-piratory sound signal and signal processing techniques are needed for automatic diagnosis of respiratory sound disor-der. The artifact suppression was done by using frequency domain and adaptive filtering technique. The characteriza-tion of respiratory sound signal for diagnosis of respiratory disorder was done by spectral analysis and power density spectrum.

2. RESPIRATORY SOUNDS

Respiratory sounds originate in the large airways where air velocity and turbulence induce vibrations in the airway walls. These vibrations are then transmitted through the lung tissue and thoracic wall to the surface where they may be heard readily with the aid of a stethoscope. Respira-tory sounds can be classified into two categories, either normal or abnormal (adventitious).

2.1 Normal Breath Sounds: In normal breathing respiration sounds are generated during inspiration and expiration with louder inspiration phase. Respiratory sounds are normally heard over the trachea and larynx.

2.2 Abnormal Breath Sounds: These sounds are generated in certain pathological conditions of airways or lungs. Abnor-mal sounds may resemble a musical wind instrument. Identi-fication of abnormal respiratory sound can be used for the diagnosis of respiratory disorders. Sometimes more than one abnormal sound is simultaneously present. Further diagnosis is needed for separation of these sounds.

3. METHOD

The sound was picked up by using microphone. The transducer was placed on the throat region. The signal was amplified and fed to computer through audio input.

Combination of Adaptive and frequency domain Filtering

There are three main components of the combination method of adaptive and frequency domain filtering, the input or pri-mary signal, the noise signal (heart signal & other muscle artifact) or reference signal and filtered output signal. In this technique the recorded signal during normal breathing is a primary input and the signal recorded with breath holding is a reference input signal. respiratory sound is considered as a primary signal analysis, partial overlap between respiratory signal and heart signal and provide the primary input for adaptive noise cancellation technique. The heart sound sig-nal acquired from sensor is taken as reference input and con-taminated respiratory signal is taken as primary input. The noise signal contain heart signal and other high frequency noise To remove the high frequency noise FIR filter is ap-plied to both the signal. The output is the signal which is determined by subtracting reference signal from primary signal. Symbols (b(n) = breath sounds, m(n) = heart sound and other high frequency noise, e(n) = de-noised breath sounds

4. EXPERIMENTAL SET-UP

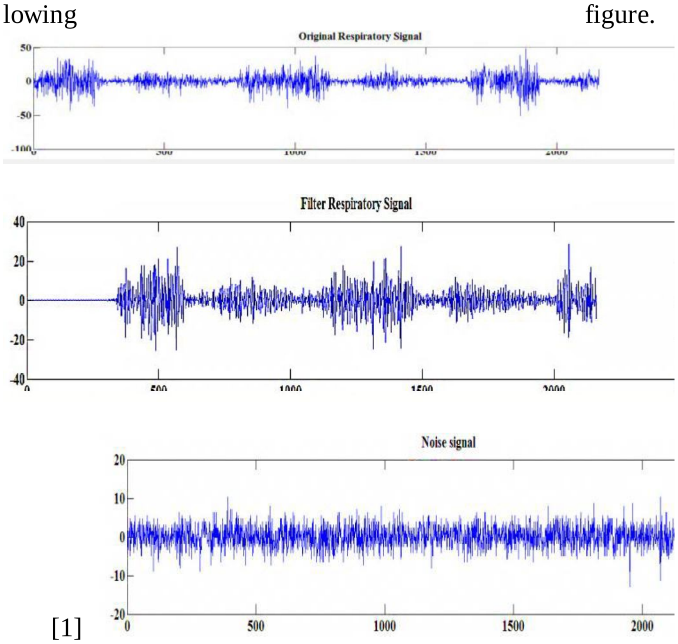
Recordings were taken from ten randomly selected subjects. Respiratory sound analysis instrument was used for recording the respiratory sound signal. A microphone based developed sensor used for recording, the sensor placed on the lower part of the throat for recording of signal with and without breath holding. Two signals was taken one signal with breathing (Respiratory signal) and other is signal without breathing (Noise signal)

5. ALGORITHM IMPLEMENTATION

The processing and analysis of respiratory sound signal is difficult due the interference of heart sound. Combination method of adaptive filtering & frequency domain filtering was used for removal of heart sound signal and other artifacts. The processing of signals was done in MATLAB Signals were recorded with and without breath holding consecutively. In MATLAB FFT was plotted to observe different frequency range of the signal according to that 40- 160 Hz selected. Then applied band pass filter for (40 -160) Hz to remove undesired frequency harmonics while saving the original information. After band pass filtered applied to the signal with breath holding remaining signal is heart signal. Then the signal recorded with breath holding was subtracted from the signal without breath holding. This way removal of sound artefacts is achieved. Power spectrum density verses frequencies of filtered respiratory sound signal was plotted. Peak of the waveform was found as the maximum value of power density. The corresponding frequency was calculated. Depending upon this frequency the abnormalities in respiratory sound signal decided.

6. RESULTS AND ANALYSIS

6.1 Filtering The recorded signal and band pass filtered signals and signal after applying combine method of adaptive and frequency domain filtering are shown in following figure.



[1]

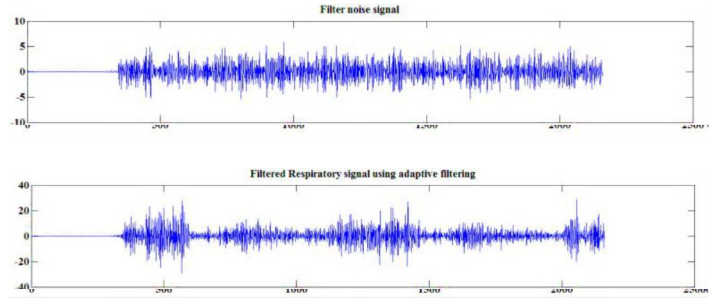


Fig.6.1 Typical Respiratory Sound Signal a. Original Respiratory Signal b. Filtered Respiratory Signal c. Noise Signal d. Filtered Noise Signal e. Filtered Respiratory Signal After Applied Combination Method of Adaptive and Frequency Domain Filtering

6.2 ANALYSIS OF RESPIRATORY SOUND FOR EXTRACTION OF DIAGNOSTIC INFORMATION

The noise suppressed respiratory sound signal from 10 different subjects certain frequency range for abnormalities in respiration sound signal was decided. For analysis of respiratory sound after adaptive and frequency domain filtering graph of power distribution over the frequency range was plotted. For normal subject the peak power was observed above 130 Hz. For fine crackles sound signal the maxima was observed in the range of 60-138 Hz. The maximum power was observed below 60 Hz for pleural sound

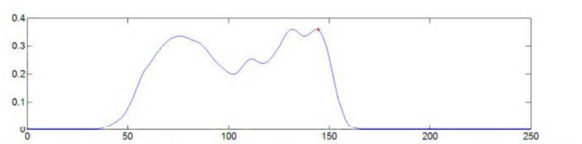


Fig. 6.1 Power Distribution over the Range of Frequency

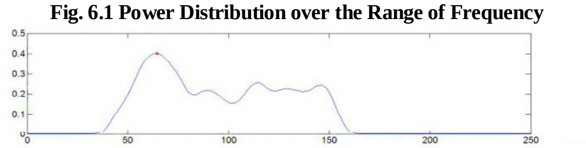


Fig. 6.2 Power Distribution Over the Range of Frequency

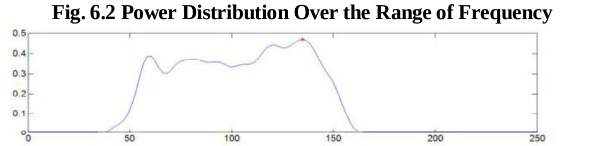


Fig. 6.3 Power Distribution over the Range of Frequency

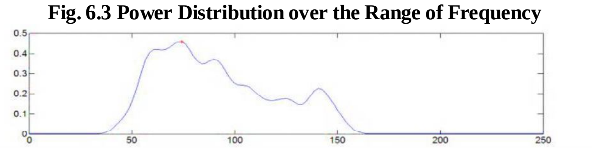


Fig. 6.4 Power Distribution over the Range of Frequency

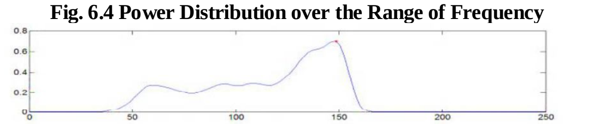


Fig. 6.5 Power Distribution over the Range of Frequency

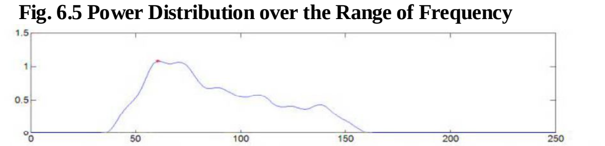


Fig. 6.6 Power Distribution over the Range of Frequency t 6)

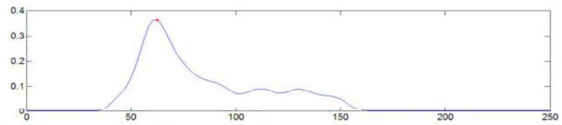


Fig. 6.7 Power Distribution over the Range of Frequency

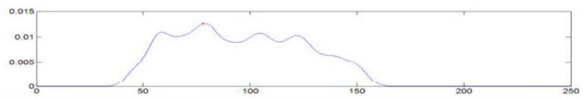


Fig. 6.8 Power Distribution over the Range of Frequency

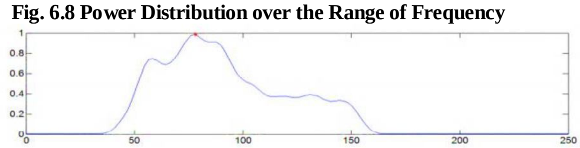


Fig. 6.9 Power Distribution over the Range of Frequency

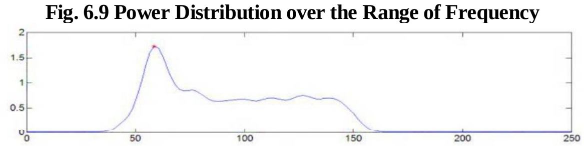


Fig. 6.10 Power Distribution over the Range of Frequency

7. CONCLUSION

The effective suppression of the heart sound signal was observed in combination method for noise filtering (Adaptive plus Frequency domain Filtering). After getting clear signal the power verses frequency plot is considered for respiratory sound analysis. From these graphs variations of relative frequency of maximum power was observed which is relative to the visual observations made from sound signal. Hence depending on position of maxima the respiratory sound is classified for indicating state of subject as normal or abnormal.

REFERENCES

[1] A. R. A. Sovijarvi, J. Vanderschoot, and J. E. Earis, “Standardization of Computerized Respiratory Sound Analysis,” Eur Repair Rev, vol 10 (77), pp. 58, 2000.

[2] J.E. Earis, B. M. G. Cheetham, “Current Methods used for Computerized Respiratory Sound Analysis,” Eur Repair Rev; vol 10 (77), pp. 585-590, 2000.

[3] Z. M. K. Moussavi, “Respiratory Sound Analysis,” IEEE Engineering in Medicine & Biology Magazine 0739-5175, 2007.

[4] Z. M. K. Moussavi, “Separating heart sound from lungs sounds,” IEEE Engineering in Medicine & Biology Magazine, 2007.

[5] Z. M. K. Moussavi, “Fundamentals of Respiratory Sound Analysis,” Morgan & Claypool, First Edition.

[6] W. F. Ganong, “Review of Medical Physiology (Seventeenth Edition), New Jersey, Prentice Hall, 1995.

[7] Guyton & Hall, ”A textbook of Medical Physiology”, Tenth Edition, printed in India at Gopsons Paper Ltd., Noida 201 301. [8] Johnson, L.R.,”Essential Medical Physiology (Second Edition). Philadelphia, Lippincott Williams & Wilkins, 1998.

[8] MedicineNet.com, article.htm.

[9] T.E. Ayoob Khan, Dr. P. Vijayakumar, ”Separating Heart sound from lung sound using LavView”, International Journal of Computer and Electrical Engineering, Vol. 2, No. 3, June, 2010 1793-8163.

[10] Rahman Jamal, Mike Cerna, John Hanks Designing Filters Using the Digital Filter Design Toolkit. [12] Emmanuel. C. Ifeachor, Barrie. W. Jervis, Digital Signal Processing [13] Edward P. Cunningham, Digital Filtering An Introduction, John Wiley & Sons, Inc.1995.

[11] Haykin S, Adaptive Filter Theory, 2nd Edition, United Stated of American: Prenticem Hall, Inc. 1991

[12] Krister Landernas, Implementation of Digital Filters Using Digital-Series Arithmetic’s, Department of electronics, Licentiate thesis No. 23, 2004.

Sensor Technology and Its use in Irrigation Management

Er.Sukhjit Singh[1], Er. Neha Sharma2
[1]Sukhjit Singh, M-Tech Scholar, ACET, Amritsar
[2]Neha Sharma, Asst.Professor(ECE),ACET, Amritsar
Email: er.sukhjit85@gmail.com, er.neha.ruchi@gmail.com

***Abstract*—A WSN is a system consisting of radio frequency (RF) transceivers, sensors, microcontrollers and power sources. Recent advances in wireless sensor networking technology have led to the development of low cost, low power, multifunctional sensor nodes. Sensor nodes enable environment sensing together with data processing. Instrumented with a variety of sensors, such as temperature, humidity and volatile compound detection, allow monitoring of different environments. They are able to network with other sensor systems and exchange data with external users. Sensor networks are used for a variety of applications, including wireless data acquisition, environmental monitoring, irrigation management, safety management, and in many other areas.**

Keywords—Precision Agriculture, wireless sensor Networks, drip irrigation , water level monitoring.

I. INTRODUCTION

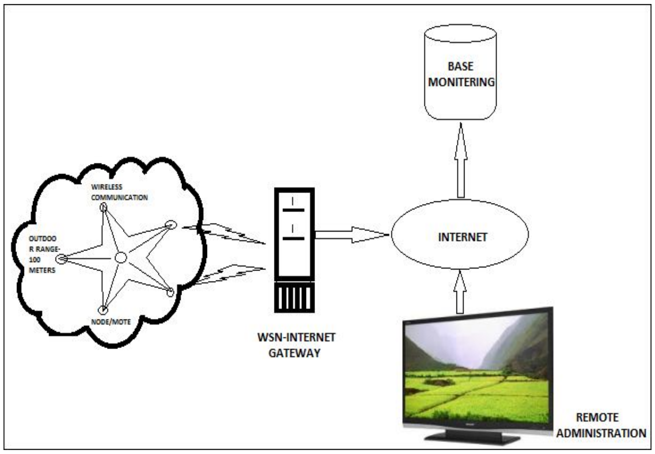
A general WSN protocol consists of the application layer, transport layer, network layer, data link layer, physical layer, power management plane, mobility management plane and the task management plane[1]. Currently two standard technologies are available for WSN: ZigBee and Bluetooth. Both operate within the Industrial Scientific and Medical (ISM) band of 2.4 GHz, which provides licensefree operations, huge spectrum allocation and worldwide compatibility. In general, as frequency increases, bandwidth increases allowing for higher data rates but power requirements are also higher and transmission distance is considerably shorter[2,3]. Multi-hop communication over the ISM band might well be possible in WSN since it consumes less power than traditional single hop communication[3].

II. PLANTATION MANAGEMENT USING WIRELESS SENSOR NETWORK

For developing an efficient system of plantation management, the foremost input is the availability of accurate data. This data includes soil properties, agronomic data, physicochemical parameters, atmospheric data, etc, preferably on a day-to-day basis or even hourly basis. Normal laboratory analysis of these parameters and manual decision-making take a long time even with the most sophisticated analytical techniques.

Most of the sampling procedures are not in-situ and samples have to be brought from the field to laboratories for analysis, a lot of time. By the time the results are available and decisions taken, the farm conditions might change making the decision inappropriate. Quick and quality decision-making at the farm level can enhance agricultural productivity and quality manifold. Computer-aided decision-making process can handle and analyse several input parameters at the same time involving large databases. Monitoring of physical and environmental parameters including soil moisture, soil temperature, soil pH, leaf temperature, relative humidity, air temperature, rainfall, vapour pressure and sunshine hours is done through a wireless sensor network(WSN).

WSN comprises spatially distributed sensors to monitor physical or environmental conditions. It is a comprehensive system that integrates sensing , wireless and processing technologies and is capable of spatially and temporally sensing different physical parameters without loss in the sensing accuracy. The parameters are processed and wirelessly transmitted to a centralised data storage system through a gateway from where they may be remotely accessed and analysed by the user.



Block Diagram of wireless network system

The system architecture of a WSN-based system consists of different sensors interfaced to electronic hardware with data processing capabilities. The electronic hardware is also equipped with wireless communication modules allowing the sensed data to be processed and transmitted according to a selected protocol. These hardware nodes are called motes in WSN terminology. Each of the motes is interfaced with a set of sensors depending on the application domain. The sensors may be programmed to sense at specific intervals or periodically in a day.

III. WSN IN AGRICULTURE

WSN technology can broadly be applied into three areas of agriculture : 1) Fertiliser control, 2) Irrigation management and 3) Pest management. The sensors that can be interfaced to the mote are temperature, relative humidity, solar radiation, rainfall, wind speed and direction, soil moisture and temperature, leaf wetness and soil pH sensors. These sensor-readings can be integrated with a decision support system that aids the management of resources to the crop.

IV. DRIP IRRIGATION AUTOMATION

Conventional irrigation methods like overhead sprinklers and flood-type feeding systems usually wet the lower leaves and stem of the plants. The entire soil surface is saturated and often stays wet long after irrigation is completed. Such a condition promotes infections by leaf mold fungi. Flood-type methods consume a large amount of water, but the area between crop rows remains dry and receives moisture only from the incidental rainfall.The drip irrigation technique slowly applies a small amount of water to the plant's root zone. Water is supplied frequently, often daily, to maintain favorable soil moisture condition and prevent moisture-stress in the plant with proper use of water resources.

WSN-based drip irrigation system is a real-time feedback control system which monitors and controls all the activities of the drip irrigation system. A typical system includes a delivery system, filters, pressure regulators, valve or gauges, chemical injectors, measuring sensors/ instruments and controllers.WSN framework installed in the field may gather various physical parameters related to irrigation. These includes ambient temperature, ambient humidity, soil temperature, drip water temperature, soil moisture, soil pH, water pressure, flow rate, amount of water, energy calculation(power),chemical concentration and water level.

The data is sent to the central server wirelessly through the motes and gateways. Based on the data ranges, the central server generates necessary control actions, which are routed to the respective controllers through control buses enabling implementation of closed-loop automation of the drip irrigation system.The basic feature of the product is to enable switching on and off of the motor remotely. The device ensures that all the fault conditions are checked and only then the motor is started.

V. CONCLUSION

Conventional Flood-type methods consume a large amount of water, but the area between crop rows remains dry and receives moisture only from the incidental rainfall whereas the drip irrigation technique slowly applies a small amount of water to the plant's root zone. So by using the wireless sensor drip irrigation technique, firstly we can control the wastage of water and secondly by using wireless sensor there is no need of laborers.

REFERENCES

[1] Callaway, E.H. "Wireless sensor networks: architectures and protocols." Auerbach Publications: New York, NY, USA, 2004.

[2] Qingshan, S.; Ying, L.; Gareth, D.; Brown, D. Wireless intelligent sensor networks for refrigerated vehicle. In IEEE 6th Symposium on Emerging Technologies Mobile and Wireless Communication, Shanghai, China, 2004.

[3] Mahir Dursun and Semih Ozden, "A wireless application of drip irrigation automation supported by soil moisture sensors", Scientific Research and Essays Vol. 6(7), pp. 1573-1582, 4 April, 2011. Available online at

http://www.academicjournals.org/SRE ISSN 1992-2248 ©2011 Academic Journals.

[4] Baggio A (2005). "Wireless sensor networks in precision agriculture". In: On-line Proc. of the Workshop on Real-World Wireless Sensor Networks, pp. 50-51.

[5] Kim Y, Evans RG, Iversen WM, Pierce FJ (2006). "Instrumentation and control for wireless sensor network for automated irrigation". ASAE Annual Int. Meeting, p. 061105.

[6] Steven R. Evett, R. Troy Peters, and Terry A. Howell,"contr-olling water use efficiency with irrigation automation: cases from drip and center pivot irrigation of corn and soybean".Southern Conservation Systems Conference, Amarillo TX, June 26-28, 2006 .

[7] Evans, R.G. 2000," Controls for Precision Irrigation with Self-Propelled Systems". In Proc. of the 4th Decennial National Irrigation Symposium. Robert G. Evans, Brian L. Benham, and Todd P. Trooien (eds.) November 14-16, Phoenix, AZ.

[8] Iqbal Singh, Meenakshi Bansal," Monitoring Water Level in Agriculture Using Sensor Networks".International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, Issue-5, November 2011.

A Review on Different Step Analysis of Document Using Image Processing

Varinder Singh
Department of Electronic and Communication , B.K.S.J.E.C , Amritsar , Punjab
Email: vinder200@yahoo.com

Abstract **--In this paper, we analyze the current trends in the document image processing. Document image processings refers to algorithms and techniques that are applied to images of documents to obtain a computer-readable description from pixel data. A well-known document image processing product is the Optical Character Recognition (OCR) software that recognizes characters in a scanned document. Analysis of document images for information extraction has become very prominent in recent past. Wide variety of information, which has been conventionally stored on paper, is now being converted into electronic form for better storage and intelligent processing. This needs processing of documents using image analysis, processing methods. This article provides an overview of various methods used for digital image processing using three main components: Pre-processing, Feature extraction and the Classification. Pre-processing includes Image acquisition, initiation, identification, Layout analysis, feature extraction and classification. Classification is an important step in Office Automation, Digital Libraries, and other document image analysis applications. This article examines the various methods used for document image processing in order to achieve a processed document having high quality, accuracy and fast retrieval.**

Keywords - OCR, Document, Analysis, Processing, Classification

1. INTRODUCTION

The objective of document image analysis is to recognize the text and graphics components in images of documents, and to extract the intended information as a human would. Two categories of document image analysis can be defined (see figure).Textual processing deals with the text components of a document image. Some tasks here are: determining the skew (any tilt at which the document may have been scanned into the computer), finding columns, paragraphs, text lines, and words, and finally recognizing the text (and possibly its attributes such as size, font etc.) by optical character recognition (OCR). Graphics processing deals with the non-textual line and symbol components that make up line diagrams, delimiting straight lines between text sections, company logos etc. Traditionally, our main form of transmission & storage for information has been by paper documents. These documents include many common types: business letters, forms, engineering drawing & maps, text books, technical manual, music notations & other symbolic data. Though paper was the exclusive medium in past, many documents now originate on the computers & often reside exclusively in electronic form. In spite of this it is unclear whether the computer has decreased/increased the amount of paper document produced, as these are printed out for reading, dissemination, markup predictions of paperless offices made so frequently during the early 1980 has given way to the realization that the objective is not elimination of paper but the ability to deal with the flow of electronic & paper document in effective & integrated way. Document processing in any organization whether having its operations manual or computerized, forms an essential activity in its functioning's. Within document processing, the key activity prior to all other activities is the recognition of documents and hence their categorization [1] [7] [10]. Several good solutions exist for document processing and analysis, this paper tries to give general idea for document processing and the various steps/methods used for that. This will give an overview for processing, analysis and classification of document images

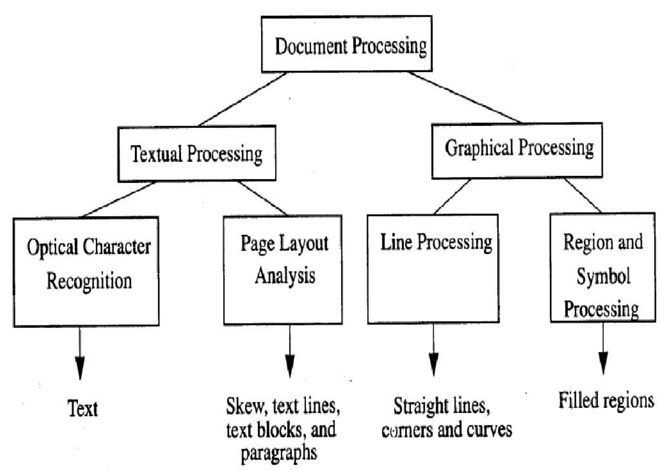


Fig. 1: A hierarchy of document processing sub areas listing the types of document components

2. DOCUMENT IMAGE ANALYSIS

The objective of Document Image analysis is to recognize the text & graphics components in image of documents & to extract intended information from them. Two categories of document image analysis can be defined.

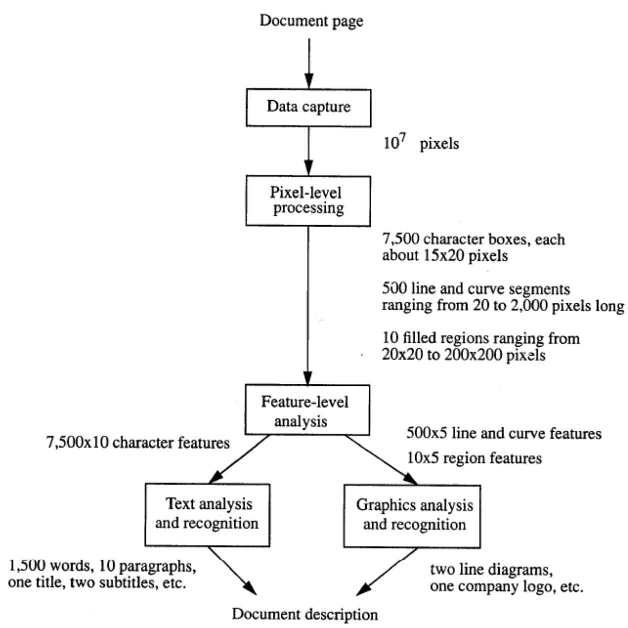


Fig. 2: A typical sequence of steps for document analysis

Text Processing **:-** Deals with the textual components of a document image & its task are; - Determining the skew (any tilt at which the documentmay have been scanned in the computer).- Finding columns, paragraphs, textual lines, words,recognizing the text (Possibly its attributes such as size, font etc) by OCR.

Graphical processing **:-** Deals with the non-textual elements (tables, lines, images, symbols, delimiters, company logo etc) Pictures are also included in this category; they are different from graphics in that they are often photographically or artistically generated.

3. PIXEL LEVEL PROCESSING

A hierarchy of document processing subareas listing the types of document components deal within each subarea. A hierarchy of document processing subareas listing the types of document components processing.

Processing of document to extract their content in an automated fashion is essential task in all types of organizations for varied applications. Any document under processing is subjected to the following steps as depicted in figure 2.

1) The Pixel level processing or pre-processing Stage that enhances the quality of the input image & locate the data of interest.
2) The feature level analysis stage that captures the distinctive characteristics of the document under processing.
3) The classification stage that identifies the document; groups the according to certain classes & helps in their efficient recognition. "Figure 2"

3.1 Pixel level Processing or Pre-Processing

Pixel-level processing or low-level or pre- processing is done on the captured image to prepare it for further analysis. Such processing includes: Thresholding to reduce a grayscale or color image to a binary image, reduction of noise to reduce extraneous data, segmentation to separate various components in the image, and, finally, thinning or boundary detection to enable easier subsequent detection of pertinent features and objects of interest.

3.2 Image Acquisition

Acquire/obtain the image of document in color, gray level or binary format.

3.3 Binarization

Converts the acquired image to binary format, the objective of binarization is to automatically choose a threshold that separates the foreground and background information. Selection of a good threshold is often a trial and error process (see figure 3).A grey level of 128 is set as threshold. This becomes particularly difficult in cases where the contrast between text pixels and background is low (for example, text printed on a gray background).

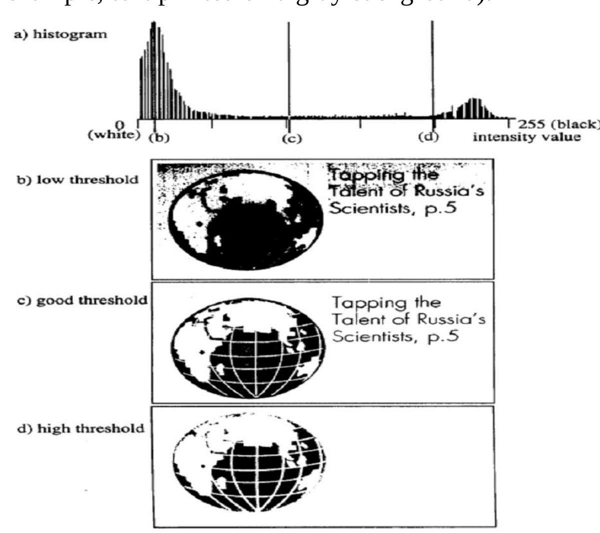


Fig. 3: Image binarization (a) Histogram of original grayscale image.

Horizontal axis shows markings for threshold values of images below. The lower peak is for the white background pixels, and the upper peak is for the black foreground pixels. Image binarized with: (b) too low a threshold value, (c) a good threshold value (d) too high a threshold value

3.4 Noise reduction

The data extraction procedure often requires binarizing the images, which discard most of the noise & replace the pixel in the image, character & the pixel in the background with binary 0 & 1 respectively. After binarization, document images are usually filtered to reduce noise. For documents, more specific filters can be designed t take advantage of the known Characteristics of the text and graph components. A document to be scanned can itself be contaminated with dust or spots etc which constitute noise. Scanning itself can introduce some amount of noise. Noise is also due to the degeneration, ageing, photocopying or during data capture. In order to make it suitable for further processing, a scanned document image is to be freed from any existing noise. This can be achieved by a method known as image enhancement-this means improvement of the image being viewed by the machine or human. It includes contrast adjustment, noise suppression & many others. Smoothing operations in document images are used for blurring and for noise reduction. Blurring is used in preprocessing steps such as removal of small details from an image. In binary (black and white) document images, smoothing operations are used to reduce the noise or to straighten the edges of the characters, for example, to fill the small gaps or to remove the small bumps in the edges (contours) of the characters. Smoothing and noise removal can be done by filtering. Filtering is a neighborhood operation, in which the value of any given pixel in the output image is determined by applying some algorithm to the values of the pixels in the neighborhood of the corresponding input pixel. Various methods are applied to reduce noise. The most important reason to reduce noise is to obtain easy way of recognition of documents.

3.5 Segmentation

Segmentation occurs on two levels. On the first level, if the document contains both text and graphics, these are separated for subsequent processing by different methods (Wong 1982; Fletcher 1988; Jain 1992). On the second level, segmentation is performed on text by locating columns, paragraphs, words, and characters; and on graphics, segmentation usually includes separating symbol and line components. For instance, in a page containing text and some illustrations similar to the pages of this journal, text and graphics are first separated. Then the text is separated into its components down to individual characters. The graphics is separated into its components such as rectangles, circles, connecting lines, symbols etc. After this step an image is typically broken down into its basic components such as an individual character.

3.6 Thinning and region detection

Thinning is an image processing operation in which binary valued image regions are reduced to lines that approximate the centre lines, or skeletons, of the regions. The purpose of thinning is to reduce the image components to their essential information so that further analysis and cognition are facilitated. For instance, a line drawing can be written with different pens giving different stroke thicknesses, but the information presented is the same. In figure 4, some images are shown whose contents can be analysed well due to thinning, and their thinning results are also shown here. Note should be made that thinning is also referred to as skeletonizing and core-line detection in the literature. We will use the term “thinning” to describe the procedure, and thinned line, or skeleton, to describe the results. A related term is the “medial axis”. This is the set of points of a region in which each point is equidistant to its two closest points on the boundary. The medial axis is often described as the ideal that thinning approaches. However, since the medial axis is defined only for continuous space, it can only be approximated by practical thinning techniques that operate on a sampled image in discrete space. Several thinning algorithms have been described by Arcelli & Sanniti di Baja (1985, 1993), and Sanniti di Baja (1994) and an algorithm that thins by several pixels in each pass is described by O’Gorman (1990). Reviews of thinning methods have been given by Lam et al (1992) and Lam & Suen (1995).

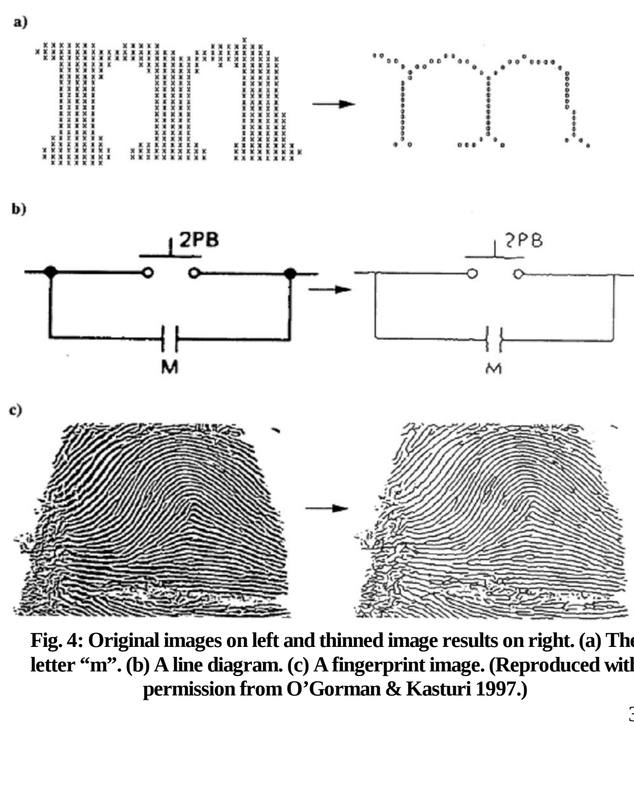
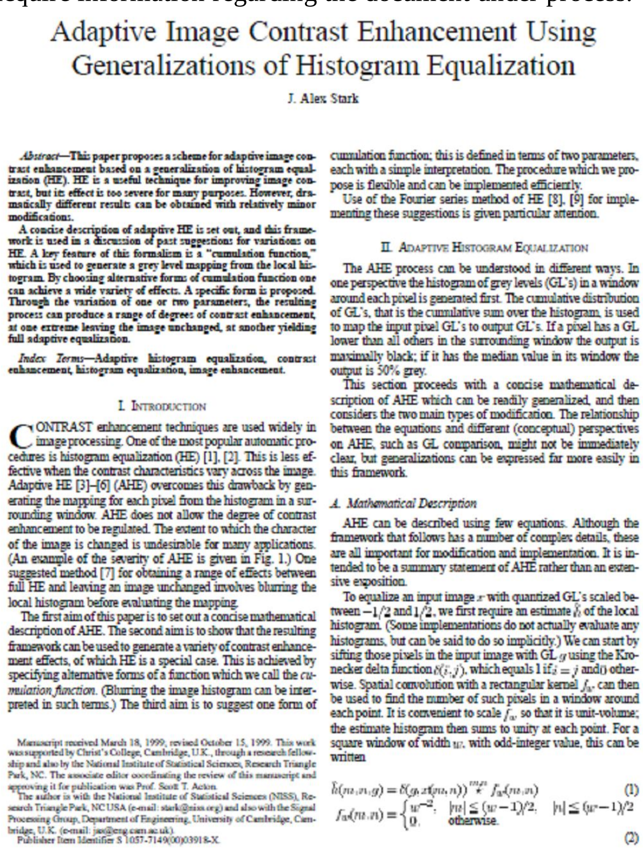


Fig. 4: Original images on left and thinned image results on right. (a) The letter “m”. (b) A line diagram. (c) A fingerprint image. (Reproduced with permission from O’Gorman & Kasturi 1997.)

3.7 Feature Level Analysis

Feature level analysis involves the extracting the meaningful information from the document image. So that it reduces the storage required. The features that are extracted from whole image are known as the global features & the features that are extracted from blocks identified during segmentation or from subdivision (sub sectioning) of the document are known s local features. They can be divided into several categories: textural, geometric, component, structural and content based [15].The extractions of global and local features provide input to classification algorithm/techniques. One of the most important advantages of feature extraction is that; it significantly reduces the information (compared to original image) to represent an image for understanding the context of that image. Simple feature extraction methods, like calculating the difference between the minimum and maximum coordinates of the document image and the shape of the document obtained by comparison of breadth and length of the document image are of prime importance to acquire information regarding the document under process.



Adaptive Image Contrast Enhancement Using Generalizations of Histogram Equalization

J. Alex Stark

Abstract—**This paper proposes a scheme for adaptive image contrast enhancement based on a generalization of histogram equalization (HE). HE is a useful technique for improving image contrast, but its effect is too severe for many purposes. However, dramatically different results can be obtained with relatively minor modifications.**

A concise description of adaptive HE is set out, and this framework is used in a discussion of past suggestions for variations on HE. A key feature of this formalism is a "cumulation function," which is used to generate a grey level mapping from the local histogram. By choosing alternative forms of cumulation function one can achieve a wide variety of effects. A specific form is proposed. Through the variation of one or two parameters, the resulting process can produce a range of degrees of contrast enhancement, at one extreme leaving the image unchanged, at another yielding full adaptive equalization.

Index Terms—**Adaptive histogram equalization, contrast enhancement, histogram equalization, image enhancement.**

I. INTRODUCTION

CONTRAST enhancement techniques are used widely in image processing. One of the most popular automatic procedures is histogram equalization (HE) [1], [2]. This is less effective when the contrast characteristics vary across the image. Adaptive HE [3]–[6] (AHE) overcomes this drawback by generating the mapping for each pixel from the histogram in a surrounding window. AHE does not allow the degree of contrast enhancement to be regulated. The extent to which the character of the image is changed is undesirable for many applications. (An example of the severity of AHE is given in Fig. 1.) One suggested method [7] for obtaining a range of effects between full HE and leaving an image unchanged involves blurring the local histogram before evaluating the mapping.

The first aim of this paper is to set out a concise mathematical description of AHE. The second aim is to show that the resulting framework can be used to generate a variety of contrast enhancement effects, of which HE is a special case. This is achieved by specifying alternative forms of a function which we call the *cumulation function*. (Blurring the image histogram can be interpreted in such terms.) The third aim is to suggest one form of cumulation function; this is defined in terms of two parameters, each with a simple interpretation. The procedure which we propose is flexible and can be implemented efficiently.

Use of the Fourier series method of HE [8], [9] for implementing these suggestions is given particular attention.

Manuscript received March 18, 1999; revised October 15, 1999. This work was supported by Christ's College, Cambridge, U.K., through a research fellowship and also by the National Institute of Statistical Sciences, Research Triangle Park, NC. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Scott T. Acton.

The author is with the National Institute of Statistical Sciences (NISS), Research Triangle Park, NC USA (e-mail: stark@niss.org) and also with the Signal Processing Group, Department of Engineering, University of Cambridge, Cambridge, U.K. (e-mail: jas@eng.cam.ac.uk).

Publisher Item Identifier S 1057-7149(00)03918-X.

II. ADAPTIVE HISTOGRAM EQUALIZATION

The AHE process can be understood in different ways. In one perspective the histogram of grey levels (GL's) in a window around each pixel is generated first. The cumulative distribution of GL's, that is the cumulative sum over the histogram, is used to map the input pixel GL's to output GL's. If a pixel has a GL lower than all others in the surrounding window the output is maximally black; if it has the median value in its window the output is 50% grey.

This section proceeds with a concise mathematical description of AHE which can be readily generalized, and then considers the two main types of modification. The relationship between the equations and different (conceptual) perspectives on AHE, such as GL comparison, might not be immediately clear, but generalizations can be expressed far more easily in this framework.

A. Mathematical Description

AHE can be described using few equations. Although the framework that follows has a number of complex details, these are all important for modification and implementation. It is intended to be a summary statement of AHE rather than an extensive exposition.

To equalize an input image x with quantized GL's scaled between $-1/2$ and $1/2$, we first require an estimate $\hat{h}$ of the local histogram. (Some implementations do not actually evaluate any histograms, but can be said to do so implicitly.) We can start by sifting those pixels in the input image with GL g using the Kronecker delta function $\delta(i, j)$, which equals 1 if $i = j$ and 0 otherwise. Spatial convolution with a rectangular kernel f_w can then be used to find the number of such pixels in a window around each point. It is convenient to scale f_w so that it is unit-volume; the estimate histogram then sums to unity at each point. For a square window of width w, with odd-integer value, this can be written

$$\hat{h}(m, n, g) = \delta(g, x(m, n)) \overset{m,n}{*} f_w(m, n) \quad (1)$$

$$f_w(m, n) = \begin{cases} w^{-2}, & |m| \leq (w-1)/2, \quad |n| \leq (w-1)/2 \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

Fig. 5(a) : Original Document Page

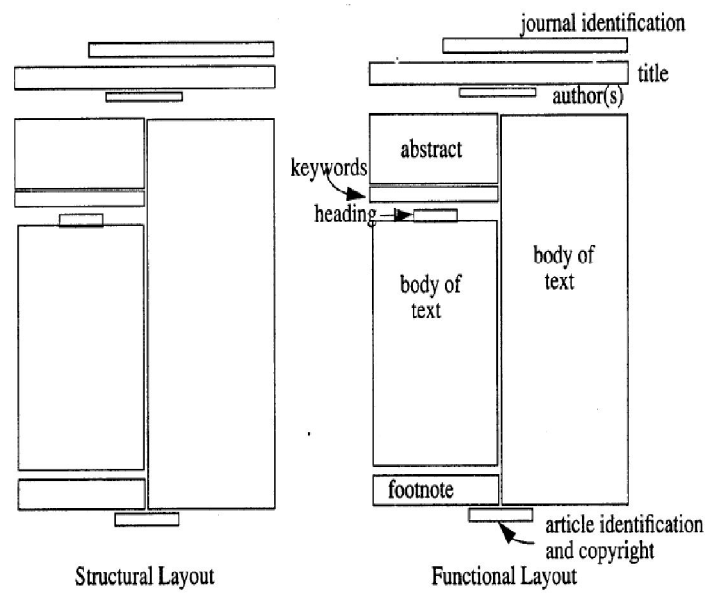


Figure 5 (b) Structural and Functional layout

3.8 Classification

Image classification is a complex process and may be affected by many factors. The classification of document being processed is required for their efficient recognization as it reduces number of searches, easy recognization of document and also reduces the chance of error at different stages during processing. A classifier associates the document with class; labeling an observed document image according to the class, region in to which it falls. The classification stage identifies each input document image by considering the detected features like spatial arrangements with respect to one another, layout of document, size of the document, color of the paper, texture. The categorization (Indexing) of images greatly enhance the performance of document by filtering out the relevant document and the class to which it belongs. Classification of main document is done first followed by the sub-sections. A class prototype is stored in knowledge base .the incoming document is assigned to one of the classes, depending on the value of measure of nearness with the class prototype. This value is obtained by comparison of document under study and the class prototype. The document is assigned to the class with which highest value of the
measurement is obtained.

3.8.1 Document classification use

Document classification is an important task in document Processing.

• Document classification allows the automatic distribution or archiving of documents. For example, after classification of business letters according to sender and message type (such as order, offer, or inquiry), the letters are sent to the appropriate departments for processing.

• Document classification improves indexing efficiency in Digital Library construction. For example, classification of documents into table of contents page or title page can narrow the set of pages from which to extract specific meta-data, such as the title or table of contents of a book.

• Document classification plays an important role in document image retrieval. For example, consider a document image database containing a large heterogeneous Collection of document images. Users have many retrieval demands, such as retrieval of papers from one specific journal, or retrieval of document pages containing tables or graphics. Classification of documents based on visual similarity helps narrow the search and improves retrieval efficiency and accuracy.

• Document classification facilitates higher-level document analysis. Due to the complexity of document understanding, most high-level document analysis systems rely on domain-dependent knowledge to obtain high accuracy. Many available information extraction systems are specially designed for a specific type of document, such as forms processing or postal address processing, to achieve high speed and performance. To process a broad range of documents, it is necessary to classify the documents first, so that a suitable document analysis system for each Specific document type can be adopted.

4. CONCLUSION

The processing of documents for the purpose of discovering knowledge from them in an automated fashion is a challenging task and hence an open issue for the research community. In this article we provide brief summary of basic building blocks that comprise of document image processing system which modifies pictures to improve them (enhancement, restoration), extract information (analysis, recognition), and change their structure (composition, image editing). Today information technology has proved that there is a need to store, query, search and retrieve large amount of electronic information efficiently and accurately. So document image processing is very challenging field of research with the continuous growth of interest and increasing security requirements for the development of the modern society. Sequences of data pre-processing operations are normally applied to the images of the documents in order to put them in a suitable format ready for information extraction.

REFERENCES

[1] Casey, R. G., Wong, K.Y.,(July 1990) "Document Analysis Systems and Techniques, Image Analysis Applications", Image Analysis Applications, pp.1-35.

[2] Castleman, K. R., Digital Image Processing. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1979.

[3] C.C.Chang and D.C. Lin.s, (1996) "A Spatial Data Representation: an Adaptive 2D-H string", Pattern Recognition Letters 17(1996) 175-185, Elsevier.

[4] Claude Faure, Nicorevincent, "Document Image Analysis for Active Reading", International Workshop on Semantically Aware Document Processing and Indexing, ISBN 978-1-59593-668-4, pp 7-14, 2007.

[5] Dr. Mehraj-Ud-Din Dar "Document image classification: A Cognition Based Approach", J&K Science Congress University of Kashmir, 25-27 July, 2006.

[6] Gonzalez, Rafael C. and Woods Richards E., (1999), "Digital Image Processing", Addison Wesley.

[7] Guru, D.S., (2001) "Classification of documents: An overview, the challenges and future avenues", NCDAR, Proceedings of the Pre-conference Workshop on Document Processing, 12th July, Mandya, India.pp28-34.

[8] H. Arai and K. Odaka. Form reading based on background region analysis. In Proceedings of the 4th International Conference on Document Analysis and Recognition. Ulm, Germany, 1997, pp. 164–169.

[9] K.Y.Wong, F.M Wahl, "Document Image Analysis System" IBM journal of research and development, pp 647-656, 1982.

[10]Nawei Chen · Dorothea Blostein, A survey of document image classification: problem statement, classifier architecture and performance evaluation, IJDAR (2007).

[11]O Gorman, L., Kasturi, R., (July 1992), "Document Image Analysis Systems", Computer, 25, pp.5-8.

[12]RANGACHAR KASTURI1, LAWRENCE, and O'GORMAN2, Document image analysis: A primer, S¯adhan¯a Vol. 27, Part 1, February 2002, pp. 3–22.

[13]Samet, H, (1990), "Applications of Spatial Data Structure", Addison-Wesley, Reading,

[14]Sonka Milan, Hlavac and Roger Boyle, (1999), "Image Processing Analysis and Machine Vision", Brooks/Cole Thomson Learning.

[15]T. Young, Gerbrands, "Fundamentals of Image Processing", Paper Back, ISBN 90-756, 9th January 2007.

[16]Ye-In Chang and Hsing-Yen Ann., (1999), "A Note on Adaptive 2D-H Strings", Pattern Recognition Letters 20(1990) 15-20, Elsevier.

[17]Y.Y.Tang and C.Y.Suen, "Document Structure: A Survey", in International Conference on Document Image Analysis and Recognition, pp 99-102, 1993.

ADHOC Network

Shushank Dogra
M.Tech, ECE, Amritsar College of Engineering & Technology, Amritsar

***Abstract*-Deployed in 1990's, Mobile Ad-hoc networks have been widely researched for many years. Mobile Ad-hoc networks are a collection of two or more devices equipped with wireless communications and networking capability. These devices can communication with other nodes that immediately within their radio range or one that is outside their radio range. For the later, the nodes should deploy an intermediate node to be the router to route the packet from the source toward the destination. Recently, the introduction of new technologies such as the Bluetooth, IEEE 802.11 and Hyperlan are helping enable eventual commercial MANET deployments. These recent evolutions have been generating a renewed and growing interest in the research and development of MANET. This paper attempts to provide a comprehensive overview of this dynamic field. It first explains the important role that mobile ad hoc networks play in the evolution of future wireless technologies. Then, it reviews the latest research activities in these areas, including a summary of MANET(Mobile Ad Hoc Network),its characteristic features, Technologies, Applications.**

I. INTRODUCTION

Mobile wireless data communication is advancing both in terms of technology and usage thanks to the Internet and the success of second-generation cellular systems. In the near future, the role and capabilities of short-range data transaction are expected to grow because of already well established and universally accepted technologies such as Bluetooth because these provide effortless and inexpensive deployment of wireless communication.

In terms of price, portability and usability, the computing and communication devices like PDA and mobile phones possess all the characteristics with context to ad hoc networking. As advances in technology continue, these attributes will be enhanced even further. Mobile users can use their cellular phone to check e-mail, browse internet; travellers with portable computers can surf the internet from airports, railway stations, Starbucks and other public locations; tourists can use Global Positioning System (GPS) terminals installed inside rental cars to locate driving maps and tourist attractions, researchers can exchange files and other information by connecting portable computers via wireless LANs while attending conferences; at home, users can synchronize data and transfer files between portable devices and desktops.

II. EVOLUTION OF AD HOC NETWORK

The roots of ad hoc networking can be traced back as far as 1968, when work on the ALOHA network was initiated (the objective of this network was to connect educational facilities in Hawaii).1 Although fixed stations were employed, the ALOHA protocol lent itself to distributed channel access management and hence provided a basis for the subsequent development of distributed channel-access schemes that were suitable for ad hoc networking. The ALOHA protocol itself was a single-hop protocol that is, it did not inherently support routing. Instead every node had to be within reach of all other participating nodes.Inspired by the ALOHA network and the early development of fixed network packet switching, DARPA began work, in 1973, on the PRnet (packet radio network). Although many experimental packet radio networks were later developed, these wireless systems did not ever really take off in the consumer segment. When developing IEEE 802.11 standard for wireless local area networks (WLAN) the Institute of Electrical and Electronic Engineering (IEEE) replaced the term packet-radio network with ad hoc network.

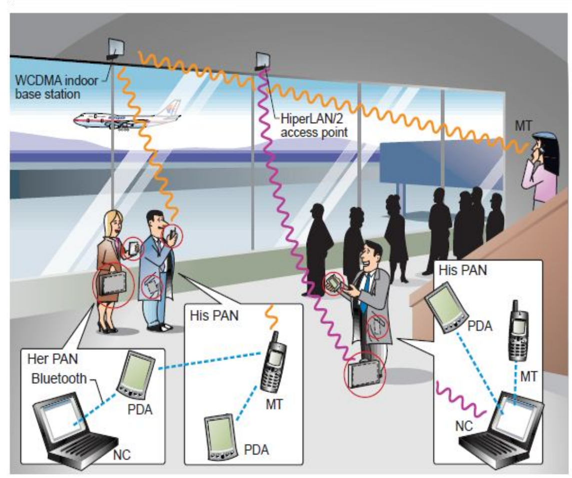


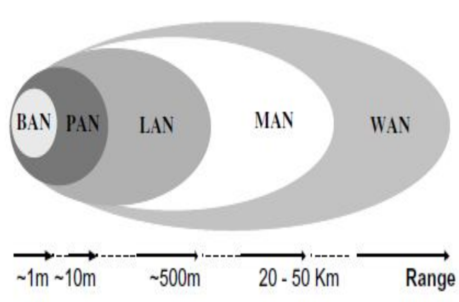
Figure: At an airport, people can access LAN and WAN through ad hoc bluetooth connection using devices with Bluetooth technology.

A mobile ad hoc network is a network formed without any central administration which consists of mobile nodes that use a wireless interface to send packet data. Since the nodes in a network of this kind can serve as routers and hosts, they can forward packets on behalf of other nodes and run user applications. Today, our vision of ad hoc networking includes scenarios such as those depicted in Figure below where people carry devices that can network on an ad hoc basis. A user device can both interconnect with one another and connect to local information points for example, to retrieve updates on flight departures, gate changes, and so on. The ad hoc devices can also relay traffic between devices that are out of range. Present day cellular systems rely heavily on infrastructure. Coverage is provided by base stations, radio resources and services are integrated into the sytem.

Fully decentralized radio, access, and routing technologies enabled by Bluetooth, IEEE 802.11 ad hoc mode, PRnet stationless mode, mobile ad hoc network (MANET), and concepts such as the personal area network (PAN) or PAN-to-PAN Communication fit more or less entirely into the ad hoc domain. The MANET initiative by the Internet Engineering Task Force (IETF) also aims to provide services via fixed infrastructure connected to the Internet.

Technologies Used In Ad hoc Networking

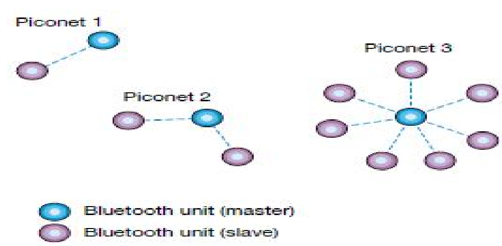
As shown in Fig., we can classify ad hoc networks, depending on their coverage area, into several classes: Body (BAN), Personal (PAN), Local (LAN),Metropolitan (MAN) and Wide (WAN) area networks. Wide and Metropolitan-area ad hoc networks are mobile multi-hop wireless networks that present many challenges that are still to be solved (e.g., addressing, routing, location management, security, etc.), and their availability is not on immediate horizon. On the other hand, mobile ad hoc networks with smaller coverage can be expected to appear soon. Specifically, ad-hoc single hop BAN, PAN and LAN wireless technologies are already common on the market, these technologies constituting the building blocks for constructing small, multi-hop, ad hoc networks that extend their range over multiple radio hops. For these reasons, BAN, PAN and LAN technologies constitute the technologies used in this networking.



BLUETOOTH

Worldwide, the industry has shown tremendous interest in techniques that provide short-range wireless connectivity. Bluetooth technology can be used to operate ad hoc networks. The main purpose of Bluetooth is to replace cables between electronic devices, such as telephones, PDAs, laptop computers, digital cameras, printers, and fax machines, by using a low-cost radio chip. The Bluetooth system can manage a small number of low-cost point-to-point, and point to multi-point communication links over a distance of up to 10 m with a transmit power of less than 1mW. It Operates in the globally available unlicensed ISM (industrial, scientific, medical) frequency band at 2.4 GHz.

Into a Bluetooth network, one station has the role of master, and all other Bluetooth stations are slaves. The master decides which slave is the one to have the access to the channel. A slave is authorized to deliver a single packet to the master only if it has received a polling message from the master. The units that share the same channel (i.e., are synchronized to the same master) form a piconet, the fundamental building block of a Bluetooth network.

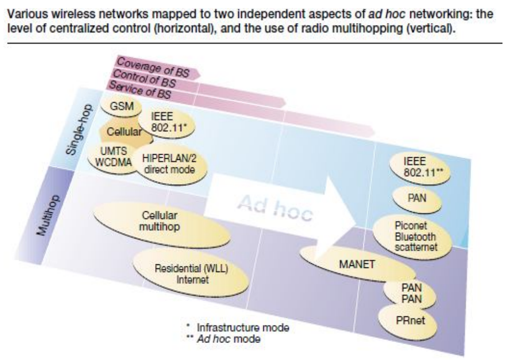


A piconet has a bit rate of 1 Mbit/s that represents the channel capacity including the overhead introduced by the adopted protocols, and polling scheme. A piconet contains a master station, and up to seven active (i.e., participate in data exchanging) slaves, contemporarily.

III. IEEE 802.11(Wi-Fi)

The IEEE 802.11 specification is a wireless LAN standard that specifies a wireless interface between a client and a base station or access point, as well as between wireless clients. IEEE 802.11 defines two physical characteristics for radio-based wireless local area networks: direct-sequence spread spectrum (DSSS), and frequency-hopping spread spectrum (FHSS), both of which operate on the2.4 GHz ISM band.

Two network architecture modes have been defined in the IEEE 802.11 standard, namely the point coordination function (PCF) mode and the distributed coordination function (DCF) mode. The former use centralized approach in which a network access point controls all traffic in the network, The DCF mode supports direct communication between wireless clients.



The IEEE 802.11 standard defines two operational modes for WLANs: infrastructure-based and infrastructure-less or ad hoc. Network interface cards can be set to work in either of these modes but not in both simultaneously. Infrastructure mode resembles cellular infrastructure-based networks. It is the mode commonly used to construct the so-called Wi-Fi hotspots, i.e., to provide wireless access to the Internet. In the ad hoc mode, any station that is within the transmission range of any other, after a synchronization phase, can start communicating. No AP is required, but if one of the stations operating in the ad hoc mode has a connection also to a wired network, stations forming the ad hoc network gain wireless access to the Internet.

IV. AD HOC NETWORKING ISSUES

Mobile ad hoc networks are formed dynamically by an autonomous system of mobile nodes that are connected via wireless links without using the existing network infrastructure or centralized administration. The nodes are free to move randomly and organize themselves arbitrarily; thus, the networks wireless topology may change rapidly and unpredictably. Such a network may operate in a standalone fashion, or may be connected to the larger Internet. Mobile ad hoc networks are infrastructure-less networks since they do not require any fixed infrastructure, such as a base station, for their operation. In general, routes between nodes in an ad hoc network may include multiple hops, and hence it is appropriate to call such networks as ‘‘multi-hop wireless ad hoc networks’’. Each node will be able to communicate directly with any other node that resides within its transmission range.

Ad hoc wireless networks inherit the traditional problems of wireless communications and wireless networking :

• The wireless medium has neither absolute, nor readily observable boundaries outside of which stations are known to be unable to receive network frames.

• The channel is unprotected from outside signals.

• The wireless medium is significantly less reliable than wired media.

• The channel has time-varying and asymmetric propagation properties.

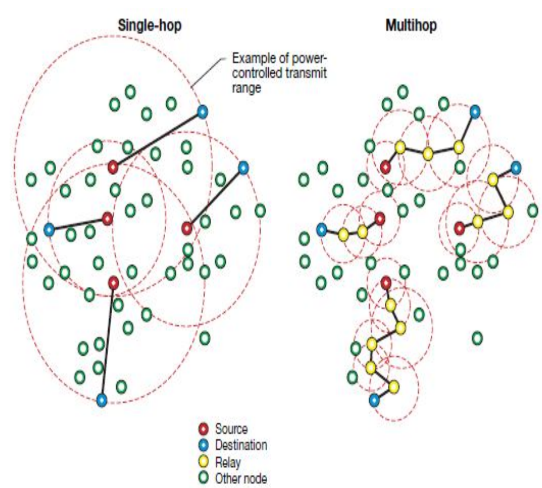
• Hidden-terminal and exposed-terminal phenomena may occur.

V. ROUTING IN AD HOC NETWORKS

For mobile ad hoc networks, the issue of routing packets between any pair of nodes becomes a challenging task because the nodes can move randomly within the network. A path that was considered optimal at a given point in time might not work at all a few moments later.

Traditional routing protocols are proactive in that they maintain routes to all nodes, including nodes to which no packets are being sent. They react to any change in the topology even if no traffic is affected by the change, and they require periodic control messages to maintain routes to every node in the network. The rate at which these control messages are sent must reflect the dynamics of the network in order to maintain valid routes. Thus, scarce resources such as power and link bandwidth will be used more frequently for control traffic as node mobility increases.

Multi-Hop Routing**:** Moving information through a network from a source to a destination are not with in mutual transmission range



Reliabiltily**:** Nodes in an adhoc network are not 100% reliable and algorithms need to find alternate routes when nodes are failing**.**

VI. CHARACTERISTICS AND REQUIREMENTS

In contrast to traditional wireline or wireless networks, an ad hoc network could be expected to operate in a network environment in which some or all the nodes are mobile. In this dynamic environment, the network functions must run in a distributed fashion, since nodes might suddenly disappear from, or show up in, the network.
Below, we discuss some typical operational characteristics and how they affect the requirements for related networking functions:

Distributed operation: A node in an ad hoc network cannot rely on a network in the background to support security and routing functions. Instead these functions must be designed so that they can operate with efficiently under distributed conditions.

Dynamic network topology**:** In general, the nodes will be mobile, which sooner or later will result in a varying network topology.

Fluctuating link capacity**:** The effects of high bit-error rates might be more profound in a multi hop ad hoc network, since the aggregate of all link errors is what affects a multi hop path.

Low-power devices**:** In many cases, the network nodes will be battery-driven, which will make the power budget tight for all the power-consuming components in a device.

VII. TYPICAL APPLICATIONS

Mobile ad hoc networks have been the focus of many recent research and development efforts. So far, ad hoc packet-radio networks have mainly been considered for military applications, where a decentralized network configuration is an operative advantage or even a necessity. In the commercial sector, equipment for wireless, mobile computing has not been available at a price attractive to large markets. However, as the capacity of mobile computers increases steadily, the need for unlimited networking is also expected to rise. Commercial ad hoc networks could be used in situations where no infrastructure is available.
Ad hoc networking could also serve as wireless public access in urban areas, providing quick deployment and extended coverage. Some of the access points would also provide gateways via which users might connect to a fixed backbone network. At the local level, *ad hoc* networks that link notebook or palmtop computers could be used to spread and share information among participants at a conference. They might also be usedin home networks where devices can communicate directly to exchange information, such as audio/video, alarms, and configuration updates. Short-range *ad hoc* networks can simplify intercommunication between various mobile devices (such as a cellular phone and a PDA)

CONCLUSION

In this article we have tried to survey ad hoc networking mainly from a technical point of view. We have also made an attempt to clarify what an ad hoc network actually is and found that the definitions vary. Typically, these networks operate with distributed functions and allow traffic to pass over multiple radio hops between source and destination. Furthermore, we have discussed some of the typical characteristics of and technologies used, by ad hoc networks. In any case, the most attractive property of an ad hoc networking model is perhaps its independence from centralized control and, thus, the increased freedom and flexibility it gives the user. Due to its inherent flexibility, ad hoc networking is easy to deploy and would fit nicely into, say, an office setting, where users could set up ad hoc networking groups using fewer LAN access points and potentially less transmitting power. However, the products that apply the concepts of ad hoc network would work in personal area range (Network) or PAN. The ad hoc network functionality will also enable the interconnection of different user devices. Ability to create generic, small-scale, ad hoc networks in portable devices represents an entirely new area for future ad hoc-based applications. In coming years, mobile computing will keep on extending, so Ad hoc networking is at the center of the evolution towards the 4th generation wireless technology. Its flexibility, ease of maintenance, lack of required infrastructure, auto-configuration, self administration capabilities, and significant costs advantages make it a prime candidate for becoming the most widely accepted technology for personal communication.

REFERENCES

1. N. Abramsson ÒThe ALOHA system another alternative for computer communications
in AFIPS Conf. Proc., vol 37, FJCC, 1970, pp. 695-702.
2. Y.D.Lin and Y.C.Hsu, ÒMultihop Cellular: A new architecture for wireless communications in IEEE INFOCOM 2000, pp. 1273-1282.
3. Mobile *Ad hoc* Networks (MANET). URL: http://www.ietf.org/html.charters/manetcharter.
html. (2000-05-28). Work in progress.
4. What’s Behind Ricochet: A Network Overview, http://www.ricochet.net/ricochet_advantage/tech_overview.
5. Charles E. Perkins. Mobile IP. IEEE Communications Magazine, pages 84D99, May 1997.
6. Haartsen, J.: Bluetooth - The universal radio interface for *ad hoc*, wireless connectivity. Ericsson Review Vol. 75(1998):3, pp. 110-117.
7. Per Johansson et al. Short-Range Radio-Based *Ad hoc* Networking: Performance and Properties ICC99, Vancouver, 1999.
8. IEEE Computer Society LAN MAN Standards Committee, Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Speci-

fications, IEEE Std 802.11-1997. The Institute of Electrical and Electronics Engineers, New York.
9. G. Anastasi, E. Borgia, M. Conti, E. Gregori, IEEE 802.11 ad hoc networks: performance measurements, in: Proceedings of the Workshop on Mobile and Wireless Networks (MWN 2003) in conjunction with ICDCS 2003, May 19, 2003.
10. Elizabeth Belding-Royer, Routing approaches in mobile ad hoc networks, in: S. Basagni, M. Conti, S. Giordano, I. Stojmenovic (Eds.), Ad Hoc Networking, IEEE Press Wiley, New York, 2003.

Performance Analysis of 32 Channel DWDM Metro System at High Bit Rate and Power using MetroCor and SMF-28 Fibers

Atul Mahajan [1], Anu Sheetal[2], Jaswinder Singh[3]
[1]Lecturer, ACET, Amritsar, Punjab, India.
[2]Assistant Professor, GNDU Regional Campus, Gurdaspur, Punjab, India
[3]Associate Professor, BCET, Gurdaspur, Punjab, India

Abstract-In this paper, we evaluate the performance of 32 channels DWDM Metro system for high Bit rates and power with 100GHz channel spacing using Directly Modulated Lasers (DML)and the Negative Dispersion Fibre (NDF) such as MetroCor and a positive dispersion fiber such as SMF-28. Simulative analysis has been performed by varying the bit rate from 2Gb/s to 5Gb/s and power from 0dBm to 5dBm. Furthermore, an optimum power level has been achieved at 3dBm. Results are obtained in the form of Eye diagrams, BER and input-output spectra's for SMF-28 and Metrocor fibers.

Keywords: DML-2, NDF, MetroCor, SMF-28, RZ data formats.

1. INTRODUCTION

Recently, in order to satisfy the growing demand for bandwidth in the access network, therehas been increased interest in low-cost DWDM systems for use in especially metropolitan area networks. The characteristics of the cost sensitive metropolitan area networks in terms of the transmission distance and bit rate are not so demanding compared to the long haul networks Low-cost directly modulated lasers (DML) recently attracted much attention for use as transmitters in 2.5-Gb/s metro area applications. However, DMLs have some major drawbacks. The output power waveform is not an exact replica of the modulation current and the instantaneous optical frequency varies with time depending on the change of the optical power (an effect also known as frequency chirp). The interaction of the positive chirp with the positive dispersion of conventional standard single-mode fibers (like Corning SMF-28 fiber or any other fiber with similar dispersion characteristics) deteriorates the optical signal and limits the maximum achievable transmission distance[1]. A nonzero dispersion-shifted fiber (NZDSF) with negative dispersion (Corning MetroCor fiber) was introduced to take advantage of the positive chirp characteristics of low-cost transmitters that are likely to be used in metro area networks.This dispersion-optimized optical fiber can eliminate the need for dispersion compensation and will enable the introduction of DWDM technologies in the metro environment [1].

Horche et.al [4] showed that the effect of directly modulated laser chirp can be compensated by a negative dispersion fibre, but that only occurs at specific range of DML output and pulse broadened by the positive dispersion can be equalized using SPM in optical fibre.

Chung et.al [5] demonstrated the error free transmission of DMLs at 2.5,10 and 40 Gb/s Coarse Wavelength Division Multiplexing (CWDM) and DWDM systems over negative dispersion fibres.

Sheetal et.al [6] presented the simulative analysis of 40Gb/s long haul (500-2000km) DWDM system with a ultra high capacity for carrier suppressed return to zero (CSRZ), Duobinary Return-to-zero (DRZ) and Modified DRZ (MDRZ).The DWDM system had been analysed for the symmetrical dispersion compensation schemes for 16 channels with 25 GHz channel spacing.

Yao.e.al.[12] discussed the exponential growth in data, IP, Traffic in MAN and in order to overcome the problems of increasing traffic demands in metropolitan area networks an optical packet switched network was designed .But in order to compete with high speed IP routers the packet switch should operate beyond 10Gb/s.Optical regeneration is necessary as bit rate approaches 40Gb/s.

However, simulative analysis of the DML in the DWDM metro systems for 32 channels with more than 2.5Gb/s and 0dBm without using any external modulator is not available as such in literature. Moreover most of the simulations in literature are at 0dBm, no optimum power level has been introduced in the literature. In section 2, the system description and simulation Parameters have been described. In section 3, comparison of results of the simulated systems has been reported and finally in section 4, conclusions are made.

2. SYSTEM DESCRIPTION AND SIMULATION

Here, we present the simulation setup for 32 channel DWDM metro system which is shown in Figure 1. Each channel of the transmitter section consists of Pseudo Random Bit Sequence (PRBS) followed by NRZ pulse generator, DML-2 as laser source and WDM multiplexer. The logical sequence generated by PRBS at 2.5Gb/s is converted to electrical signal using NRZ electrical pulse generator. The rise and fall time of NRZ is taken to be 0.5 bit with amplitude of 1a.u. Here, 32 DML-2 sources used have frequency range = 192.4 - 195.5THz with channel spacing = 100GHz within the channels and input power

= 1mW for each laser source. All the 32 channels are fed to WDM multiplexer operating at 1550nm with a bandwidth of 10GHz and channel spacing of 100GHz.

The optical channel consists of MetroCor fibre with attenuation = 0.25dB/km, Dispersion = -5.6 ps/nm/km, Dispersion slope = 0.12ps/km-nm^2, Effective core area(A_{eff}) = 50µm^2 and Non linear refractive index n_2 = 2.6x10^{-20} m^2/W. An Erbium Doped Fibre Amplifier (EDFA) has been used with a Gain = 26dB and Noise Figure = 4dB. The fiber parameters of MetroCor and SMF-28 fibers are given in Table 1 respectively.

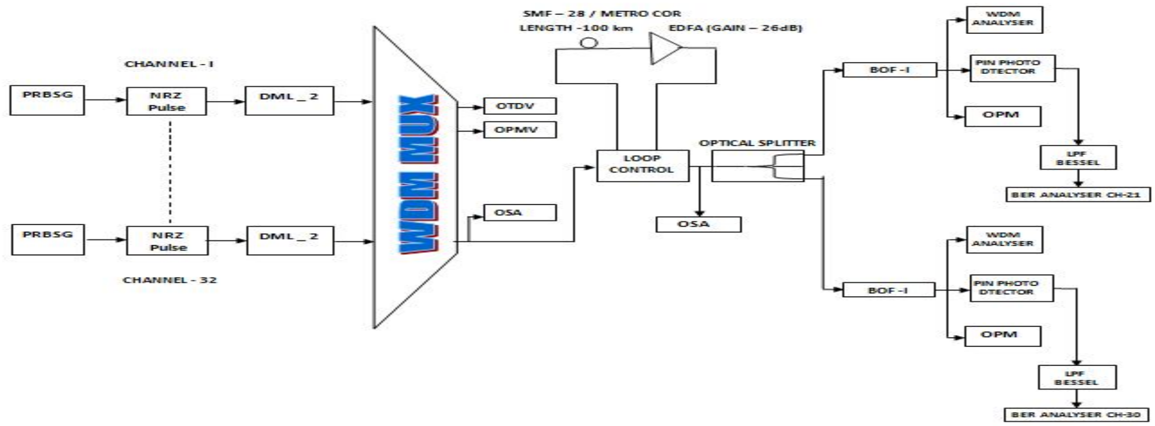


Figure **1.** Schematic of Simulation Setup

Table 1. Fibre Parameters

Fibre	Reference wavelength (nm)	Attenuation (dB/km)	Dispersion (ps/nm/km)	Dispersion Slope (ps/km-nm^2)	Effective core area A_{eff}(µm^2)	Non linear refractive index n_2 (10^{-20} m^2/W)
MetroCor	1550	0.25	-5.6	0.12	50	2.6
SMF-28	1550	0.25	16.75	0.075	80	2.6

3. RESULTS AND DISCUSSION

To estimate the performance, the BER, Q value [dB] and the eye diagrams for the simulative model has been considered. Figure 2(a) & (b) show the graphical representation of BER value as a function of Power [dBm] and Bit rate[Gb/s] respectively for two different optical fibers. The graphs clearly depicts that the performance of MetroCor is superior to that of the SMF-28.

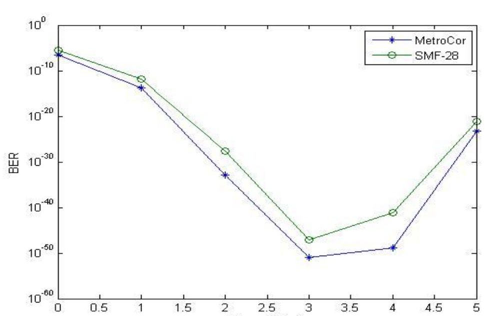


Figure 2(a). BER vs. Power for MetroCor and SMF-28 Fiber

The graph 2(a) clearly shows that at the power of 3dBm we achieve a minimum value of BER i.e 8.94e^{-52}for the MetroCorfiber and 7.88e^{-48} for the SMF-28 is achieved. It is seen that for the increasing power the performance of the MetroCor predominates the SMF-28 fiber. Further, it is seen that with increase of power from 0dBm to 5dBm the BER achieves its minimum value at 3dBm beyond which the value of BER starts increasing.

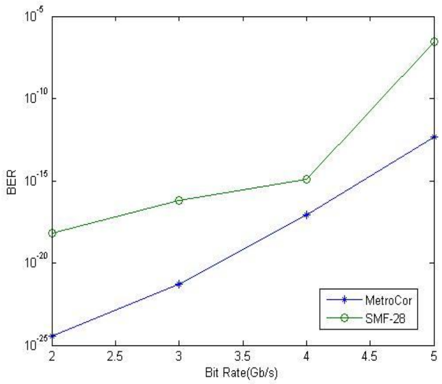


Figure 2(b). BER vs. Bit Rate for MetroCor and SMF-28 fiber

Similarly, in graph 2(b) with the increasing bit rates the performance of the SMF-28 deteriorates as compared to the MetroCor fiber. The Eye diagrams along with their input and output spectras for the two fibers for the varying power and bit rate are shown below.

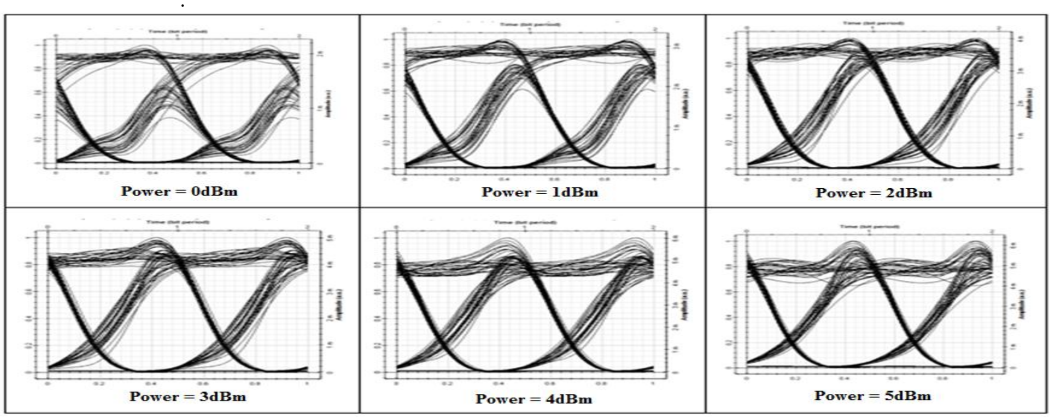


Figure 3(a).Eye Diagrams for the MetroCor Fiber by varying Input Power from 0dBm to 5dBm.

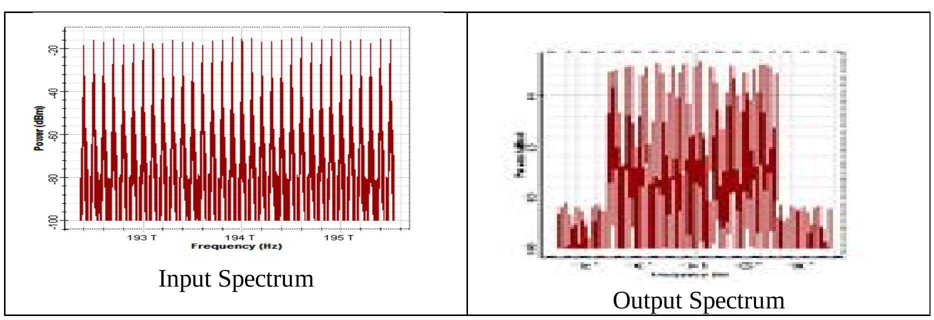
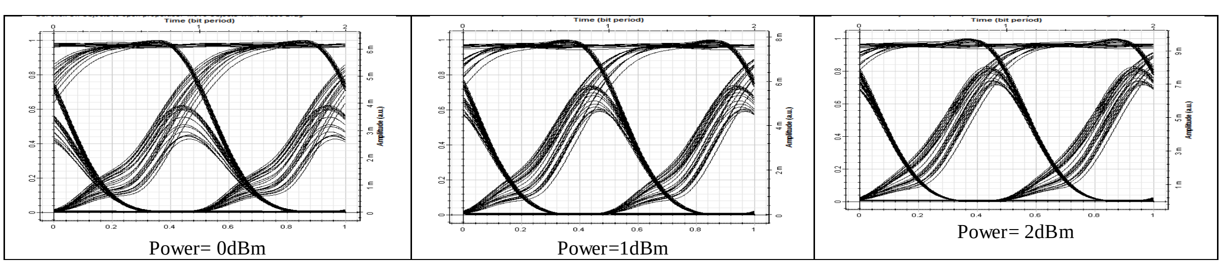


Figure 3(b) Input and output Spectrums for the MetroCor Fiber for the varying Input Power



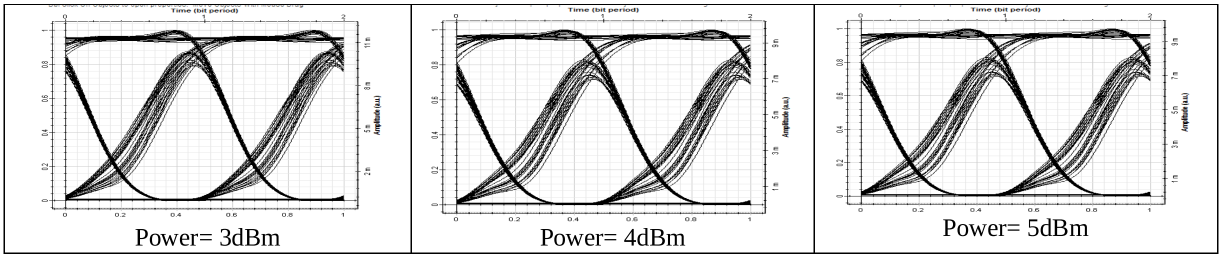


Figure 4(a).Eye Diagrams for the SMF-28 Fiber for the varying Input Power from 0dBm to 5dBm

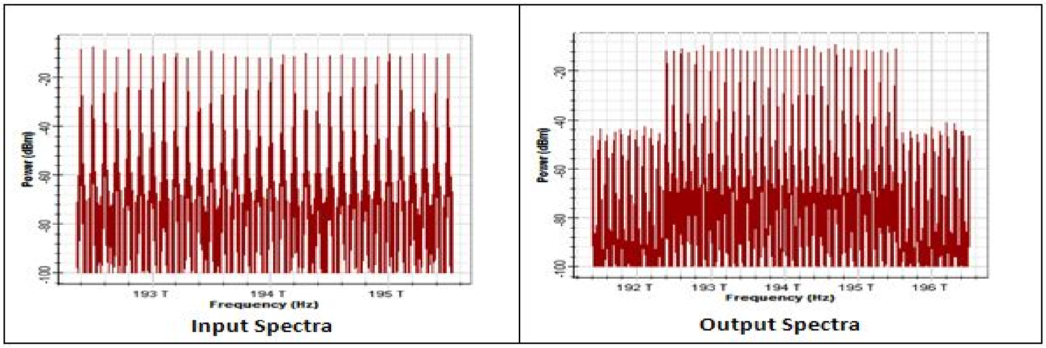


Figure 4(b) Input and output Spectrums for the SMF-28 Fiber for the varying Input Power

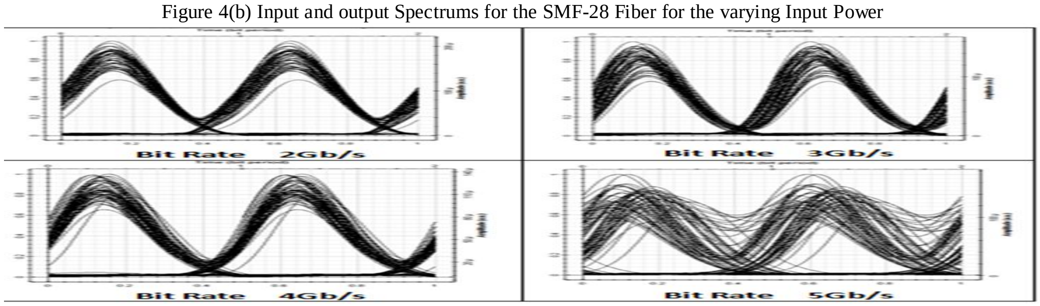


Figure 5.Eye Diagrams for the MetroCor Fiber for the varying Bit Rate from 2Gb/s to 5Gb/s

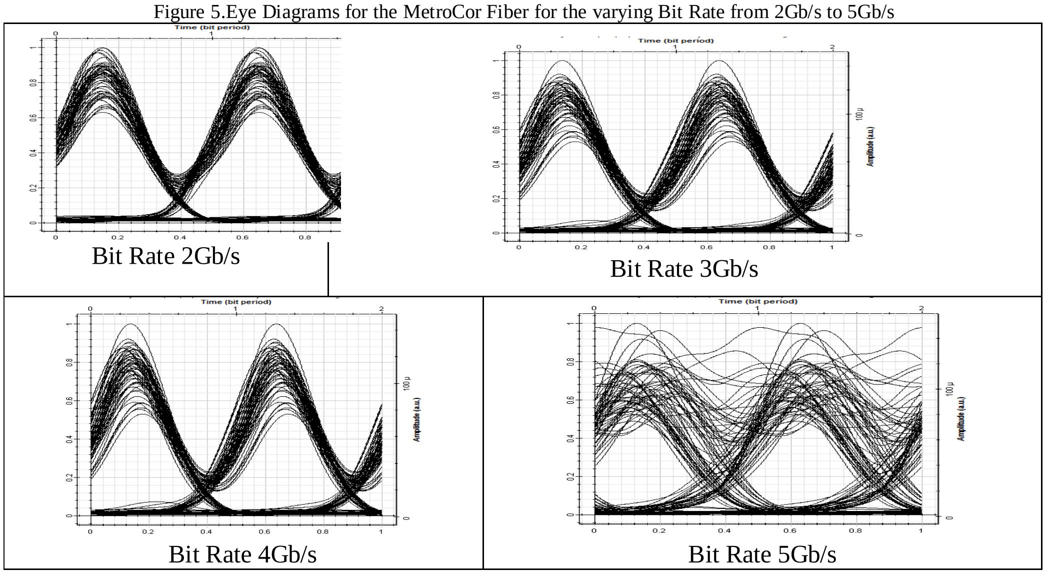


Figure 6.Eye Diagrams of SMF-28 Fiber for the varying Bit Rate from 2Gb/s to 5Gb/s

The output optical spectrums in Fig.3(b) and 4(b) for the DM laser sources show large number of spurious signals, due to the Interchannel Four Wave Mixing (IFWM) effect the initial signal bandwidth has been expanded after the transmission through a nonlinear optical fibre. The wave frequencies interacting through the FWM leads to the generation of frequencies known as spuriouswaves[6]

CONCLUSIONS

We have simulated 32 channel DWDM metro systems with 100GHz channel spacing for the high bit rate and varying input power for the NDF MetroCor and SMF-28 which is a positive dispersion fiber. It is seen that at high bit rates, the performance of the MetroCor predominates the SMF-28. For both the fibers DM Lasers are used upto 5Gb/s, but further increase in the bit rate for such a system an external modulator is required. The optimum input power of 3dBm has been achieved for faithful transmission using both the fibers. Also, it is concluded that the performance of MetroCor is superior as compared to that of SMF-28.

REFERENCES

[1] I. Tomkos et. al., "Demonstration of negative dispersion fibres for DWDM metropolitan networks", IEEE J. of Select. Top.inQuan. Elec. 7, 2001.

[2] Yao. et.al., "All-Optical Packet Switching for Metropolitan Area Networks Opportunities and Challenges",IEEE Communication Magazine,March 2001.

[3] Chung et.al,"Directly modulated CWDM/DWDM system using negative dispersion fibre for metro network application," Elsevier Science ,October 2004.

[4] Furukawa.et.al., "Field trial of 160 Gbit/s DWDM-based optical packet switching and transmission",Optical Society of America,2008.

[5] Horche et.al ",Enhanced performance of WDM system using directly modulated lasers on positive dispersion fibres", Science Direct/Elsevier/November 2007.

[6] Sheetal.et.al. "Simulation of High Capacity 40Gb/s Long Haul DWDM System Using Different Modulation Formats and DispersioCompensation Schemes in the Presence of Kerr's Effect". Science Direct/Elsevier/November 2008.

[7] Tomkos.et.al. "Transport performance of an 80-Gb/s WDM regional area transparent ring network utilizing directly modulated lasers" Lightwave Technology Journal,August 2002

[8] Hakki.et.al. "Evaluation of transmission characteristics of chirped DFB lasers in dispersive optical fiber" Lightwave Technology Journal,August 2002.

[9] Tomkos.et.al. "Extraction of laser rate equation parameters for representative simulations of metropolitan area transmission system and network" Optics Communication /Elsevier /April 2001.

[10] Sharma.et.al. "Enhancing the performance of Directly Modulated Laser system using negative dispersion fiber for metro applications"Corning Incorporations

[11] G. P. Aggrawal, Fibre Optic Communication Systems. New York:Wiley, 1992

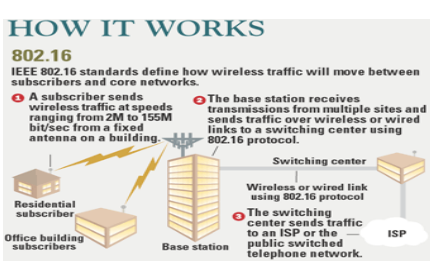
WiMAX

Tanvir Singh Buttar
M.Tech, ECE, Amritsar College of Engineering and Technology, Amritsar
Tanvir_buttar@ymail.com

Abstract: **- WiMAX is a wireless digital communications system, also known as IEEE 802.16, that is intended for wireless "metropolitan area networks". WiMAX (Worldwide Interoperability for Microwave Access) is a wireless communications standard designed to provide 30 to 40 megabit-per-second data rates,with the 2011 update providing up to 1 Gbit/s for fixed stations. It is a part of a "fourth generation," or 4G, of wireless-communication technology. WiMax far surpasses the 30-metre wireless range of a conventional Wi-Fi local area network (LAN), offering a metropolitan area network with a signal radius of about 50 km.This paper intends to give an overview over WiMAX.I make this paper with the of project report based on wimax.**

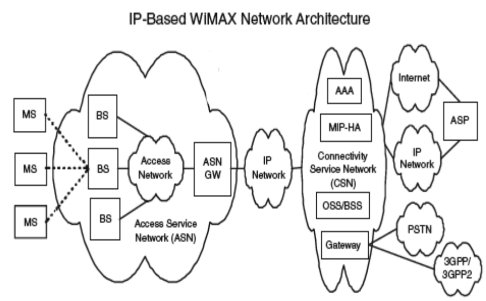
INTRODUCTION

The Mobile WiMAX (Worldwide Interoperability for Microwave Access) standard of 802.16e is divergent from Fixed WiMAX. It attracted a significant number of Forum members towards an opportunity to substantively challenge existing 3G technology purveyors. While clearly basedon the same OFDM base technology adopted in 802.16-2004, the 802.16e version is designed to deliver service across many more sub-channels. The 802.16e standard adds OFDMA 2KFFT, 512-FFT and 128-FFT capability. Sub-channelization facilitates access at varying distance by providing operators the capability to dynamically reduce the number of channels while increasing the gain of signal to each channel in order to reach customers farther away. The reverse is also possible. WiMAX systems are based on the IEEE 802.16-2004 and IEEE 802.16e-2005standards which define a physical (PHY) layer and the medium access control (MAC) layer for broadband wireless access systems operating at frequencies below 11 GHz. The first of these standards, published in 2004, addresses fixed services, and the second, published in 2005, is intended for mobile services.



Arcitecture

The IEEE 802.16e-2005 standard provides the air interface for WiMAX but does not define the full end-to-end WiMAX network. The WiMAX Forum's Network Working Group (NWG), is responsible for developing the end-to-end network requirements, architecture, and protocols for WiMAX, using IEEE 802.16e-2005 as the air interface.TheWiMAX NWG has developed a network reference model to serve as an architecture framework for WiMAX deployments and to ensure interoperability among various WiMAX equipment and operators.The network reference model envisions a unified network architecture for supporting fixed, nomadic, and mobile deployments and is based on an IP service model. Below is simplified illustration of an IP-based WiMAX network architecture. The overall network may be logically divided into three parts:



1.Mobile Stations (MS) used by the end user to access the network.

2.The access service network (ASN), which comprises one or more base stations and one or more ASN gateways that form the radio access network at the edge.

3.Connectivity service network (CSN), which provides IP connectivity and all the IP core network functions.

The network reference model developed by the WiMAX Forum NWG defines a number of functional entities and interfaces between those entities. Fig below shows some of the more important functional entities.

Base station (BS): The BS is responsible for providing the air interface to the MS. Additional functions that may be part of the BS are micromobility management functions, such as handoff triggering and tunnel establishment, radio resource management, QoS policy enforcement, traffic classification, DHCP (Dynamic Host Control Protocol) proxy, key management, session management, and multicast group management.

Access service network gateway (ASN-GW): The ASN gateway typically acts as a layer 2 traffic aggregation point within an ASN. Additional functions that may be part of the ASN gateway include intra-ASN location management and paging, radio resource management and admission control, caching of subscriber profiles and encryption keys, AAA client functionality, establishment and management of mobility tunnel with base stations, QoS and policy enforcement, foreign agent functionality for mobile IP, and routing to the selected CSN.

Connectivity service network (CSN): The CSN provides connectivity to the Internet, ASP, other public networks, and corporate networks. The CSN is owned by the NSP and includes AAA servers that support authentication for the devices, users, and specific services. The CSN also provides per user policy management of QoS and security. The CSN is also responsible for IP address management, support for roaming between different NSPs, location management between ASNs, and mobility and roaming between ASNs.

COMPARISON OF MOBILE WIMAX WITH ITS COMPETING WIRELESS TECHNOLOGIES

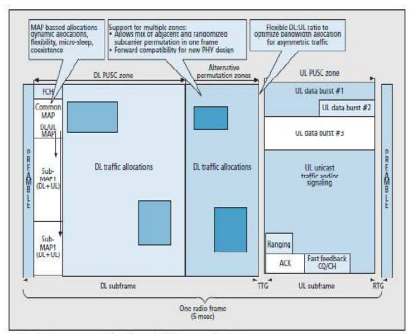
A Comparison of Wimax & 3G.

In this section WiMAX is compared with other wireless technologies. According to several industry sources, the key features of Mobile Wimax arethat it uses OFDMA, MIMO, Beamforming and a number of other recent technology advancements that are labeled as features in 4G. It supports several new features necessary for delivering mobile broadband services at vehicular speeds greater than 120 km/hr with QoS. Some of WiMAX key new features and benefits over other wireless technologies are:

1. Introduces OFDMA, which improves spectrum efficiency (amount byte transferred on given width of frequency) around two times more than current 3G technologies or Wi- Fi. Wimax only need about half of the base station as would for HSPA.
2. Enables a wide range of advanced antenna systems including MIMO, beam-forming, Spacetimecoding and spatial multiplexing. It thus increases the covering range of Wimax; it also candynamically allocate frequency band (from 1.5 to 20 MHz) based onUser's signal strength, bandwidth requirement. By this it makes better use of availablefrequency to support more users, so have better spectral efficiency.
3. Dynamic Power Conservation Management ensures power efficient operation of battery operated mobile handheld and portable devices in Sleep and Idle modes. This may be critical for small devices like cell phones.
4. With 5 millisecond latency between hand hold devices and cellular tower, plus the support of QoS, make Wimax good for high quality VOIP, this wireless data network also competes with 2G and 3G on voice service.
5. Wimax is an open standard, which means there will be no or very little royalty. This is one of the biggest advantages of Wimax.

FRAME STRUCTURE AND PHYSICAL CHANNELIZATION

The IEEE802.16e PHY supports both TDD and FDD operation. The FDD mode also defines ahalf duplex FDD mode to support lower-complexity terminals in which one radio front unit is time-shared between UL(uplink) and DL(downlink). While Mobile WiMAX Release 1.0 includes only the TDD profile, in Release 1.5 both TDD and FDD systems are supported. The OFDMA frame structure for TDD mode, where each 5 ms radio frame is flexibly divided into DL and UL sub frames



Frame structure and channelization for TDD system in release The DL and UL sub frames are separated by small transmit/receive and receive/transmit transition gaps (TTG and RTG, respectively) to prevent DL and UL transmission collisions. This frame structure defines the following physical channels:

• Preamble: broadcast in the first orthogonal frequency-division multiplexed (OFDM) symbol of the frame in DL and used by the MS initial and handover related scanning as well as PHY synchronization with the BS.

• Frame control header (FCH): follows the preamble and provides the frame configuration information, such as MAP message length and coding scheme and usable sub channels.

• DL-MAP and UL-MAP: provide resource allocation and other control information for the DL and UL sub frames, respectively. The MAP is typically broadcast across the cell using a robust modulation and coding scheme (MCS). To reduce the MAP overhead, the system may also define one or more multicast sub-MAPs that can carry traffic allocation messages at higher MCS levels for users closer to the BS and with higher CINR conditions.

•UL ranging: The UL ranging sub channel is allocated for an MS to perform closed-loop time, frequency, andpower adjustment as well as bandwidth requests**.**

•UL CQICH: The UL CQICH channel is allocated for the MS to feedback channel state information.

•UL ACK: The UL ACK is allocated for the MS to feedback DL HARQ ACKs.

COMPARISION WITH WIFI

Confusion between WiMAX and Wi-Fi are frequent because both are related to wireless connectivity and Internet access.

•WiMAX is a long range system, covering many kilometres, that uses licensed or unlicensed spectrum to deliver connection to a network, in most cases the Internet.

•Wi-Fi uses unlicensed spectrum to provide access to a local network.

•Wi-Fi is more popular in end user devices.

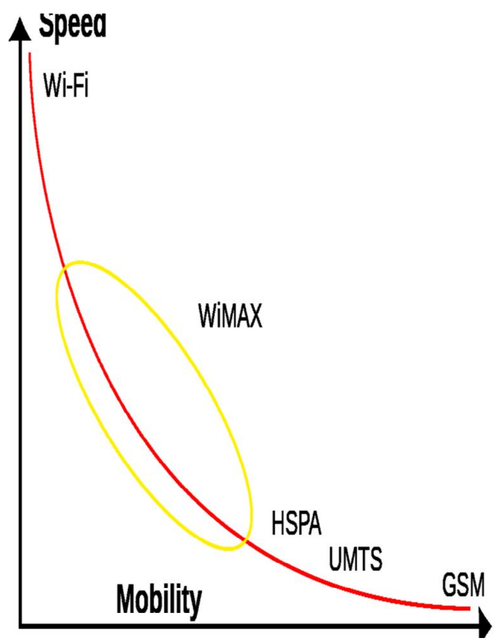
•Wi-Fi runs on the Media Access Control's CSMA/CA protocol, which is connectionless and contention based, whereas WiMAX runs a connection-oriented MAC.

•WiMAX and Wi-Fi have quite different quality of service (QoS) mechanisms:

•WiMAX uses a QoS mechanism based on connections between the base station and the user device. Each connection is based on specific scheduling algorithms.

•Wi-Fi uses contention access - all subscriber stations that wish to pass data through a wireless access point (AP) are competing for the AP's attention on a random interrupt basis. This can cause subscriber stations distant from the AP to be repeatedly interrupted by closer stations, greatly reducing their throughput.

•Both 802.11 (which includes Wi-Fi) and 802.16 (which includes WiMAX) define Peer-to-Peer (P2P) and ad hoc networks, where an end user communicates to users or servers on another Local Area Network (LAN) using its access point or base station. However, 802.11 supports also direct ad hoc or peer to peer networking between end user devices without an access point while 802.16 end user devices must be in range of the base station.



Speed vs. mobility of wireless systems:

QUALITY OF SERVICE OF MOBILE WIMAX

The MAC layer of the WiMAX architecture is responsible for Qos. Sub channelization anddifferent coding schemes enable end-to-end QoS. High data rate and flexible scheduling can enhance the QoS [13]. To allow quality-of-service (QoS) differentiation, the uplink traffic flows are grouped into four types of applications for 802.16 MAC:

• Unsolicited grant services (UGS): UGS is designed to support constant bit rate services, such as T1/E1 emulation and voice over IP (VoIP) without silence suppression.

• Real-time polling services (rtPS): It is used to support real-time variable bit rate services, such as MPEG video and VoIP with silence suppression.

• Non real-time polling services (nrtPS): It is used to support non real time variable bit rate services, such as FTP.

• Best-effort (BE) services: With BE services, packets are forwarded on a first-in-first-out basis using the capacity not used by other services. Web browsing is one example of it. The 802.16 MAC is connection oriented and every traffic flow is mapped into a connection, which is identified by a CID and assigned to one of the above four service types with a set of QoS and traffic parameters. The UGS traffic flow has the highest priority while the BE service has the lowest**.**

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, gaming) and the user's service level. The standard defines service flows which can be mapped to fine, granular IP sessions or coarsely differentiated services code points to enable end-to-end IP based QoS. Additionally, sub-channelization and media access protocol (MAP) based signaling schemes provide a flexible mechanism for optimal scheduling of broadcast and unicast traffic on a frame-by frame basis.

BENIFITS OF WiMAX

The various benefits of wimax are:

- Speed - Faster than broadband service
- Wireless - Not having to lay cables reduces cost. Easier to extend to suburban and rural areas.
- Broad Coverage- Much wider coverage than WiFi hotspot.

Benefits to Service Providers

- Allow service providers to deliver high throughput broadband based services like VoIP, high-speed Internet and Video
- Facilitate equipment compatibility
- Reduce the capital expenditures required for network expansion
- Provide improved performance and extended range

Benefits to Customers

Range of technology and service level choices from both fixed and wireless broadband operators

•DSL-like services at DSL prices but with portability

•Rapidly declining fixed broadband prices

•No more DSL "installation" fees from incumbent.

USES

The bandwidth and range of WiMAX make it suitable for the following potential applications:

• Providing portable mobile broadband connectivity across cities and countries through a variety of devices.

• Providing a wireless alternative to cable and digital subscriber line (DSL) for "last mile" broadband access.

• Providing data, telecommunications (VoIP) and IPTV services (triple play).

• Providing a source of Internet connectivity as part of a business continuity plan.

• Smart grids and metering

LIMITATIONS

A commonly-held misconception is that WiMAX will deliver 70 Mbit/s over 50 kilometers. In reality, WiMAX can either operate at higher bitrates or over longer distances but not both: operating at the maximum range of 50 km (31 miles) increases bit error rate and thus results in a much lower bitrate. Conversely, reducing the range (to under 1 km) allows a device to operate at higher bitrates. There are no known examples of WiMAX services being delivered at bit rates over around 40 Mbit/s. is also important to consider that a throughput of 2 Mbit/s can mean 2 Mbit/s symmetric simultaneously, 1 Mbit/s symmetric or some asymmetric mix, each of which required slightly different network equipment and configurations. Higher-gain directional antennas can be used with a WiMAX network with range and number of users that may be served as required. Like most wireless systems, available bandwidth is shared between users in a given Typically, fixed WiMAX networks have a higher-gain directional antenna installed near the client (customer) which results in greatly increased range and throughput. Mobile WiMAX devices typically have omni directional antennae which are of lower-gain compared to directional antennas but are more portable. In current deployments, the throughput may reach 2 Mbit/s symmetric at 10 km with fixed WiMAX and a high gain antenna.

CONCLUSION

The latest developments in the IEEE 802.16 group are driving a broadband wireless access evolution, thanks to a standard with unique technical characteristics. In parallel, the WiMAX forum, backed by industry leaders, helps the widespread adoption of broadband wireless access by establishing a brand for the technology.

When WiMAX chipsets are integrated into laptops and other portable devices, it will provide high-speed data services on the move, extending today's limited coverage of public radio sector, so performance could deteriorate in the case of many active users in a single sec will be added to the base station to increase the throughput benefits but the obvious loss of practical mobility.

WLAN to metropolitan areas. Integrated into new generation networks with seamless roaming between various accesses, it will enable end users to enjoy an "Always Best Connected" experience. The combination of these capabilities makes WiMAX attractive for a wide diversity of people: fixed operators, mobile operators and wireless ISPs, but also for many vertical markets and local authorities.

REFERANCES

[1] IEEE Stds. 802.16e-2005 and 802.16-2004/Cor 1-2005, "Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands" and Corrigendum 1, 2004.

[2] Carl Eklund, Nokia Research Center Roger B. Marks, National Institute of Standards and Technology Kenneth L. Stanwood and Stanley Wang, Ensemble Communications Inc. IEEE Standard 802.16: A Technical Overview of the WirelessMAN™ Air Interface for Broadband Wireless Access June 2002

[3] M. Kornfeld, "DVB-H — The Emerging Standard for Mobile Data Communication," IEEE Int'l. Symp. Consumer Elect., vol. 1, Sept. 2004, pp. 193–98.

[4]Christian Lottermann, MAC-Layer and Mobile WiMAX as an Example of Implementation, July 03, 2009

[5]Beth N. Komu, Mjumo Mzyece and Karim Djouani, "Formal Verification of Initial Network Entry in WiMAX Networks".

[6] Evren Eren , "WiMAX Security Architecture – Analysis and Assessment", IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications 6-8 September 2007.

[7] "Sprint Eyes WiMax Backhaul" (http:/ / www. lightreading. com/ document. asp?doc_id=104349). lightreading.com. .Retrieved 2008-03-22.

Approach to Detect Sinkhole Attack in Wireless Sensor Networks

Ekta[1], Manisha sharma[2], Harpreet Kaur[3], Ramanjyot Kaur[4]
Sant Baba Bhag Singh Institute of Engineering and Technology, Jalandhar
E-mail : [1] ekta.2917@yahoo.com, [2]manisha_sharma91@yahoo.com,
[3]erharpreetarora@gmail.com,[4]er.ramanjyot@gmail.com

***Abstract*-One of the reasons that the research of intrusion detection in wireless sensor networks has not advanced significantly is that the concept of "intrusion" is not clear in these networks. In this paper we investigate in depth one of the most severe attacks against sensor networks, namely the sinkhole attack, and we emphasize on strategies that an attacker can follow to successfully launch such an attack. Then we stated the recent approach for resolving the sinkhole attack. The stated scheme is to defend against sink hole attacks using mobile agents. Mobile agent is a program segment which is self controlling. They navigate from node to node not only transmitting data but also doing computation. They are an effective paradigm for distributed applications, and especially attractive in a dynamic network environment.**

1. INTRODUCTION

Securing sensor networks against the threats is a challenging research area, necessary for commercially attractive deployments. Encryption and authentication mechanisms provide reasonable defense for mote-class outsider attacks. However, cryptography is inefficient in protecting against laptopclass and insider attacks. It remains an open problem for additional research and development since the presence of insiders significantly lessens the effectiveness of link layer security mechanisms. This is because an insider is allowed to participate in the network and have complete access to any messages routed through the network and is free to modify, suppress, or eavesdrop on the contents. What makes it even easier for attackers is the fact that most protocols for sensor networks are not designed having security threats in mind. As a consequence, deployments of sensor networks rarely include security protection and little or no effort is usually required from the side of the attacker to perform the attack. This paper investigates one of the most severe routing attacks in sensor networks, namely the sinkhole attack from the attacker's point of view. Our goal is to describe the recent technique to detect the sinkhole attack in WSN.

2. THE SINKHOLE ATTACK

The *sinkhole* attack is a particularly severe attack that prevents the base station from obtaining complete and correct sensing data, thus forming a serious threat to higher-layer applications. In a Sinkhole attack [3], a compromised node tries to draw all or as much traffic as possible from a particular area, by making itself look attractive to the surrounding nodes with respect to the routing metric. As a result, the adversary manages to attract all traffic that is destined to the base station. By taking part in the routing process, she can then launch more severe attacks, like selective forwarding, modifying or even dropping the packets coming through.

2.1. Sinkhole Attack on MintRoute

MintRoute[1] uses link quality estimates as the routing cost metric to build the routing tree toward the base station. For the calculation of these link estimates, MintRoute uses the packet *error rate*. The nodes periodically transmit a packet, called "*route update*" and each node estimates the link quality of its neighbours based on the *packet loss* of the packets received from each corresponding neighbour. The list of these estimates for each neighbour is broadcasted by the node periodically in its route update packets. Every node maintains a Neighbour Table and updates it when it receives a route update packet. This table stores a list with the IDs of all neighbouring nodes and their corresponding link costs. The node chooses its "parent node" to be the one with the best link quality in the Neighbour Table. Note that the hop distance of each neighbour to the base station is not taken under consideration in choosing the parent, unless two nodes have the same link quality. The parent changing mechanism is triggered every time the link quality of one or more nodes becomes 75% better than the link quality of the current parent, or the link quality of the current parent drops below 25 in absolute value (with 255 being the maximum value). In such case, the node with the highest quality becomes the new parent. However, if two of such candidate nodes happen to have the same link quality, the new parent will be the one with the smaller hop count to the base station. In the case of a routing protocol, like MintRoute that uses link estimates as the routing metric, the compromised node launching the sinkhole attack will try to persuade its neighbours to change their current parents and choose the sinkhole node as their new one. There are two ways to do that:

1) Advertise an attractive link quality for itself,

2) Make other nodes look like they have worse link quality than itself.

2.2 Sinkhole Attack on MultiHopLQI

The weakness of MintRoute is that each node is based on the advertised link quality from other nodes to decide on its parent. In MultiHopLQI[1], the nodes calculate the link quality based on their own hardware. Each node periodically broadcasts a beacon message and the receivers extract the LQI given by their radio chip. This number is given to a function that calculates the cost of the corresponding link. The cost is inversely proportional to the LQI. The most attractive link is the one with the lowest cost. The payload of the beacon message includes the sender's current parent and a cost for the whole path to the base station (i.e., the path cost). This cost is calculated as the sum of all the costs of the links that make the path.

3. DETECTION OF SINKHOLE ATTACK IN WSN

3.1 COMPARISON OF FEW SINKHOLE DETECTION TECHNIQUES[2]

3.2(B). Basic definition

Definition 1: DAB -> Distance between two neighboring nodes(Say A & B).

DAB = (R-d) / V

Where R ->Transmission range

d ->Distance between Node A and Node B.

V ->Average speed of the node.

Definition 2: Counter of agent

It tells how many times the agent finds the particular Node as a one hop neighbor or as a child node to the previous Node. One mobile agent has agent ID, agent Program, agent Briefcase (It contains some condition parameters such as DAB , Counter)

An agent is capable of sharing its briefcase with other agents and nodes. The state variables may be updated if Necessary when an agent leaves a node.

Definition 3: Counter of every Node

Existing proposal name	Approach	Problem found
Packet leashes[4]	It is based on geographical and temporal packet leashes. The use of geographical leashes assumes knowledge of the node location.	The use of temporal leashes requires all nodes to have tightly synchronized clocks and demands computational power, which according to the authors, is beyond the capability of sensors.
SECTOR[5]	It is based on measurement of the time of flight of a message in a challenge–reply scheme.	Such a scheme assumes that sensors are able to execute time measurements of nanosecond precision and, hence, this scheme requires very accurate clocks at each sensor.
Sink hole attack detection[6]	It finds a list of suspected nodes, and then carries out a Network flow graph identifying a sink attack by observing data missing from an attacked area.	The method is based on a central processing unit,which is not convenient in a wireless sensor network.

3.2 Recent approach to detect sinkhole attack

This system is designed to make every node aware of the entire network so that a valid node will not listen the cheating information from malicious or compromised node which leads to sink hole attack. It is achieved with the help of mobile agents[2]. The mobile agents are developed using the Aglet. Aglet is a Java based system developed by IBM. Agents are called aglets in this System. The system proposes two Algorithms . Agent navigation algorithm tells how does a mobile agent gives network information to nodes and visits every node. Data routing algorithm tells how a node uses the global network information to route data packets.

3.2(A) . Protection of mobile agent

To protect the Mobile agent Dummy data is stored in the data base of the agent. It is not modified while the agent performs its functions. If dummy data is not modified after the agent's returns, then one can have confidence that legitimate data also has not been corrupted.

It tells how many times this node has been visited by an agent .i.e. It represents frequency of the visits by agents. The structure of every node is shown in the table 2.

TABLE 2

Structure of node

Node	1	2	. . .	I	. . .	N	Counter	History
1								
2								
. . .								
J								
. . .								
N								

This cell has two information. (If it is a non zero value):

- Dij
- Agent packet counter information tells how many times agent visited j after i.e., how many times agent found j as a one hop neighbour of i.

3.2(C) Agent Navigation Algorithm

The primary goal of agent is to deliver information of one node to others in the network. In order to achieve this goal with the least overload, we put forward a least visited neighbour first algorithm to control the navigation of mobile agent. An agent applies the algorithm to the information of node on which it currently resides, and decides its next destination. Each node has an information cache that agents can update with more recent values. Nodes access this shared cache whenever they require information about the network. When the agent reaches a node i, agent program performs the following steps:

Step1:

Update the information cache of node i with any newer information available in its own briefcase. The counter of all nodes stored in the information cache of node i is compared with the corresponding counter carried in the briefcase of this agent. If the counter of some nodes, say j, in the host node's information cache happens to be less than that in the agent's briefcase, obviously, the agent is carrying more recent information about node j. In that case, the entire information about node j in the host node's cache is overwritten by information in the agent's briefcase.In this step, information means how many times agent finds this particular node as a one hop neighbor to the previous node. Otherwise it means that how many times agent finds this particular node as a child node of the previous node.

Step2:

Determine which neighboring node has the least counter. It is the least visited neighbor.

Step3:

If this neighbor of i hasn't been visited in recent 3 times, the agent selects this neighbor as its next destination.History information about the last 3 visits can be found in the node's information cache. In case node selected has been visited in the recent past, the agent selects the second least visited neighbor, and so on.

Step4:

After choosing the next destination, the agent updates its next destination's ID with the chosen destination node ID, and changes the history variables in the host node's information cache with the next destination node.

3.2 Data Routing Algorithm

The node's information matrix can be acquired through mobile agent navigation algorithm. When the data packets wanted to be sent to node B, it can be transmitted by the Data routing algorithm according to node A's information matrix.

Suppose node A is the source node, node B is the destination node. Communicating with node B, node A performs as follows:

Step1:

Examine TabValAB of A's matrix. If TabValAB is not equal to 0 (has a non zero value), there is a connection between A and B. The data packets are sent to B directly. End routing. Otherwise, go to step2.

Step2:

check the column, B, in the cache of node B, and find out all the items which are not equal to zero in B, these items are the child nodes of node B;

If TabValB! = 0 for node1 to node n where n can be 1 to max number of nodes present in the network. So, node1 to node n are child nodes of B.

If all the items in B are equal to zero, then, there is no valid route between node A and nod B, and the routing ends.

Step3:

Set the maximum number of hops to reach the destination as n.

Step4:

Initialize k =1. Where k = current number of hops.

(After finding child nodes of node1 to node n, will be incremented by one. k can reach up to n.)

This process continues till it reaches to A.

Here maximum repeated hop with less weight is selected every time. i.e., Maximum agent counter value with less TabValAB. For every neighbouring nodes A & B. It limits the chance of paths containing sink hole.

4. CONCLUSION

In this paper we propose a mobile agent based approach that can be used to provide necessary knowledge to every sensor node in a Wireless Sensor Network not to believe the false path so that sink hole attack can be avoided at certain extent.

REFERENCES

[1]. Ioannis Krontiris, Thanassis Giannetsos, Tassos Dimitriou Athens Information Technology, 19002 Peania, Athens, Greece"(Launching a Sinkhole Attack in Wireless Sensor Networks; the Intruder Side)".

[2]. D. Sheela, Nirmala. S, Sangita Nath and Dr. G Mahadevan"(A Recent Technique to Detect Sink Hole Attacks in WSN)".

[3]. C. Karlof and D. Wagner, "Secure routing in wireless sensor networks: Attacks and countermeasures," *AdHoc Networks Journal*, vol. 1, no. 2–3, pp. 293–315, September 2003

[4]. Y. Hu, A. Perrig, and D. Johnson, Packet Leashes: A Defense Against Wormhole Attacks in Wireless Ad Hoc Networks, in: Proc. Of Infocom 2003. San Francisco, CA, USA, April 2003.

[5] S. Capkun, L. Buttyan, J. Hubaux, SECTOR: Secure Traking of Node Encounters in Muti-hop Wireless Networks, in: proc. Of SASN 2003. Fairfax, Virginia, October 2003.

[6] E. C. H Ngai, J. Liu and M R. Lyu, "On the intruder Detection for Sinkhole Attack in Wireless Sensor Networks," Proc. IEEE ICC, 2006.

A Comparative Analysis of Image Steganography Techniques

Gurmeet kaur[1], Aarti Kochhar[2], Sanjeev Kumar[3]
[1]Department of Electronics and Communication Engineering, CEM, Kapurthala
[2]Department of Electronics and Communication Engineering, DAVIET, Jalandhar
[3]Department of Electronics and Communication Engineering, ACET, Amritsar
er.meet_308@yahoo.com, aarti.kochher2@gmail.com, sanjeev_be@indiatimes.com

***Abstract*—In recent years, Steganography and Steganalysis are two important areas of research that involve a number of applications. These two areas of research are important especially when reliable and secure information exchange is required. Steganography is an art of embedding information in a cover image without causing statistically significant variations to the cover image. Steganalysis is the technology that attempts to defeat Steganography by detecting the hidden information and extracting. In this paper a comparative analysis is made to demonstrate the effectiveness of the proposed methods. The effectiveness of the proposed methods has been estimated by computing Mean square error (MSE) and Peak Signal to Noise Ratio (PSNR).The analysis shows that the BER and PSNR is improved in the case of DWT than DCT.**

***Index Terms*— Steganography, Discrete Wavelet Transform (DWT), code, Discrete Cosine Transform (DCT), LSB (Least Significant bit).**

1. INTRODUCTION

An important aspect of the modern way of life is communication. Many devices present today have the ability to transmit various information between themselves using different ways of communication, like insecure public networks, different types of wireless networks and the most used: the Internet. In some cases it is needed to keep the information travelling through different kinds of channels secret. Mainly there are two ways of concealing information: cryptography and steganography. Cryptography's main aspect is that the information is somehow distorted, scrambled by the sender using normally an encryption key also known only by the intended receiver who decrypts the message. The problem with cryptography is that a user intercepting the message, although he cannot decrypt it, he might detect that there is encrypted, secret information .On the other hand steganography is able even to hide this aspect making sure that even the fact that there is secret information, is concealed. Steganography's main aspect is that it is embedding the secret message into another message.The basic structure of Steganography is made up of three components: the "carrier", the message, and the key.

IMAGE STEGANOGRAPHY TECHNIQUES

Based on the analyses of steganography tools' algorithms, we partition these tools into two categories: (1). Spatial domain based steganography (2) Transform domain based steganography

Spatial Domain Based Steganography**:** Spatial steganography mainly includes LSB (Least Significant Bit) steganography Least significant bit (LSB) insertion is a common, simple approach to embedding information in a cover image . The least significant bit (in other words, the 8th bit) of some or all of the bytes inside an image is changed to a bit of the secret message.

Pixel: (10101111 11101001 10101000)
(10100111 01011000 11101001)
(11011000 10000111 01011001)
Secret message: 01000001
Result: (10101110 11101001 10101000)
(10100110 01011000 11101000)
(11011000 1000011

1 01011001)

Cover image Stego image

Transform Domain Based Steganography: Basically there are many kinds of power level transforms that exist to transfer an image to its frequency domain, some of which are Discrete Cosine Transform, KL Transform and Wavelet Transform.

The Discrete Cosine Transform (DCT): This method is used, but similar transforms are for example the Discrete Fourier Transform (DFT). These mathematical transforms convert the pixels in such a way as to give the effect of "spreading" the location of the pixel values over part of the image [5]. The DCT transforms a signal from an image representation into a frequency representation, by grouping the pixels into 8 × 8 pixel blocks and transforming the pixel blocks into 64 DCT.

Wavelet Method: The Haar wavelet basis was chosen. The basic process involves a low pass filter (l[n]) and a high pass filter (h[n]) the image is processed in 4 ways (producing 4 separate images as output). Its rows are convolved with l or h, as are its columns. The 4 image outputs have l rows and l columns, h rows and l columns, l rows and h columns, and h rows and h column. These images are then down sampled by 2, meaning that every other row and every other column is eliminated. The 4 images are then combined into one having the same dimensions as the original source image**.**

*Spread Spectrum:*In spread spectrum techniques, hidden data are spread throughout the cover-image making it harder to detect .Spread spectrum communication can be defined as the process of spreading the bandwidth of a narrowband signal across a wide band of frequencies. This can be accomplished by adjusting the narrowband waveform with a wideband waveform, such as white noise. After spreading, the energy of the narrowband signals in anyone frequency band is low and therefore difficult to detect. In spread spectrum image steganography the message is embedded in noise and then combined with the cover image to produce the stego image. Since the power of the embedded signal is much lower than the power of the cover image, the embedded image is not perceptible to the human eye or by computer analysis without access to the original image.

II.RELATED WORK

Vignesh Kumar Munirajan, Eric Cole, Sandy Ring in Steganography is a means of data hiding in images for covert transmission. Though steganography aims at transmitting images without visual degradation or changes for a naked observer, it cannot dispense with altering spatial and transform level details in order to embed the data [1, 2, 3]. Even though these alterations may not be captured by visual observation, they do manifest themselves for detailed analysis. In this paper we analyze a Fuzzy logic based technique that could be applied for detaining images with steganography. The images we are looking at are JPEG images and the data. T. Morkel, J.H.P**.** Eloff, M.S. Olivier [6] give an overview of image steganography, its uses and techniques. It also attempts to identify the requirements of a good steganographic algorithm and briefly reflects on which steganographic techniques are more suitable for which applications**.**

K Suresh Babu, K B Raja [6] proposes an image Steganography that can verify the reliability of the information being transmitted to the receiver. The method can verify whether the attacker has tried to edit, delete or forge the secret information in the stego-image. The technique embeds the hidden information in the spatial domain of the cover image and uses two special AC coefficients of the Discrete Wavelet Transform domain to verify the veracity (integrity) of the secret information from the stego image. The analysis shows that the BER and PSNR is improved in the case of DWT than DCT.

Hao-tian Wu,Jiwu Huang **[9]** in steganographic algorithm is proposed for JPEG Image by modifying the block DCT coefficients. Firstly, an embedding algorithm called LSB+ matching is generated to approximately preserve the marginal distribution of DC coefficients. We further divide the DCT coefficients into four frequency bands, including the direct current (DC), low frequency, middle-frequency, and high-frequency. Via matrix encoding, low data hiding rate and high embedding efficiency are achieved in high-frequency band, while the hiding rate is increased in the middle-frequency and DC bands, and highest in the low-frequency band. In addition, a coefficient selection strategy is employed to make the hidden message less detectable. The proposed algorithm is implemented on a set of 10000 images and tested with four steganalytic algorithms. Mamta Juneja, Parvinder Singh Sandhu [11] represented Robust image steganography technique based on LSB (Least Significant Bit) insertion and RSA encryption technique. Steganography is the term used to describe the hiding of data in images to avoid detection by attackers. Steganalysis is the method used by attackers to determine if images have hidden data and to recover that data. The application discussed in this paper ranks images in a users library based on their suitability as cover objects for some data. By matching data to an image, there is less chance of an attacker being able to use steganalysis to recover the data. Before hiding the data in an image the application first encrypts it. The steganography method proposed in this paper and illustrated by the application is superior to that used by current steganography tools.

Vladimir Bancoi, Gabriel Bugari, Dusan Levicky[12] present a novel method called Code Book in steganography system based on CDMA(Code Division Multiple Access) techniques having regard to perceptibility of stego-image. These techniques and their variations are widely used in radio telecommunication systems. As it will be shown later, aptly imposed features of CDMA techniques are in consonance with imperatives claimed upon steganography systems.

Yi-zhen Chen, Zhi Han, Shu-ping Li, Chun-hui Lu, Xiao-Hui Yao [13] an improved adaptive steganography algorithm—SVBA algorithm, which fully analyzes the statistical properties and adopts HVS features. SVBA algorithm first divides the image into 8*8 blocks and analyzes the mean, variance and entropy value of gray by block, then sets a sensitivity vector for each block with considering HVS features and adjusts the steganography schema dynamically according to the block sensitivity vectors. Simulation experiment results onMatlab7.0 shows this algorithm has a balanced performance on efficiency, capacity, imperceptibility and robustness., Mohammad Javad Khosravi, Samaneh Ghandali [20] novel steganography technique based on the combination of a secret sharing method and wavelet transform is presented. In this method, a secret imageis shared into some shares. Then, the shares and Fletcher- 16checksum of shares are hidden into cover images using aninteger wavelet based steganography technique.

III. MODEL

A. Definitions

(*i*)*Cover image***:** It is defined as the original image into which the required information is embedded. It is also termed as carrier image. The information should be embedded in such a manner that there are no significant changes in the statistical properties of the cover image.

(*ii*)*Stegoimage*: It is an unified image obtained by the combination of the payload and cover image.

(*iii*)*Perceptibility***:** It describes the ability of a third party (not the intended recipient) to visually detect the presence of hidden information in the stego image. The embedding algorithm is imperceptible when used on a particular image if an innocent third party, interested in the content of the cover image, is unaware of the existence of the payload. Essentially this requires that the embedding process not degrade the visual quality of the cover image.

(*iv*)*Robustness*: It characterizes the ability of the payload to survive the embedding and extraction process, even in the face of manipulations of the stego image such as filtering, cropping, rotating and compression.

(*v*)Security: It is inability of adversary to detect hidden images accessible only to the authorized user. The quality factor can enhance the security of the image. A steganographic image is perfectly secure when statistical data of the cover and stego images are
identical.

	LSB	**DCT**	**DWT**	**Spread Spectrum**
Invisibility	Low	High	High	High
Payload capacity	High	Medium	Medium	Medium
Robustness against statistical attacks	Low	Medium	High	High
Robustness against image manipulation	Low	Medium	Medium	Medium
Independent of file format	Low	Medium	Medium	Medium
PSNR	High	Medium	Medium	Medium
MSE	Less	Medium	Medium	Medium

Table 1 Comparative performance of all the methods

B. Error Analysis:

(*i*) *Bit Error Rate***:** For the successful recovery of the hidden information the communication channel must be ideal but for the real communication channel, there will be error while retrieving hidden information and this is measured by BER. The cover image is represented as covand stego image as steg in the given equation.

$$BER = \frac{1}{|image^{covg}|} \sum_{i=0}^{all\,pixels} |image^{covg} - image^{steg}|$$

where *i* is the pixel position

(*ii*) *Mean Square Error***:** It is defined as the square of error between cover image and the stego image. The distortion in the image can be measured using MSE.

$$\sum_{I=1}^{all\,pixels} \sum_{j=1}^{all\,pixels} \frac{(cov(i,j) - steg(i,j))^2}{N * N}$$

(*iii*) *Peak Signal to Noise Ratio*: It is the ratio ofthe maximum signal to noise in the stego image.

$$PSNR = 20\log_{10} \frac{255}{\sqrt[2]{MSE}}$$

IV. CONCLUSION

Steganography is the art and science of writing hidden messages in such a way that no one, apart from the sender and intended recipient, suspects the existence of the message, a form of security through obscurity. It is therefore a book on magic. It is emerging in its peak because it does not attract anyone by itself. In this paper a comparative analysis of several methods has been successfully implemented and results are delivered. The MSE and PSNR of all the methods are also compared and also this paper presented a background discussion and implementation on the major algorithms of steganography deployed in digital imaging.

REFERENCES

[1] Anil K Jain, “Fundamentals of Digital Image Processing”, University of California-Davis, Prentice Hall, 1988

[2] Marvel, L.M., Boncelet Jr., C.G. & Retter, C., “Spread Spectrum Steganography”, IEEE Transactionson image processing, 8:08, 1999

[3] Chang, C.C., Chen, T.S. and Chung, L.Z., “A steganographic method based upon JPEG and quantization table modification”, Information Sciences, 2002, 141(1-2), pp.123-38.

[4] T. Morkel, J.H.P. Eloff , M.S. Olivier,"An Overview of Image Steganography," in Proceedings of the Fifth Annual Information Security South Africa Conference (ISSA2005), Sandton, South Africa, June/July 2005 .

[5] Ankur M. Mehta, Steven Lanzisera, and Kristofer S. J. Pister, “Steganography in 802.15.4 Wireless Communication”.

[6] K Suresh Babu, K B Raja, Kiran Kumar K, Manjula Devi T H, Venugopal K R, L M Patnaik,“Authentication of Secret Information in Image Steganography” Microprocessor Applications Laboratory, Indian Institute of Science,Bangalore

[7] Proceedings of the 2006 International Conference on “Intelligent Information Hiding and Multimedia Signal Processing “(IIH-MSP'06)0-7695-2745-0/06 © 2006 IEEE.

[8] Asghar Shahrzad Khashandarag and Naser Ebrahimian, “A new method for color image steganography using SPIHT and DFT, sending with JPEG format”, International Conference on Computer Technology and Development, IEEE, 2008

[9] CHEN Zhi-li, HUANG Liu-sheng, YU Zhen-shan, LI Ling-jun and YANG wei, “A Statistical Algorithm for Linguistic Steganography Detection Based on Distribution of Words”, 3RD International Conference on Availability, Reliability and Security, IEEE, 2008.

[10]K Suresh Babu, K B Raja, Kiran Kumar K, Manjula Devi T H, Venugopal K R, L M Patnaik, “Authentication of Secret Information in Image Steganography” Microprocessor Applications Laboratory, Indian Institute of Science, Bangalore in 2008

[11] Mamta Juneja Parvinder Singh Sandhu, “Designing of Robust Image Steganography Technique Based on LSB Insertion and Encryption” 2009 International Conference on Advances in Recent Technologies in Communication and Computing.

[12]Vladimír Banoci, Gabriel Bugar Dusan Levicky, “Steganography Systems by using CDMA Techniques” 978-1-4244-3538-8/09/©2009 IEEE

[13]Yi- Zhen Chen, Zhi Han, Shu-ping Li, Chun- hui Lu, Xiao- Hui Yao , “An Adaptive Steganography Algorithm Based on Block Sensitivity Vectors Using HVS Features “ 2010 3rd International Congress on Image and Signal Processing.

[14] James Cane, TheoYi- zhen Chen, Zhi Han, Shu-ping Li, Chun- hui Lu, Xiao- Hui Yao , “An Adaptive Steganography Algorithm Based on Block Sensitivity Vectors Using HVS Feature” 2010 3rd International Congress on Image and Signal Processing.

[15] Adel Almohammad, Gheorghita Ghinea, “Image Steganography and Chrominance Components” 2010 10th IEEE International Conference on Computer and Information Technology (CIT 2010)

[16]Jinsuk Baekl, Cheonshik Kim, Paul S. Fisherl, and Hongyang Cha0, “(N. 1) Secret Sharing Approach Based on Steganography with Gray Digital Images.” 978-1-4244-5849-3/10/$26.00 ©20 1 0 IEEE

[17] Suhaila Abd Halim and Muhammad Faiz Abdullah Sani , “Embedding Using Spread Spectrum ImageSteganography with GF)”Proceedings of the 6th IMT-GT Conference on Mathematics, Statistics and its Applications (ICMSA2010)

[18]Tetsuya Kojima , Yoshiya Horii , “A Steganography Based on CT-CDMA Communication SchemeUsing Complete Complementary Codes” arXiv:1001.2623v1 [cs.IT] 15 Jan 2010

[19]T. Morkel, J.H.P. Eloff, M.S. Olivier, “Information and Computer Security Architecture (ICSA) Research Group”.

[20]Mohammad Javad Khosravi, Samaneh Ghandali, “A Secure Joint Wavelet Based Steganography and Secret Sharing Method” 978-1-4577-2155

A Comparative study of Discrete Cosine Transformation and Haar Transformation

Harmandeep Singh Chandi
Amritsar College of Engineering & Technology
Punjab, India
Email:harmandeep.cgc@gmail.com

***Abstract*-With the rapid spread of image and video processing applications and the further development of multimedia technologies, compression standards become more and more important every day. For image compression, it is very necessary that the selection of transform should reduce the size of the resultant data as compared to the original data set.This paper addresses the area of data compression, as it is applicable to image processing. An analysis of Discrete Cosine Transformation and Wavelet Transformation (Haar Transformation) are made. These two image compression strategies are examined for their relative effectiveness. Finally a performance comparison is made between these techniques (on variety of Standard Test Images considering parameters such as data reduction ratios).**

Keywords: DCT, DWT, Image compression, Image processing

1. INTRODUCTION

Compression deals with techniques for reducing the storage required for saving an image or the bandwidth required for transmitting it. To reduce the increasing demand of storage space and transmission time, compression techniques are the need of the day. The main motive of image reduction process is to remove the redundant data. In the specific area of still image compression, many efficient compression techniques, with considerably different features, have been developed [1]-[5]. Two of these are- discrete cosine transform (DCT) based JPEG [1], and discrete wavelet transform (DWT) based SPIHT [4].

JPEG is a popular and continuous tone still image compression mechanism established by first Joint Photographic expert Group in 1992. In JPEG, both encoder and decoder are DCT based. This is a block based technique; where the original image is divided into small n x n (usually 8x8) blocks and then DCT transformation is applied. The data compression is achieved via quantization followed by variable length coding (VLC). The disadvantage of JPEG is the "blocking artifacts" in reconstructed image. JPEG is very efficient coding method but the performance of block-based DCT scheme degrades at high compression ratio. In recent time, much of the research activities in image coding have been focused on the Discrete Wavelet Transform (DWT). DWT offers adaptive spatial-frequency resolution (better spatial resolution at high frequencies and better frequency resolution at low frequencies) that is well suited to the properties of human visual system (HVS). It can provide better image quality than DCT, especially at higher compression ratio [8]. Set Partitioning in Hierarchical Trees (SPIHT) coding algorithm introduced by Said and Pearlman [4] is a very efficient technique for wavelet image compression. SPIHT is improved and extended version of Embedded Zero tree Wavelet (EZW) coding algorithm introduced by J. M. Shapiro [3] and is one of the best wavelet coders available to date. It produces an embedded bit stream from which the best reconstructed images in the mean square error sense can be extracted at various bit rates.

2. THE DISCRETE COSINE TRANSFORMATION (DCT)

Several techniques can transform an image into frequency domain, such as DCT, DFT [1] and wavelet transform. Each transform has its advantages. First here the DCT technique is discussed.
The most common DCT definition of a 1-D sequence of length N is:

$$Y[k] = C[k]\sum_{n=0}^{N-1} X[n]\cos[\frac{(2n+1)k\pi}{2N}]$$

For k= 0, 1, 2,…, N− 1. Similarly, the inverse DCT transformation is defined as

$$X[n] = \sum_{k=0}^{N-1} C[k]Y[k]\cos[\frac{(2n+1)k\pi}{2N}]$$

For k= 0, 1, 2,…, N− 1 . In both equations (1.1) and (1.2) C[n] is defined as

$$C[n] = \begin{bmatrix} \sqrt{\frac{1}{N}} & for\ n = 0 \\ \sqrt{\frac{2}{N}} & for\ n = 1,2,\ldots,N-1 \end{bmatrix}$$

The 2-D DCT is a direct extension of the 1-D case and is given by:

$$y[j,k] = C[j]C[k]\sum_{m=0}^{N-1}\sum_{n=1}^{N-1} x[m,n]\cos[\frac{(2m+1)j\pi}{2N}]$$

Where: j, k = 0, 1, 2,…, N −1 and. The inverse transform is defined as:

$$x[m,n] = \sum_{j=0}^{N-1}\sum_{k=1}^{N-1} C[j]C[k]y[j,k]\cos[\frac{(2m+1)j\pi}{2N}]\cos[\frac{(2n+1)k\pi}{2N}]$$

Where: m, n = 0, 1, 2, …, N −1. And c[n] is as it is as in 1-D transformation.

Discrete cosine transform (DCT) is widely used in image processing, especially for compression algorithm for encoding and decoding in DCT technique is shown below.

2.1 Encoding System

There are four steps in DCT technique to encode or compress the image

Step1. The image is broken into N*N blocks of pixels. Here N may be 4, 8, 16, etc.

Step2. Working from left to right, top to bottom, the DCT is applied to each block.

Step3. Each block's elements are compressed through quantization means dividing by some specific value.

Step4. The array of compressed blocks that constitute the image is stored in a drastically reduced amount of space.

So first the whole image is divided into small N*N blocks then DCT is applied on these blocks. After that for reducing the storage space DCT coefficients [5] are quantized through dividing by some value or by quantization matrix. So that large value is become small and it need small size of space. This step is lossy step. So selection of quantization value or quantization matrix [10] is affecting the entropy and compression ratio. If we take small value for quantization then we get the better quality or less MSE (Mean Square Error) but less compression ratio. Block size value also affects quality and compression ratio. Simply the higher the block size higher the compression ratio but with loss of more information and quality.

2.2 Decoding System

Decoding system is the exact reverse process of encoding. There are four steps for getting the original image not exact but identical to original from compressed image.

Step1. Load compressed image from disk

Step2. Image is broken into N*N blocks of pixels.

Step3. Each block is de-quantized by applying reverse process of quantization.

Step4. Now apply inverse DCT on each block. And combine these blocks into an image which is identical to the original image.

In this decoding process, we have to keep N's value same as it used in encoding process. Then we do de-quantization process by multiplying with quantization value or quantization matrix. As earlier said that this is lossy technique so output image is not exact copy of original image but it is same as original image. So this process' efficiency is measure by compression ratio. Compression ratio [3] is defined by ratio of storage bits of original image and storage bits of compressed image.

C.R.=n1/n2

Where n1 is number of bits to store original image and n2 is number of bits to store compressed image.

Loss of information is measure by Mean square Error (MSE)[1,5] between reconstructed image and original image. If MSE of reconstructed image to original image is greater than the information lost is more.

$$MSE = \sum_{i=1}^{M}\sum_{j=1}^{N}\left(x(i,j) - x'(i,j)\right)^2$$

Where M, N is dimension of image. x(i, j) is pixel value of (i,j) coordinate of original image while x'(i,j) is the reconstructed image's pixel value.

3. DWT TECHNIQUE

Wavelet analysis [1,3,7] can be used divided the information of an image into approximation and detailed sub signal[3]. The approximation sub signal shows the general trend of pixel value, and three detailed sub signal show vertical, horizontal and diagonal details or changes in image. If these detail is

very small than they can be set to zero without significantly changing the image. If the number of zeroes is greater than the compression ratio is also greater. There is two types of wavelet is used. First one is Continues wavelet transform[1] and second one is Discrete wavelet transform.[1] Wavelet analysis is computed by filter bank. There is two type of filter
1) High pass filter [1]: high frequency information is kept, low frequency information is lost.
2) Low pass filter [1]: law frequency information is kept, high frequency information is lost.
So signal is effectively decomposed into two parts, a detailed part (high frequency) and approximation part (law frequency). Level 1 detail is horizontal detail, level2 detail is vertical detail and level 3 details is diagonal detail of the image signal.

The wavelet theory is based on Haar transformation and the Daubechies transform but here we will use the Haar transformation theory only. Wavelet image compression belongs to the transform class of methods and such differs from other methods. Wavelet methods use self similarity across different scales to reduce stored in the wavelet transform domain. The idea behind the wavelet image compression, like that of the other transforms compression technique, fairly simple. One applies transform to an image and then removes some of the coefficient data from the transformed image. The compressed image is reconstructed by decoding the coefficient, if necessary and applies the inverse to the result. As a simple-minded approach to the image compression, one could replace an image by averaging of the pixel values. Consider an "image "with two pixel values :{ X1, X2}. These two values can be replaced by the average "a "and difference "d" of the values [04].a = (x1+ x2)/2 is the low Pass Filter. The high Pass Filter is d =(x1-x2)/2, the "wavelet transform" of the original Image sequence {x1,x2}is {a,d}.No information is gained or lost in this representation.

3.1Encoding System

Six steps process for compressing an image with discrete wavelet transform is shown below.
Step1.First original image have to been passed through high pass filter and low pass filter by applying filter on each row.
Step2.now output of the both image l1 and h1 are combining into t1= [l1 h1].
Step3. T1 is down sampled by 2.
Step4. Now, again T1 has been passed through high pass filter and low filter by applying on each column.
Step5. Output of the step4 is supposed l2 and h2. Then l2 and h2 is combine into

t3= $\begin{bmatrix} l2 \\ h2 \end{bmatrix}$.

Step6. Now down sampled t3 by 2.This is our compressed image.

3.2 Decoding System

. Here decoding system's process is not exact reverse of encoding system's process. Steps are shown below.
Step1.Extract low pass filter image and high pass filter image from compressed image simply by taking upper half rectangle of matrix is low pass filter image and down half rectangle is high pass filter image
Step2. Both images are up sampled by 2.
Step3.Now we take the summation of both images into one image called r1.
Step4. Then again extract low pass filter image and high pass filter image by simply dividing vertically. First half is low pass filtered image. And second half is high pass filter image.
Step5.Take summation of both images that is out reconstructed image.

4. THE ANALYSIS

In last we compare the three techniques and the implementation is given. It is hereby to note that the compression ratio is much satisfactory in the case of Haar transformation but the compression ratio is calculated on the reduced scale. The compression ratio is different for different images in case of Simultaneous Encryption and Compression and the Compression ratio is nearly equal in Discrete Cosine Transformation. There is one point to note that the compression ratio is also dependent on the type of Images. The speed of the compression is also dependent on the processor and the type of language used to implement the algorithm i.e. the same algorithm when implemented on the lower level language it takes smaller amount of time than implemented on the high level language.

CONCLUSION

The conclusion that can be drawn from the above facts that the compression ratios are also dependent on the file size and the algorithms used. Better compression ratios can be achieved by using the appropriate file size. Lastly the high compression ratios can be achieved selecting the nature of the images.

REFERENCES

[1] Rafael C. Gonzalez, Richard E. Woods.(1992), Digital Image Processing(2nd edition), NJ:Prentice Hall

[2] LockerGnome(2011),”Real World Application Of Image Compression” http://www.lockergnome.com/nexus/windows/2006/12/25/real-world-applications-of-imagecompression/ [accesed 11 Dec 2011]

[3] Swastik Das and Rashmi Ranjan Sethy, “A Thesis on Image Compression using Discrete Cosine Transform and Discrete Wavelet Transform”, Guided By: Prof. R. Baliarsingh, dept of Computer Science & Engineering, National Institute of Rourkela.

[4] Andrew B. Watson, NASA Ames Research, “Image Compression Using the Discrete Cosine Transform”, Mathematica Journal, 4(1), 1994, p. 81-88.

[5] M. Stumpfl, “Digital Watermarking”, University of Southampton, 2001.

[6] Nikolay N. Ponomarenko, Vladimir V. Lukin, Karen Egiazarian, Jaakko Astola. DCT Based High Quality Image Compression. In Proceedings of SCIA'2005. pp.1177~1185

[7] Karen Lees "Image compression using Wavelets", Report of M.S. 2002

[8] Saeid Saryazdi and Mehrnaz Demehr(2005),”A blind DCT domain digital watermarking ”, Proceeding of 3rd International Conference: SETIT, Tunisia: march-2005

[9] G. R. Ramesh, and K. Rajgopal, "Binary Image Compression using the Radon Transform,” in IEEE XVI Annual Convention and Exhibition, pp.178-182, 1990.

[10] Nopparat Pantaesaena, M.Sangworaisl, C. Nantajiwakornchai and T. Phanpraist, “Image compression using vector quantization”, ReCCIT, Thialand,2005

Proactive and Reactive Protocols of MANET

Kulvir Kaur[1], Sandeep Kaur[2], Harpreet Kaur[3], Ramanjyot Kaur[4]
Sant Baba Bhag Singh Institute of Engineering & Technology, Jalandhar
E-mail: [1]kulvir_bal@yahoo.in,[2]sanu_gill91@yahoo.com
[3]erharpreetarora@gmail.com,[4]er.ramanjyot@gmail.com

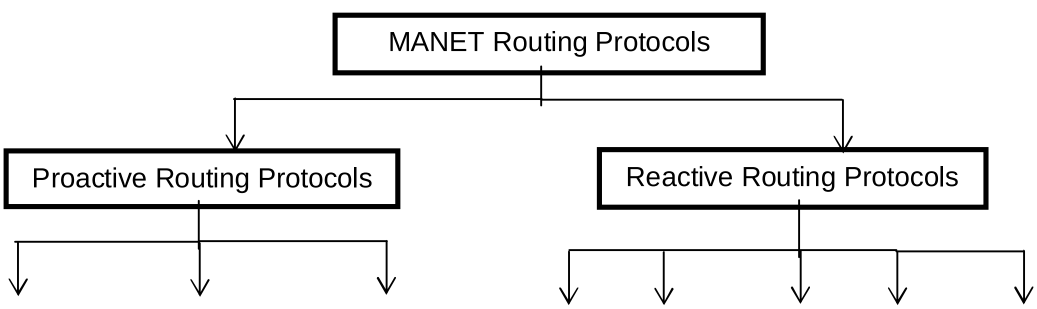
***Abstract* -MANET offers a convenient infrastructure-less communication over a shared wireless channel. These, being cost-effective and quick to install, find many applications such as military tactical operations, emergencies and law enforcement, rescue missions, and many other applications like round table conferences and classroom discussions etc. The primary goal of such an ad hoc network routing protocol is correct and efficient route establishment between a pair of nodes so that messages may be delivered in a timely manner. In Mobile Ad hoc Network (MANET), mobility, traffic and node density are main network conditions that significantly affect the performance of routing protocols.**

1. Introduction

Mobile Ad-hoc Network (MANET) is a collection of autonomous nodes that form a dynamic purpose-specific multi hop radio network in a decentralized fashion. As nodes move about in an unpredictable fashion, these networks must be configured on the fly to handle the dynamic topology. These networks with no fixed topology have been constrained with limited energy and processing capabilities of nodes and lack centralized administration. These networks also carry all the disadvantages of wireless medium like shared physical medium, higher bit error rates etc. In brief, MANET characteristics are as enumerated below.

- There is no fixed topology.
- Each node is a router.
- The transmission medium is shared.

2. Routing Protocols Used in MANET



DSDV WRP CGSR AODV DSR
TORA ABR SSR

Figure 1: Classification of routing protocol in MANET

2.1. Proactive Routing Protocols

Table driven protocols maintain consistent and up to date routing information about each node in the network. These protocols require each node to store their routing information and when there is a change in network topology updation has to be made throughout the network [2]. These types of protocols are also known as Table Driven Routing Protocols.

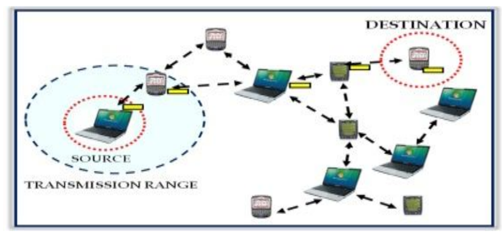


Figure 2: Working of proactive routing protocols

2.1.1. Destination-SequencedDistance-Vector Routing(DSDV)

The Destination Sequenced Distance Vector Routing protocol (DSDV) is a table-driven algorithm based on the classical Bellman-Ford routing mechanism. The improvements made to the Bellman-Ford algorithm include freedom from loops in routing tables.

Every mobile node in the network maintains a routing table in which all of the possible destinations within the net- work and the number of hops to each destination are recorded. Each entry is marked with a sequence number assigned by the destination node.

2.1.2.The Wireless Routing Protocol(WRP)

The Wireless Routing Protocol (WRP) is a table-based protocol with the goal of maintaining routing information among all nodes in the network. Each node in the network is responsible for maintaining four tables:

- Distance table
- Routing table
- Link-cost table
- Message retransmission list (MRL) table

2.1.3. Clusterhead Gateway Switch Routing(CGSR)

The Clusterhead Gateway Switch Routing (CGSR) protocol differs from the previous protocol in the type of addressing and network organization scheme employed. Instead of a "flat" network, CGSR is a clustered multihop mobile wireless network with several heuristic routing schemes [4]. A cluster head selection algorithm is utilized to select a node as the cluster head using a distributed algorithm within the cluster. The disadvantage of having a cluster head scheme is that frequent cluster head changes can adversely affect routing protocol performance since nodes are busy in cluster head selection rather than packet relaying.

2.2. Reactive Routing Protocols

In on-demand routing protocols routes are created as and when required. When a source wants to send to a destination, it invokes the route discovery mechanisms to find the path to the destinations. The route remains valid till the destination is reachable or until the route is no longer needed [6] . These types of protocols are also known as Source Initiated On-Demand Routing Protocols.

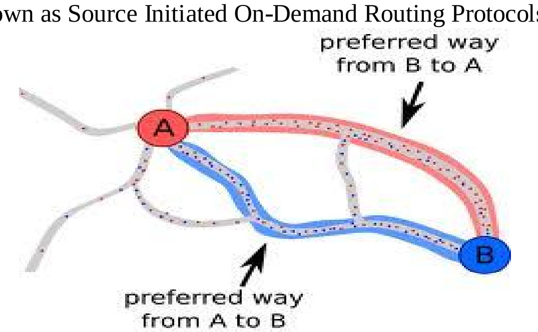


Figure 3: Working of Reactive Routing Protocols

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

The Ad Hoc On-Demand Distance Vector (AODV) routing protocol builds on the DSDV algorithm previously described. AODV is an improvement on DSDV because it typically minimizes the number of required broadcasts by creating routes on a demand basis, as opposed to maintaining a complete list of routes as in the DSDV algorithm. When a source node desires to send a message to some destination node and does not already have a valid route to that destination, it initiates a path discovery process to locate the other node. It broadcasts a route request (RREQ) packet to its neighbours, which then forward the request to their neighbours. Once the RREQ reaches the destination or an intermediate node with a fresh enough route, the destination/intermediate node responds by unicasting a route reply (RREP) packet back to the neighbour from which it first received the RREQ[4].

2.2.2. Dynamic Source Routing(DSR)

The Dynamic Source Routing (DSR) protocol is an on-demand routing protocol that is based on the concept of source routing. Mobile nodes are required to maintain route caches that contain the source routes of which the mobile is aware. Entries in the route cache are continually updated as new routes are learned [7]. The protocol consists of two major phases: route discovery and route maintenance. When a mobile node has a packet to send to some destination, it first consults its route cache to determine whether it already has a route to the destination. If it has an unexpired route to the destination, it will use this route to send the packet. A route reply is generated when the route request reaches either the destination itself, or an intermediate node which contains in its route cache an unexpired route to the destination. By the time the packet reaches either the destination or such an intermediate node, it contains a route record yielding the sequence of hops taken.

2.2.3. Temporally Ordered Routing Algorithm(TORA)

The Temporally Ordered Routing Algorithm (TORA) is a highly adaptive loop-free distributed routing algorithm based on the concept of link reversal. TORA is proposed to operate in a highly dynamic mobile networking environment [3]. It is source-initiated and provides multiple routes for any desired source/destination pair. The key design concept of TORA is the localization of control messages to a very small set of nodes near the occurrence of a topological change. To accomplish this, nodes need to maintain routing information about adjacent (one-hop) nodes.

The protocol performs three basic functions:

- Route creation
- Route maintenance
- Route erasure

2.2.4. Associativity-Based Routing(ABR)

The Associativity- Based Routing (ABR) protocol is free from loops, deadlock, and packet duplicates, and defines a new routing metric for ad hoc mobile networks. This metric is known as the degree of association stability [5]. In ABR, a route is selected based on the degree of association stability of mobile nodes. Each node periodically generates a beacon to signify its existence. Association stability is defined by connection stability of one node with respect to another node over time and space. A high degree of association stability may indicate a low state of node mobility, while a low degree may indicate a high state of node mobility. A fundamental objective of ABR is to derive longer-lived routes for ad hoc mobile networks. The three phases of ABR are:

- Route discovery
- Route reconstruction (RRC)
- Route deletion

2.2.5. Signal Stability Routing(SSR)

Another on-demand protocol is the Signal Stability-Based Adaptive Routing protocol (SSR). Unlike the algorithms described so far, SSR selects routes based on the signal strength between nodes and a node's location stability. This route selection criteria has the effect of choosing routes that have "stronger" connectivity.
SSR can be divided into two cooperative protocols:

- The Dynamic Routing Protocol (DRP)
- The Static Routing Protocol (SRP).

3. COMPARISON OF ROUTING PROTOCOLS

In general, on-demand reactive protocols are more efficient than table driven proactive ones. As described in [1] On-demand protocols minimize control overhead and power consumption since routes are only established when required. By contrast, proactive protocols require periodic route updates to keep information current and consistent; in addition, maintain multiple routes that might never be needed, adding unnecessary routing overheads. Proactive routing protocols provide better quality of service than on-demand protocols.

Overall comparisons of on-demand versus table-driven routing protocols:

Parameters	Table-driven	On-demand
Availability of routing information	Always available regardless of need	Available when needed
Routing philosophy	Mostly flat, except for CGSR	Flat
Periodic route updates	Required	Not required
Coping with mobility	Inform other nodes to achieve a consistent routing table	Use localized route discovery as in ABR and SSR
Signalling traffic generated	Greater than that of on demand routing	Grows with increasing mobility of active routes (as in ABR)
Quality of service support	Mainly shortest support shortest path QoS metric	Few can support QoS, although most support shortest path

4. CONCLUSION

In Mobile Ad hoc Network (MANET), mobility, traffic and node density are main network conditions that significantly affect the performance of the network. This issue has been reviewed in this paper. And also, in this article we provide descriptions of several routing schemes proposed for ad hoc mobile networks. We also provide a classification of these schemes according to the routing strategy (i.e., proactive and reactive). We have presented a comparison of these two categories of routing protocols, highlighting their features, differences, and characteristics. Finally, we have identified possible applications of ad hoc mobile wireless networks. The field of ad hoc mobile networks is rapidly growing and changing, and while there are still many challenges that need to be met, it is likely that such networks will see widespread use within the next few years.

REFERENCES

[1] Elizabeth M. Rayor and Chai-Keong Toh,"A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks," IEEE Personal Communications, pp.46-55, 1999.
[2] Geetha Jayakumar, and G. Gopinath," Ad Hoc Mobile Wireless Networks Routing Protocols – A Review," Journal of Computer Science, pp.574-582, 2007.
[3] Huda Al Amri, Mehran Abolhasan and Tadeusz Wysocki," Scalability of MANET Routing Protocols for Heterogeneous and Homogenous Networks," Computer & Electronics Engineering Faculty Publications, 2007.
[4] Khan, A.Y., S. Rashid, and A. Iqbal,"Mobility vs. predictive MPR selection for mobile ad hoc networks using OLSR", 2005.
[5] L. McNamara, C. Mascolo and L. Capra, "Media sharing based on colocation prediction in urban transport, in: MobiCom '08: Proceedings of the 14th ACM international conference on Mobile computing and networking", 2008.
[6] T. Claveirole, M. Boc and M. D. de Amorim, "An empirical analysis of wi-fi activity in three urban scenarios," 2009.
[7] J.Whitbeck and V. Conan, "HYMAD: Hybrid DTN-MANET Routing for Dense and Highly Dynamic Wireless Networks," in: AOC'09: Proceedings of the Third IEEE WoWMoM Workshop on Autonomic and Opportunistic Communications, 2009.
[8] A. Ker¨anen, J. Ott and T. K¨arkk¨ainen, "The ONE Simulator for DTN Protocol Evaluation," in: SIMUTools '09: Proceedings of the 2nd International Conference on Simulation Tools and Techniques, "2009.

Handwritten Gurmukhi Character Recognition

Mandeep Kaur[1], Sanjeev Kumar[2]
[1]Department of Electronics and Communication, ACET, Amritsar, Punjab, India
[2]AP, Department of Electronics and Communication, ACET, Amritsar, Punjab, India
Email: mandeepk_gill87@yahoo.co.in, sanjeev_be@indiatimes.com

***Abstract*—Recognition of Indian language scripts is a challenging area of research. Though, several research works have been focusing toward evolving newer techniques and methods that would reduce the processing time while increasing the recognition accuracy but still recognition of handwritten characters is complex due to varying writing styles. This paper will provide a path for developing recognition system for gurmukhi script. In this paper different techniques which are used for the recognition of handwritten gurmukhi characters such as pre-processing, segmentation, feature extraction and classification of the characters are discussed.**

***Keyword:-* Gurmukhi character recognition system, Pre-processing, Segmentation, Feature Extraction, Classification.**

I. INTRODUCTION
II.

Much of the world's information is held captive in hard copy documents. There are many applications in Indian offices such as bank, sales-tax, railway, embassy, etc. where both English and regional languages are used. Many forms and applications are filled in regional languages and in order to digitize the forms or other documents we have to scan directly. If we are not having handwritten character recognition system, then the image is directly captured and we have no option for editing in those documents. The handwritten Character Recognition systems are used to liberate this information by converting the text on paper into electronic form. Character recognition contributes immensely to the advancement of an automation process and can improve the interface between man and machine. Handwritten character recognition is mainly of two types online and offline. In online handwriting recognition, data is captured during the writing process with the help of a special pen on electronic surface. In offline handwriting recognition, prewritten data generally written on a sheet of paper is scanned.The process of handwritten character recognition of any script can be broadly classified into five stages i.e. Pre-processing, Segmentation, Feature Extraction, Classification and Post-processing. The first important step for recognition is preprocessing followed by segmentation and feature extraction. The selection of appropriate feature extraction method is probably the single most important factor in achieving high recognition performance. Artificial neural networks (ANN), Support Vector Machines (SVM), Structural pattern recognition, etc can be used for classification process. The neural networks have emerged as the fast and reliable tools for classification towards achieving high recognition accuracy.

II.INTRODUCTION TO GURMUKHI SCRIPT

Gurmukhi script is the most common script used for writing the Punjabi language. Gurmukhi was standardized by the second Sikh guru, Guru Angad Dev Ji. The whole of Sri Guru Granth Sahib Ji's1430 pages are written in this script.

The name gurmukhi is derived from the old Punjabi term "guramukhi" meaning "from the mouth of the Guru".Gurmukhi script is 12th most widely used script in the world. There is rich literature in this language in the form of scripture, books, poetry, etc. Gurmukhi is the first official script adopted by Punjab state. So it is important to develop a recognition system for such a rich and widely used language.Writing style of Gurmukhi script is from top to bottom and left to right. In gurmukhi script, there is no case sensitivity and most of the characters have a horizontal line at the upper part called headline and characters are connected with each other through this line. It consists of 41 consonants which are shown in figure 1.

Figure 1: Basic characters of Gurmukhi Script

III.THE PROPOSED RECOGNITION SYSTEM

The proposed recognition system consists of the phase's digitization, preprocessing, feature extraction and classification. The detail of each phase of handwritten Gurmukhi character recognition system is given in the proceeding sections. The block diagram of proposed recognition system is given in figure 3.

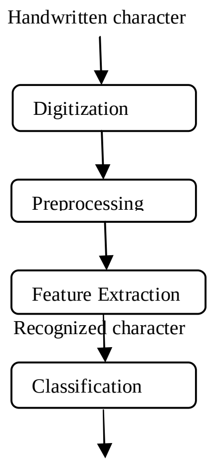


Figure 2: Block diagram of the character recognition system

Digitization

Digitization is the process of converting the paper based handwritten data into electronic form with the help of scanner and an electronic representation of the original document, in the form of bitmap image, is produced. Digitization produces the digital image, which is fed to the next phase of the recognition system i.e. preprocessing phase.

Preprocessing

Preprocessing is a series of operations such as binarization, noise reduction, thinning, skew correction and segmentation performed on the digitized image. In binarization, the grayscale image is converted to binary image by using the most common technique i.e. thresholding. Digital images are prone to various types of noises. Noise reduction techniques improve the quality of the document. Some morphological operations such as dilation (to bridge unconnected pixels) and erosion (to remove isolated pixels) are applied. Thinning provides tremendous reduction in data size and extracts the shape information of the characters.

Segmentation

Text segmentation is a process in which the text image is segregated to form characters. Any error in segmenting the basic shape of characters will produce a different representation of the character component. The segmentation provides the separation of different logical parts, like lines of a paragraph, words of a line and characters of a word. The white space or inter-character gap is used to segment the characters.

IV .FEATURE EXTRACTION

Extraction of good features is the main key to correctly recognize an unknown character. Feature extraction [2] [3]is the important step which is used to extract the most relevant information which is further used to classify the objects. Features can be broadly classified into following categories: -

1) *Local features:* Local features are those which are usually geometric (e.g. concave/convex parts, number of endpoints, branches, joints, etc).
2) *Global features:* Global features are those which are usually topological (connectivity, projection profiles, number of holes, etc) or statistical (invariant moments etc.)
3) *Structural features*: Structural features are involved of structural elements like loop, line, crossing point, curve, end point and stroke etc.
4) *Statistical features:* Statistical features are computed by some statistical operations on image pattern and these include features like zoning, projection, profiling, histogram and distance etc.

After a careful analysis of the shapes of the Gurmukhi characters for different fonts and sizes, different sets of features were developed which are described below:

Primary Feature Set

The features used in Primary set such as number of junctions with the headline, presence of sidebar, presence of a loop and no loop formed with the headline are as following:

1. Number of junctions with the headline: Each character in Gurmukhi has either 1 or more than 1 junction with the headline .This feature has been used to divide the complete Gurmukhi character set into almost 2 equal sized subsets. This feature is true if the number of junctions is 1 else it is false.
2. Presence of sidebar**:** The presence or absence of sidebar is another very robust feature for classifying the characters. This feature is true if a vertical line is present on the rightmost side of the sub-symbol else it is false.
3. Presence of a loop**:** The presence of a loop in the sub-symbol is another important classification feature. The loop should not be formed along the headline.
4. No Loop formed with headline**:** This feature is true if the character is open at top along the headline or in other words if there is no loop containing headline as its subpart.

Secondary Feature Set

The second feature set, called secondary feature set, is a combination of local and global features, which are aimed to capture the geometrical and topological features of the characters and efficiently distinguish and identify the character from a small subset of characters

The Secondary Feature Set consists of following features:

Number of endpoints and their location: A black pixel is considered to be an end point if there is only one black pixel in its n x n neighbourhood in the resolution of the character image. In order to determine the position of an endpoint, the character image is divided into n x n equal zones .Using these zones, the position of the endpoints in terms of their positions in quadrants and their numbers are noted.

1. Number of junctions and their location:

A black pixel is considered to be a junction if there are more than two black pixels in its n x n neighbourhood in the resolution of the character image. The number of junctions as well as their positions in terms of n x n quadrants is recorded.

2. Aspect Ratio:

Aspect ratio which is obtained by dividing the sub-symbol height by its width was found to be very useful for classifying the sub-symbols lying in lower-zone.

2.3. Statistical Feature Set:

Statistical features[1] are computed by some statistical operations on image pattern and these include features like zoning, projection, histogram and distance etc.

1. Zone Density:

   In zoning, the character image is divided into N*M zones. From each zone, features are extracted to form the feature vector. By dividing the number of foreground pixels in each zone by total number of pixels in each zone, the density of each zone is obtained.

2. Projection histograms:

   Projection histograms count the number of pixels in specified direction. Three types of projection histograms-Horizontal, Vertical, Diagonal (left diagonal and right diagonal) can be used. The projection histograms are counted by counting the number of foreground pixels. In horizontal histogram these pixels are counted by row wise. In vertical histogram, the pixels are counted by column wise. In left diagonal histogram, the pixels are counted by left diagonal wise. In right diagonal histogram, the pixels are counted by right diagonal wise.

3. BDD Feature:

   In the background directional distribution (BDD), the neighboring background pixels are considered to the foreground pixels. To calculate the directional distribution values of background pixels for each foreground pixel, the masks for each direction are used. The pixel at center 'X' is considered as a foreground pixel to calculate directional distribution values of background. The weight for each direction is computed by using specific mask in particular direction depicting cumulative fractions of background pixels in particular direction.

V. CLASSIFICATION:

Each pattern having feature vector is classified in predefined classes using classifier. Classifier is first trained by a training set of pattern samples to prepare a model which is later used to recognize the test samples. K-Nearest Neighbor (K-NN) classifier uses the instance based learning by relating unknown pattern to the known according to some distance or some other similarity function [1]. It classifies the object by majority vote of its neighbour. Because it considers only neighbour object to a particular level, it uses local approximation of distance function. It means instance learning is used in K-NN. K specifies the number of nearest neighbors to be considered and the class of majority of these neighbors is determined as the class of unknown pattern. The distance function used to find nearest neighbors can be specified as Euclidean.

CONCLUSION

In this paper, we have discussed different feature extraction and classification schemes for handwritten Gurmukhi character recognition system. A very small set of easy to compute features has been used by G S Lehal ,Chandan Singh and Kartar Singh Siddharth for classification of characters. Artificial neural networks (ANN), Support Vector Machines (SVM), Structural pattern recognition, etc can be used for classification process. The neural networks have emerged as the fast and reliable tools for classification towards achieving high recognition accuracy.

REFERENCES

[1] Kartar Singh Siddharth, Mahesh Jangid, Renu Dhir Rajneesh Rani, "Handwritten Gurmukhi Character Recognition Using Statistical and Background Directional Distribution Features" Proceedings of et al./International Journal on Computer Science and Engineering,Vol.3 No.6 June 2011

[2] G. S. Lehal, C. Singh, " A shape based post processor for Gurmukhi OCR" ,Proceedings 6t International Conference on Document Analysis and Recognition, Seattle, USA, 2001 1105-1109.

[3] G S Lehal and Chandan Singh, "Feature extraction and classification for OCR of Gurmukhi script", Vivek, Vol. 12, No. 2, 1999, pp. 2-12.

[4] J. Mantas, "An overview of character recognition methodologies", Pattern Recognition, Vol. 19, pp 425-430 (1986).

[5] V. K. Govindan and A. P. Shivaprasad, "Character recognition – A survey ",Pattern Recognition, Vol. 23, pp 671-683(1990).

[6] B. Al-Badr and S.A. Mahmoud, "Survey and bibliography of Arabic optical text recognition", Signal Processing, Vol.41, pp. 49-77(1995).

Honeypot – A Security Solution

Neha choudhary

M.Tech, CSE,Sri Sai college pathankot

nehachoudharydaviet@gmail.com

Abstract—**There are a few security threats like viruses, worms etc over the internet.Tools are also available against these problems such as anti-virus software, firewalls and intrusion detection systems, but these systems cannot give us information about the attackers and the tools being used to attack ,they only react to or prevent against attacks. Now there is a way to detect attackers and the tool they use .Honeypots are used in computer and internet security. It can be deployed to attack and divert an attacker from their real targets. This review paper provides an overview, types and pros-cons of a honeypot.**

***Keywords*— Honeypot , viruses , firewall , security.**

I. Introduction

The first step to understand honeypot is to define what a honeypot is .A honey pot is a computer system on the Internet that is basically set up to attract and trap people who attempt to penetrate other computer systems..Unlike firewalls or intrusion detection systems (ids), honeypots do not work on a specific problem. They are flexible tool that are available in many shapes and sizes. They can perform many tasks from detecting encrypted attacks in IPv6 networks to capturing the latest in on-line credit card fraud. This flexibility gives honeypots their true power. This flexibility can also make them challenging to define and understand. Conceptually almost all honeypots work in a same way. They are a resource that do not have their authorized activity, they do not have any production value. Theoretically, a honeypot has no legitimate activity. This means any type of interaction with a honeypot is unauthorized or malicious activity. Any connection attempts to a honeypot are basically a probe, attack, or compromise. While this concept sounds very simple, because of this simplicity honeypots have tremendous advantages and few disadvantages.

In computer terminology, a honeypot is a trap setup to detect and deflect any type of attempts of unauthorized use of information systems. Basically it consists of a computer, data, or a network site that appear as the part of a network, but actually it is isolated and monitored, and which contains information and a resource of value to attackers.Honeypots are a security system whose value lies in being intruded or attacked. Honeypots are generally used to protect themselves from any type of identity theft. It is a internet-based server that acts as a decoy, luring potential hackers so that security specialists can monitor and study how system break-ins occur. Any prospective user of the honeypot technology should know about the numerous benefits a honeypot . Moreover, it is also important for users to understand the disadvantages or risks related with implementing the honeypot technology to one's network system. Other aspects such as how honeypots exactly work and other important information related with honeypots, should also be understood in order to calculate how one's network system will have benefits form this technology . Honey Pot Systems are decoy servers or systems setup to collect all information regarding an attacker or intruder into your system. It is important to know that Honey Pots do not replace other traditional Internet security systems; they are an additional level of security. Honey Pots can be setup inside or outside of a firewall design or even in all of the other locations. Most of the time it is deployed inside of a firewall for control purposes.

II. ADVANTAGES

Honeypots have very simple concept, which gives them some very powerful strengths and advantages. Followings are some of the important and useful advantages of honeypot technology :

A. Data value

One of the challenges the security specialists faces is gaining correct value from data. Organizations collect large amounts of data every day including firewall logs, system logs and other intrusion detection alerts. The amount of information collected is extremely difficult to derive any value from the data .On the other hand honeypot collect small amount of data and that data have high value. The general concept of honeypot of no expected production activity basically reduces the noise level. Instead of gathering gigabytes of data every day, the honeypots collect few megabytes of data per day, if even that much.

B. Simplicity

Simplicity is the biggest single advantage of honeypots. There is no need to develop algorithms , no large databases to maintain, no rule bases to configure. We have to take the honeypot, place it somewhere in our organization . Some of the honeypots like research honey-pots, can be more complex. Experienced security professionals can easily tell about the concept of simplicity and reliability of honeypots. With some complexities of honeypot technology ,breakdowns and failures can occur .Finally we conclude that honeypots are conceptually very simple.

C. Minimal resources

Honeypots require minimum resources because it only capture bad activity in our network system.

D. Encryption or IPv6

Most of the security technologies like IDS systems do not work in encrypted or IPv6 environments where as honeypots work very fine within these environments. Any type of attack is easily detected and captured by the honeypot.

E. Information

Honeypots can collect deep and relevant information than any other technology related with security of system .

III . DISADVANTAGES

With advantages we have disadvantages and honeypots also have a few. Honeypots can only track activity that interacts with it. It cannot capture attacks of other systems unless it interacts with the honeypot as well. Like other security technologies honeypots also have few risks related with them. Different types of honeypots consists of different levels of risk. Followings are some of the disadvantages of honeypot :

A. Narrow field of view

The biggest disadvantage of honeypots is that they have a very narrow field of view. They only see the activities which are directed against them. If an attacker breaks into your network system and make attacks on variety of systems, your honeypot will be totally unaware of the activity unless it is attacked directly. If the attacker has identified your honeypot he or she can easily avoid that system and can enter gradually in your organization. As we know that honeypots have a microscopic effect on the value of the data you collect and enables you to focus completely on data of known value .Like a microscope, the honeypot have very limited field of view.

B. Finger Printing

Another disadvantage of honeypots is fingerprinting. Fingerprinting is a situation in which an attacker can identify the true identity of a honeypot . Whenever an attacker get connected to a specific type of honeypot, the Web server responds it by flashing a common error message using standard HTML. This is the expected response for any type of web server. Basically the honeypot misspells one of the HTML commands, such as spelling the word security as securiyt. This type of misspelling becomes a fingerprint for the honeypot and attacker can quickly identify it .

C .Risk

The next disadvantage of honeypots is risk. They can include some risk to your environment. Here the word risk means that a honeypot once attacked can be used to attack or harm other systems. As we know that different honeypots have different levels of risk. Some consists of very little risk while the others give the attacker complete platform to launch new attacks. The simplicity of the honeypot introduce less risk. Risk vary from system to system and depends on how we build and deploy the honeypot.All security technologies introduce risks with them. Firewalls have risk of penetration and encryption. The biggest risk of IDS is failing to detect attacks. Like other technologies honeypots also have the risk of being trapped by the bad guy and being used to harm other systems.

IV. CATEGORIES OF HONEYPOT

Honeypots can be classified on the basis of the type of deployment and level of involvement. Based on deployment the honeypots can be classified as:

A. Production

Production honeypots are quite easy to use as they capture only limited information, and also used by several companies or corporations. Production honeypots are generally placed inside the production network along with the other production servers by an organization to improve their overall level of the security of system .Basically production honeypots are low-interaction honeypots and they are very easy to deploy. They give less information about the attacks or attackers as compared to research honeypots .

B. Research

Research honeypots generally gather information about the motives and tactics of the Blackhat community which usually target different systems over the networks. These type of honeypots do not add direct value to a specific organization. They only research the threats which the organizations face and learn how to better protect against various threats in the network. Research honeypots are really complex to deploy and maintain. They capture extensive information and are used by the research, military or government organizations.

V. TYPES OF HONEYPOTS

A. Honeyd: Low-interaction honeypot

Low interaction honeypots allow only little interaction for an attacker or malware. All services provided by the low interaction honeypots are emulated. Hence low interaction honeypots are not vulnerable and will not become easily affected by attacks .Honeyd is a low-interaction honeypot and developed by Niels Provos. Honeyd is open source setup and designed to work on Unix systems. Honeyd follow the concept of monitoring unused IP space. When it discovers a connection attempt to an unused IP, it stops the connection and then interacts with the attacker. By default honeyd discover and logs any type of connection to any UDP or TCP port. You can also configure other services to monitor ports like FTP server which monitor TCP port 21. When an attacker connects to the system the honeypot captures all of the attacker's interaction with the specific service. In the case of the FTP server we can gather the attacker's login and password but it all depends on the level of emulation of the honeypot.

Most emulated services works in the same manner. They possess a specific type of behaviour and then they are programmed to react in a predefined way. The main limitation is in some situations it does not know how to react. Most of the low-interaction honeypots simply flash an error message.

B. Honeynets: High-interaction honeypot

Honeypots make use of various vulnerable service or software. High-interaction honeypots are usually complex because they involve real operating systems and applications. In high Interaction honeypots everything is real. High Interaction honeypots provide a clear and detailed picture of how an attack execute in real-time. Since there is no additional services the high interaction honeypots works for identifying unknown vulnerabilities. But high interaction honeypots are very prone to infections and increases the risk because attackers can easily use these real honeypot operating systems to attack and to get into the system .Honeynets is the type of high-interaction honeypot. Honeynets are not a product and also not a software solution that we can install on a computer. Actually it is an architecture ,an entire network of computers which is specially designed to attack. The idea behind it is to have an architecture that creates a highly controlled network system .In this case all activity is controlled and captured continuously. Within this network we project our victims and real computers running real applications on the system. The attacker attacks and break into these systems on their own but they do not realize they are within a honeynet. All of their activities like encrypted SSH sessions , emails and files uploads are continuously captured without them knowing it. Basically this is done by inserting kernel modules on the victim systems in order to capture all of the attacker's actions. At the same time the honeynet also controls the activities of the attackers. All this is done using a honeywall gateway. This gateway controls the traffic of network using intrusion prevention technologies. It provide flexibility to the attacker to interact with the victim systems and also it prevents the attacker from harming other computer systems.

VI. VALUE OF HONEYPOT

A honeypot's biggest value lies in its simplicity. Any type of a connection sent to the honeypot is commonly a probe, scan, or even attack. Whenever a connection is created from the honeypot it means the honeypot was compromised. There is very small amount of production traffic going to or from the honeypot. Sometimes mistakes also happen such as an incorrect DNS entry or wrong IP address assigned. Generally most of the honeypot traffic represents unauthorized activities. We have three areas related to security as followings :

A. Prevention

If we want to stop the attackers we have to secure our house. Prevention is very similar like placing dead bolt locks on our doors. Prevention is the very first step for the network security. It keep the threat out of the system. There are various methods of preventing our system but the challenge is to adopt the best method.

B. Detection

We need to detect the attackers if they get through our various security measures. With the passage of time, prevention will fail. We need to detect when failure happens. Once again using the same house's example, it would be similar to putting a security alarm and motion sensors in the entire house zone .These alarms will work till someone breaks in it. If prevention fails, we need to act immediately and get alerted to threat as soon as possible .

C. Reaction

We need to react as soon as possible to the attackers once we detect them. Detecting the failure has little value till we correctly respond to it . If someone breaks into your house and sounds the burglar alarm the next step should be that local police force's reaction. The same method applies to information security once you detected a failure in the system, you must have an effective and immediate incident response.

VII. ETHICAL ISSUES CONCERNING HONEYPOTS

The use of honeypots in our computer system is somehow a controversial topic and the main question is how ethical are they really ? According to M.E. Kabay, author of 'liability and ethics of honeypots , "As far as entrapment is concerned this is not a legal problem and this does not mean that attackers is not unethical." Basically the argument is that if both are unethical and illegal in practice then why not we should prefer to use honeypot in our computer system.Other experts consider honeypots unethical as well as a disadvantage to the computer world because more and more hackers are getting aware of honeypots and working with them.On the other hand some system security specialists have their opinion that honeypots use the attack first, before being attacked. According to B. Scottberg, author of 'Internet Honeypots - Protection or Entrapment ?' "tracking something in a honeypot provide us invaluable insights into attacker techniques and ultimately the production systems will be better protected. We can easily learn of vulnerabilities before they are exploited." This is quite a valid support concerning the ethics of honeypot applications for organizations that use them in their network or computer systems.In many cases use of honeypot cannot be viewed as being unethical because of its various advantages. The article 'Combat Viruses' by Kurt Kleiner provide us with a view that honeypots have been known to contain and fight against computer viruses. In another article 'Using honeypots to fake out an attacker', the author Mark Edmead provide us with the most common and useful advantages of using honeypots in our security systems.Honeynet.org is an organization related to spreading awareness of the vulnerabilities that exist on the internet.

VIII. COMMERCIAL HONEYPOT SYSTEMS

We have a variety of commercial Honey pot systems available in market. The operating systems supported by honeypots are Microsoft NT and Unix. Some of the commercial Honey Pot systems available are:

A. Deception ToolKit (DTK)

It is a toolkit designed to provide defenders a list of orders of magnitude advantage against the attackers.

B . FakeBO

This program act as a fake trojan servers and blocks every attempt from client . It can send fake messages and replies back to the trojan client. The trojans supported are basically Back Orifice (BO) and NetBus.

C. CyberCop Sting by Network Associates

Basically this product is designed to run on Windows NT and is able to interact with several different systems including Linux, Solaris, Cisco IOS, and NT etc. It is a kind of appeal to the hackers for looking as it already has several well-known vulnerabilities.

VIII. CONCLUSION

This paper provides us knowledge about honeypots and their contributions to the security of the computer or network. A honey pot is a computer system on the Internet that is basically set up to attract and trap people who attempt to penetrate other computer systems. They can perform many tasks from detecting encrypted attacks in IPv6 networks to capturing the latest in on-line credit card fraud. Honeypots have very simple concept, which gives them some powerful strengths.Simplicity is the biggest advantage of honeypots. There is no need to develop algorithms , no large databases to maintain, no rule bases to configure. With advantages we have disadvantages of honeypot such as it has narrow field of view and it has various risks related with it. Honeypots is divided into two categories production (low interaction honeypot) and research (high interaction honeypot)

REFERENCES

[1] http://www.tracking-hackers.com/papers/honeypots.html

[2] http://searchsecurity.techtarget.com/definition/honey-pot

[3] http://en.wikipedia.org/wiki/Honeypot_%28computing%29

[4] http://students.kennesaw.edu/~lec6557/Formal%20Report.htm

[5] http://www.informit.com/articles/article.aspx?p=30489

[6] http://www.informit.com/articles/article.aspx?p=30489&seqNum=2

[7]http://homes.cerias.purdue.edu/~kaw/research/honeynet/HoneynetTutorial/honeypots/procon.html

[8] http://www.sans.org/security-resources/idfaq/honeypot3.php

[9] http://en.wikipedia.org/wiki/Honeypot_%28computing%29

[10] http://www.omnisecu.com/security/infrastructure-and-email-security/low-interaction-honeypots-and-high-interaction-honeypots.htm

[11] http://www.symantec.com/connect/articles/value-honeypots-part-one-definitions-and-values-honeypots

[12] http://www.symantec.com/connect/articles/value-honeypots-part-one-definitions-and-values-honeypots

[13] http://ethics.csc.ncsu.edu/abuse/hacking/honeypots/study.php

[14] http://www.123eng.com/forum/f7/honey-pot-technology-82589/

[15] http://www.seminartime.com/seminar_rpt/ct_rpt/honeypots/honeypots_17.php

Night Imagery Enhancement Algorithms for Vehicle Detection and Traffic Surveillance

[1]Amritpal Singh, [2]Vijay Kumar Banga, [3]Sartajvir Singh
[1] AP, Rayat Institute of Engineering & Information Technology, Railmajra, Distt. SBS Nagar
[2]Professor, Amritsar College of Engineering & Technology, Amritsar
[3]AP, Rayat Institute of Engineering & Information Technology, Railmajra, Distt. SBS Nagar

***Abstract* - The traditional night image enhancement algorithms enhance noise signal in image mainly in night image but now, new algorithms have been developed for night imagery enhancement. Designing image enhancement systems with good nature is a goal of all researchers. Modern night-vision systems like image intensifiers and thermal cameras enable operations at night and in adverse weather conditions. In this paper, we discussed newly developed methods for night image enhancement which includes mapping optimizes the match between the multiband image and the reference, and yields a night vision image with colors similar to the daytime image. Secondly, fast bright-object segmentation process has been introduced based on automatic multilevel histogram thresholding. Third method include colorizes the night vision imagery by using a combined framework consisting of hill-climbing algorithm for color-based segmentation, non-linear diffusion, region recognition and fuzzy based image fusion techniques. This paper also describes the future scope and required improvements in night imagery enhancement algorithms for vehicle recognition/detection and traffic surveillance in future.**

I. Introduction

Detecting and recognizing moving vehicles in traffic scenes for traffic surveillance, traffic control, and road traffic information systems is an emerging research area for Intelligent Transportation Systems. We can get information about moving vehicles from loop detectors, slit sensors, or cameras. Night vision cameras can provide more information for this system [1]. The interest in the colour display of night vision imagery is increased because of the availability of fused and multiband infrared and visual night vision systems. For observer to get complete information of the scene, nighttime imagery systems captures images in thermal and visual bands, thus provides complementary information of the observed scene which gives the maximum awareness. Image fusion is method which reduces the larger strength of information while at the same time it preserves all information from the source images [2].

There are various approaches available which are broadly divided into four different categories: pixel-based search [3], statistical approaches [4], region based approach [5] and pattern recognition based approach [6]. For a natural colorization of fused image Welsh et al [3] have proposed a method to transfer color characteristics to fused grayscale image from source images. Pitie et al [7], [8] proposed a color transfer algorithm based on N-Dimensional PDF transfer. A region-based approach for color transfer in night vision image sequences has been explored by Yang et al in [5]. Researchers have developed model and feature-based techniques to detect and track slowly moving or stationary vehicle and nighttime traffic scenes vehicles. Gardner and Lawton and Sullivan applied 3-D vehicle shape models to detect and track vehicles in traffic scenes. These can provide high accuracy in detecting vehicles in less crowded traffic scenes [1]. In recent years for colorizing night vision imagery using region-based color transfer method offers high perceptual quality color image [2].

However, under conditions of bad-illuminated at nighttime, the background scenes are substantially affected by lighting effect of moving vehicles, so that those obvious cues
of background models which are effective for vehicle detection during daytime become invalid. Thus, most of the
Frame-differencing based techniques may not work well under such nighttime traffic environments. At night, as well as under darkly illuminated conditions in general, the only salient features of moving vehicles are their headlights and taillights. Huang et al. proposed a method based on block-based contrast analysis and inter-frame change information. This contrast-based method can effectively detect outdoor objects in a given surveillance area using a stationary camera. This proposed system comprises fast bright-object segmentation process based on automatic multilevel histogram thresholding is performed to extract pixels of lighting objects from the captured image sequences of nighttime traffic scenes. These lighting objects are then grouped by a spatial

clustering process to obtain groups of vehicle lights of potential moving cars and motorbikes [9].

In this paper, we have discussed in first section about recent developed algorithms of Night Imagery Enhancement Algorithms for Vehicle Detection and Traffic Surveillance. There are three algorithms namely Coloring method using a colour lookup transformation, fast bright-object segmentation process based on automatic multilevel histogram thresholding, Novel region based color transfer method also with methodology block diagram.

II. RECENT DEVELOPED METHODS

This study presents an effective system for detecting and tracking moving vehicles in nighttime traffic scene for traffic surveillance. Many researchers have developed valuable methods in which most recent technique has been discussed here.

A) Coloring method using a colour lookup transformation

Hogervorst et al [16] proposed colouring method using a colour lookup transformation as shown in Figure 1 in which two bands are fed into the red and green channels of a colour image. All two-band sensor (red-green) combinations are fused. When the colour transformation used to transform the two-band image and it is applied to an image containing all possible sensor outputs. In this method the false colour (red-green) image is translated into an indexed image using all the colours of the inset. When the colour map is replaced by a colour table containing the colours in the inset of the coloured image arises.

This technique has two advantages as explained in [16]. Firstly, such a transformation is very efficient and can be implemented in real time. Secondly, the same colour scheme can be applied to a different image. This assures that the colour of objects no longer depends on the image content. Therefore, the object colour does not change under various operations, e.g. panning and zooming. There is another feature is the use of samples as the first feature is colour lookup transformations for this new colouring method.

They have used a set of samples for which both the multiband sensor values and the corresponding natural colours (RGB value). As a set of samples one can use a combination of a multiband sensor image and a naturally coloured image of the same scene, when the images are in correspondence.

In that case, each pixel represents a sample for which both the multiband sensor values and the natural colour are known. The description of this method worked as convert the multiband sensor image into a false-colour image by taking the individual bands as input to the Red and Green channels (and Blue when the sensor contains three bands). Then convert the resulting false colour red green-image into an indexed image by use of standard techniques. Each pixel in such an image contains a single index. The index refers to an RGB value in a colour lookup table (the number of entries can be chosen by the user). In the case of a sensor image consisting of two bands the colour lookup table contains various combinations of Red and Green values (the Blue-values are zero when the sensor or sensor pair provides only two bands). For each index the corresponding natural colour equivalent is derived by locating the pixels in the image with this index and finding the corresponding pixels in the (natural colour) reference image. Next, process involves calculation the average colour over this ensemble of pixels. One could simply take the average RGB-value. However, it's better to calculate the average in a perceptually de-correlated colour space and convert the result to RGB-space. This assures that the computed average colour reflects the perceptual average colour. Averaging automatically takes the distribution of the pixels into account: colours that appear more frequently are attributed a greater weight. Finally, the red-green colour table is replaced by the colour table with natural colours. This results in a two-band image with natural colours, in which the colours are optimized for this particular sample set.

B) A fast bright-object segmentation process based on automatic multilevel histogram thresholding (6191)

In this process, Yen-Lin Chen *el al* [1] proposed effective nighttime vehicle detection and tracking approach for identifying and classifying moving vehicles by locating and analyzing spatial and temporal features of their vehicle lights for traffic surveillance.

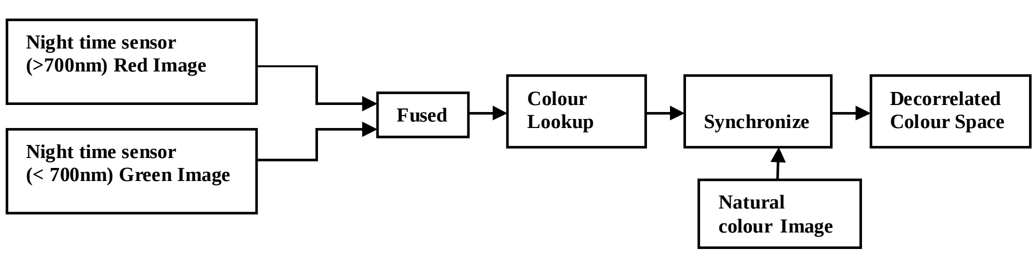


Figure1: Methodology for Coloring method using a colour lookup transformation

This system comprises the following processing stages [1]. First, a fast bright-object segmentation process based on automatic multilevel histogram thresholding is performed to extract pixels of lighting objects from the captured image sequences of nighttime traffic scenes. These lighting objects are then grouped by a spatial clustering process to obtain groups of vehicle lights of potential moving cars and motorbikes.

Next, a feature-based vehicle tracking and identification process is applied to analyze the spatial and temporal information from these potential vehicle light groups from consecutive frames, and to refine the detection results and correct for grouping errors and occlusions. Actual vehicles and their types can thus be efficiently detected and verified from these tracked potential vehicles to obtain the traffic flow information in the road scenes. Experimental results demonstrate that the proposed approach is feasible and effective for vehicle detection and identification in various nighttime environments for traffic surveillance.

This method involves [1] first task is to extract these bright objects from the road scene image to facilitate further rule-based analysis. To save the computation cost on extracting bright objects, we firstly extracted the gray scale image, i.e. the Y-channel, of the grabbed image by performing a RGB to Y transformation. For extracting these bright objects from a given transformed gray intensity image, pixels of bright objects must be separated from other object pixels of different illuminations. Thus, an effective multilevel thresholding technique is needed for automatically determining the appropriate number of thresholds for segmenting bright object regions from the traffic-scene image. For this purpose, Yen-Lin Chen *el al* have already proposed an automatic multilevel thresholding technique for image segmentation [15]. This technique can automatically decompose a grabbed road-scene image into a set of homogeneous thresholded images.

C) Novel region based color transfer method (7449)

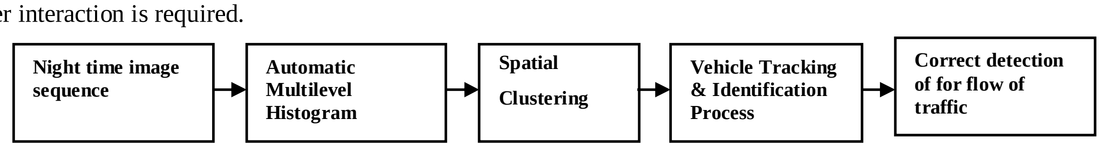
Zaveri et al [2] proposed a method, involves is to impart natural daytime colors to the grayscale fused images of night vision imagery by using a region-based approach. This method provides a new region-based framework for an optimized color transfer from natural color image to the grayscale fused night vision (NV) image. The block diagram of this method is shown in Fig. 3. This method is automatic and no user interaction is required.

The steps for algorithm as explained in [Zaveri] involves in this method shown in figure 2, First Pseudo-colored fused NV image is generated, by considering the Visible image, IR image and their averaged image, as input to the hybrid high boost filter false color fusion algorithm [11] then Color-based segmentation using hill-climbing algorithm [12] is applied to pseudo-colored NV image, which generates optimized number of regions of the NV image.

Visible and IR input images are fused using the fuzzy region feature based method [14], considering as input the optimum number of regions obtained in the previous step. Now, a natural color target image, with similar composition to that of the content in multiband NV images, is considered. The intensity band of the natural color target image is histogram matched to that of the gray scale fused image obtained in the previous step. Non-linear diffusion [12] is applied on the histogram matched natural color image, in order to reduce the number of colors.

Color-based segmentation using hill-climbing algorithm [13-14] is applied to the diffused natural color target image, and optimized number of regions are generated. Every region in the segmented source image (false colored NV image) and the segmented target image (natural color image) is represented by standard deviation of the intensity component of the region in the HSV color space. Corresponding to each region of source image, one region of target image is associated based on the matching value of standard deviation of the regions and color transfer is performed in HSV color space. The gray scale fused NV image is enhanced with adaptive histogram equalization. The resultant image is substituted in intensity band in the HSV color space. Finally, the HSV to RGB transformation is performed on the result image to produce the natural colored NV image. Thus this method offers an optimized framework for colorizing multiband night vision images.

This method combines the advantages of hill-climbing algorithm in subsequent methods like region recognition and fuzzy based image fusion techniques. The hill-climbing algorithm optimizes the performance of the fuzzy region feature based image fusion method and color transfer method, thus generating an optimized colored night vision image.



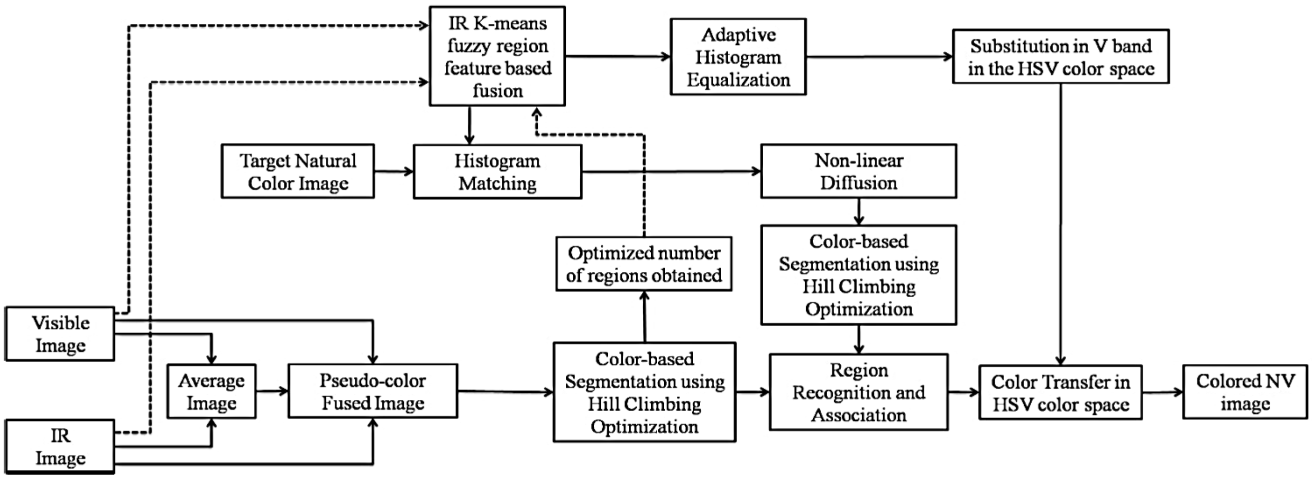


Figure 2: Methodology for Novel region based color transfer method [2]

CONCLUSION

These three recent developed methods have their own feature. In Colour method, the object colours remain invariant under panning operations and are largely independent of the scene content and also fusion method used in two prototype portable dual band real-time night vision systems. A fast bright-object segmentation process takes an average of 16 milliseconds processing time per frame. Novel region based color transfer method combines the advantages of hill-climbing algorithm in subsequent methods like region recognition and fuzzy based image fusion techniques. The hill-climbing algorithm optimizes the performance of the fuzzy region feature based image fusion method and color transfer method, thus generating an optimized colored night vision image. In future, there is need of combine the all features of different recent developed methods.

REFERENCES

[1] Yen-Lin Chen, Member, IEEE, Bing-Fei Wu, Senior Member, IEEE, Hao-Yu Huang, and Chung-Jui Fan "A Real-Time Vision System for Nighttime Vehicle Detection and Traffic Surveillance" vol. 58.

[2] Tanish Zaveri, Mukesh Zaveri, Ishit Makwana, Harshit Mehta. "An Optimized Region-based Color Transfer Method for Night Vision Application", 2010 International Conference on Signal and Image Processing

[3] T. Welsh, M. Ashikhmin, and K. Mueller. Transferring color togreyscale images. ACM Transactions on Graphics, vol. 21(3), pp. 277-280, July 2002.

[4] A. Toet. Natural colour mapping for multiband nightvision imagery. Information Fusion, vol. 4(3), pp. 155-166, 2003.

[5] B. Yang, F. Sun, and S. Li. Region-based color fusion method for visible and IR image sequences. Chinese Conference on Pattern Recognition (CCPR'08), pp. 1-6, October 2008.

[6] L. Ji, J. Ying. Pattern recognition based color transfer. Proceedings of the International Conference on Computer Graphics, Imaging and Visualization, pp. 55-60. IEEE Computer Society, July 2005.

[7] F. Pitie, A. Kokaram, R. Dahyot. Automated colour grading using colour distribution transfer. Computer Vision and Image Understanding, (2007).

[8] F. Pitie, A. Kokaram, R. Dahyot. N-Dimensional Probability Density Function Transfer and its Application to Colour Transfer. International Conference on Computer Vision (ICCV'05). Beijing, October\2005.

[9] Yen-Lin Chen1, Bing-Fei Wu2,*, and Chung-Jui Fan2 "Real-time Vision-based Multiple Vehicle Detection and Tracking for Nighttime Traffic Surveillance"

[10] Z. Yheng, J. Song, W. Zhou, R. Wang, False color fusion for multiband SAR images based on contourlet transform. ACTA Automatica Sinica, vol. 33(4), April 2007.

[11] M. V. J. Weickert, B.M. Haar Romeny. Efficient and reliable schemes for nonlinear diffusion filtering. IEEE Transactions on Image Processing, vol. 7, pp. 398-410, 1998.

[12] T. Ohashi, Z. Aghbari, and A. Makinouchi. Hill-climbing algorithm for efficient color-based image segmentation. IASTED International Conference on Signal Processing, Pattern Recognition, and Applications, pp. 17-22, July 2003.

[13] R. Achanta, F. Estrada, S. Susstrunk. Salient region detection and segmentation. Proceedings of the 6th international conference on Computer vision systems, pp. 66-75, 2008.

[14] S. S.-y. LIU Gang, JING Zhong-liang. Multiresolution image fusion scheme based on fuzzy region feature. Zhejiang University - Science and Technology, vol. 7(2), pp. 117-122, February 2006.

[15] B.-F. Wu, Y.-L. Chen, and C.-C. Chiu, "A discriminant analysis based recursive automatic thresholding approach for image segmentation," IEICE Trans. Info. Systems, vol. E88-D, no.7, pp.1716-1723, 2005.

[16] Hogervorst, M.A. , Soesterberg, "Presenting nighttime imagery in daytime colours", formation Fusion, 2008 11th International Conference.

Morphological Edge Detection Algorithms on Medical Images

Pankaj Sharma[1] and Ratinder Kaur[2]
[1]Lecturer, Department of ECE, CTIT, Jalandhar
[2]Lecturer, Department of ECE,, SIET, Manawala, Amritsar
Email: er.pankajsharma@rocketmail.com, er.ratinderkaur@gmail.com

***Abstract* — Medical images edge detection is an important work for object recognition of the human organs and it is an important pre-processing step in medical image segmentation and 3D reconstruction. Conventionally, edge is detected according to some early brought forward algorithms such as gradient-based algorithm and template-based algorithm, but they are not so good for noise medical image edge detection. In this paper, basic mathematical morphological theory and operations are introduced at first, and then a novel mathematical morphological edge detection algorithm is proposed to detect the edge of lungs CT image with salt-and-pepper noise. The experimental results show that the proposed algorithm is more efficient for medical image denoising and edge detection than the usually used template-based edge detection algorithms and general morphological edge detection algorithms.**

Keywords- Medical image, edge detection, mathematical morphology, denoising.

I. INTRODUCTION

Medical images edge detection is an important work for object recognition of the human organs such as lungs and ribs, and it is an essential pre-processing step in medical image segmentation. The work of the edge detection decides the result of the final processed image. Conventionally, edge is detected according to some early brought forward algorithms like Sobel algorithm, Prewitt algorithm and Laplacian of Gaussian operator , but in theory they belong to the high pass filtering, which are not fit for noise medical image edge detection because noise and edge belong to the scope of high frequency.In real world applications, medical images contain object boundaries and object shadows and noise. Therefore, they may be difficult to distinguish the exact edge from noise or trivial geometric features. Mathematical morphology is a new mathematical theory which can be used to process and analyze the images. It provides an alternative approach to image processing based on shape concept stemmed from set theory not on traditional mathematical modeling and analysis. In the mathematical morphology theory, images are treated as sets, and morphological transformations which derived from Minkowski addition and subtraction are defined to extract features in images. As the performance of classic edge detectors degrades with noise, morphological edge detector has been studied. In this paper, a novel mathematical morphology edge detection algorithm is proposed to detect lungs CT medical image edge. It is a better method for edge information detecting and noise filtering than differential operation, which is sensitive to noise. And it is a better compromise method between noise smoothing and edge orientation, but the computation is more complex than general morphological edge detection algorithms.

II. MATHEMATICAL MORPHOLOGICAL OPERATION

Mathematical morphology is developed from set theory. It was introduced by atheron as a technique for analyzing geometric structure of metallic and geologic samples. It was extended to image analysis Based on set theory; mathematical morphology is a very important theory, whose operation must be defined by set arithmetic. Therefore, the image which will be processed by mathematical morphology theory must been changed into set. Mathematical morphology uses structuring element, which is characteristic of certain structure and feature, to measure the shape of image and then carry out image processing. Based on set theory, mathematical morphology is the operation which transforms from one set to another. The aim of this transformation is to search the special set structure of original set. The transformed set includes the information of the special set structure and the transformation is realized by special structuring element. Therefore, the result is correlative to some characteristics of structuring element. The basic mathematical morphological operators are dilation and erosion and the other morphological operations are the synthesization of the two basic operations. In the following, we introduce some basic mathematical morphological operators of grey-scale images. Erosion is a transformation of shrinking, which decreases the grey-scale value of the image, while dilation is a transformation of expanding, which increases the grey-scale value of the image. But both of them are sensitive to the image edge whose grey-scale value changes obviously. Erosion filters the inner image while dilation filters the outer image. Opening is erosion followed by dilation and closing is dilation followed by erosion. Opening generally smoothes the contour of an image, breaks narrow gaps. As opposed to opening, closing tends to fuse narrow breaks, eliminates small holes, and fills gaps in the contours. Therefore, morphological operation is used to detect image edge, and at the same time, denoise the image.

III. PROPOSED MORPHOLGICAL EDGE DETECTION ALGORITHMS

Morphological edge detection algorithm selects appropriate structuring element of the processed image and makes use of the basic theory of morphology including erosion, dilation, opening and closing operation and the synthesization operations of them to get clear image edge. In the process, the synthesized modes of the operations and the feature of structuring element decide the result of the processed image. Detailedly saying, the synthesized mode of the operations reflects the relation between the processed image and origin image, and the selection of structuring element decides the effect and precision and the result.

Therefore, the keys of morphological operations can be generalized for the design of morphological filter structure and the selection of structuring element. In medical image edge detection, we must select appropriate structuring element by texture features of the image. And the size, shape and direction of structuring element must been considered roundly. Usually, except for special demand, we select structuring element by 3×3 square. The effect of erosion and dilation operations is better for image edge by performing the difference between processed image and original image, but they are worse for noise filtering. As opposed to erosion and dilation, opening and closing operations are better for filtering. But because they utilize the complementarity of erosion and dilation, the result of processed image is only correlative with the convexity and concavity of the image edge. Accordingly, what we get is only the convex and concave features of the image by performing the difference between processed image and original image, but not all the features of image edge.

In this paper, a novel mathematical morphology edge detection algorithm is proposed. Opening-closing operation is firstly used as preprocessing to filter noise. Then smooth the image by first closing and then dilation. The perfect image edge will be got by performing the difference between the processed image by above process and the image before dilation.

IV EXPERIMENTAL RESULT AND ANALYSIS

In this section, the proposed morphological edge detection algorithm is compared with a variety of existing methods for edge detection. Fig.1 is the original lungs CT image with salt-and-pepper noise. Fig.2 and Fig.3 are the results of processed lungs CT images after respectively applying Laplacian of Gaussian operator and Sobel edge detector. Fig.4 and Fig.5 are the lungs CT images processed by morphological gradient operation and dilation residue edge
Detector.

Fig.1. Original lungs CT image with salt-and- pepper noise.

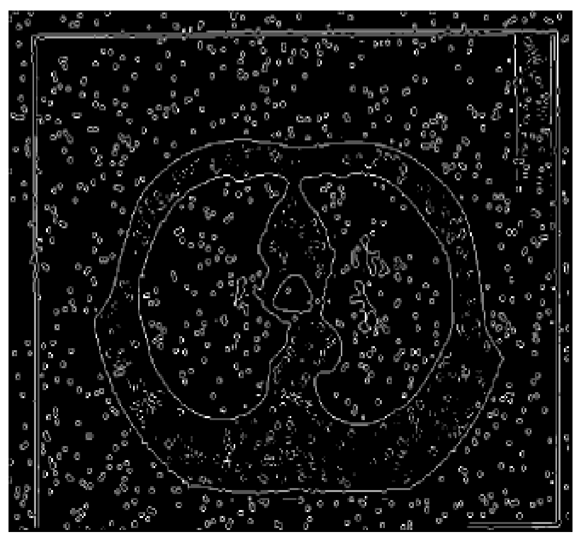


FIG.2. Lungs CT image processed by Laplacian of Gaussian operator.

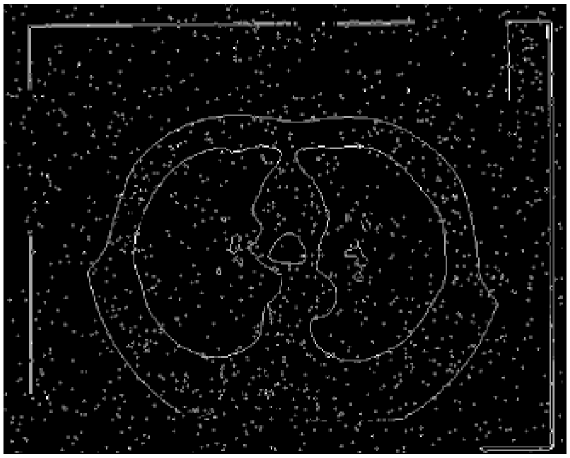


Fig.3. Lungs CT image processed by Sobel detector

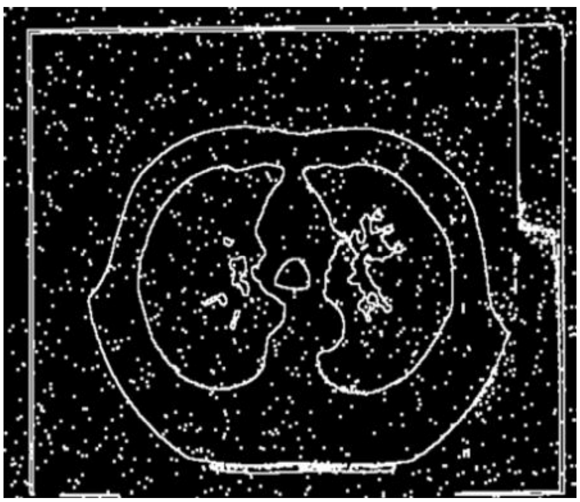


Fig.4. Lungs CT image processed by Morphological Gradient Operation.

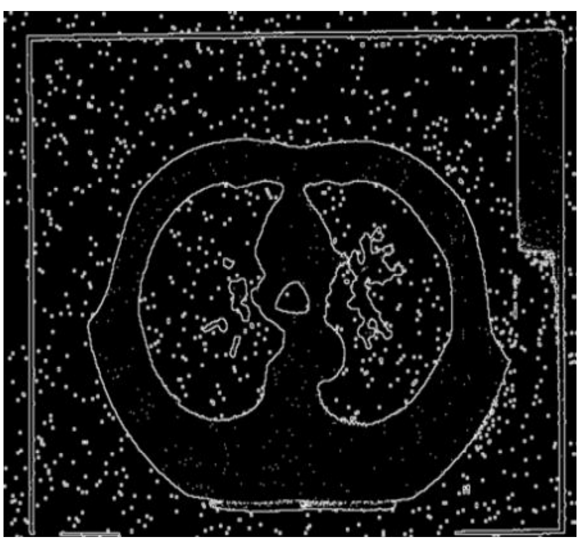


Fig.5. Lungs CT image processed by dilation residue edge detector.

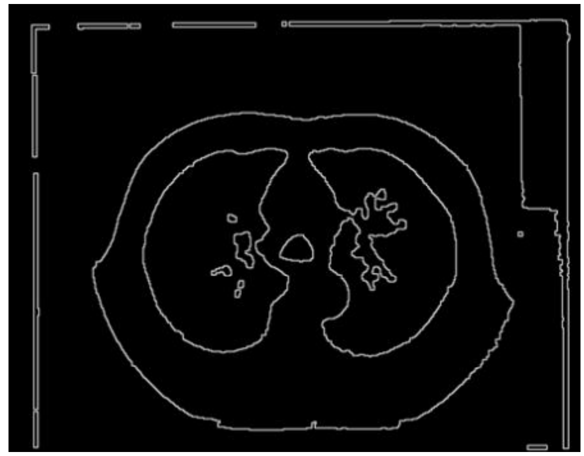


Fig.6. Lungs CT image processed by novel morphological edge detector.

According to the experiment results shown in Fig.2 and Fig.3, Laplacian of Gaussian operator and Sobel edge detector detect the lungs edges successfully, but Sobel edge detector fail to detect the outer edge of body, and both of them can't filter the noise. By Fig.4 and Fig.5, the morphological gradient operation and dilation residue edge detector are succeed in lungs and body edges detection, and the detected edges are clearer than the edges detected by Laplacian of Gaussian operator and Sobel edge detector. But both of them fail to filter the noise in despite of the latter is better for noise filtering than the former. By Fig.6, the novel morphological edge detector proposed in this paper is succeed in lungs and body edges detection, but more important than template-based edge detection algorithm and general morphological edge detection algorithm mentioned before, it also filters the noise successfully.

CONCLUSION

In this paper, a novel mathematic morphological algorithm is proposed to detect lungs CT medical image edge. The experimental results show that the algorithm is more efficient for medical image denoising and edge detecting than the usually used template-based edge detection algorithms such as Laplacian of Gaussian operator and Sobel edge detector, and general morphological edge detection algorithm such as morphological gradient operation and dilation residue edge detector.

REFERENCES

[1] M.I. Rajab, M.S. Woolfson, and S.P. Morgan, "Application of region-based segmentation and neural network edge detection to skin lesions," Computerized Medical Imaging and Graphics, vol. 28, pp.61–68, 2004.

[2] H. Tang, E.X. Wu, Q.Y. Ma, D. Gallagher, G.M. Perera, and T. Zhuang, "MRI brain image segmentation by multi-resolution edge detection and region selection," Computerized Medical Imaging and Graphics, vol 24, pp. 349–357, 2000.

[3] Huertas, A. and Medioni, G., "Detection of intensity changes with sub pixel accuracy using Laplacian-Gaussian masks," IEEE Trans. On Pattern Analysis and Machine Intelligence, PAMI, vol. 8, pp. 651–664, 1986.

[4] Maragos P., "Differential Morphology and Image Processing," IEEE Trans Image Processing, vol. 5, pp. 922–937, June 1996.

[5] Ortiz F, and Torres F, "Vectorial Morphological Reconstruction for Brightness Elimination in Colour Images," Real-Time Imaging, vol. 10, pp. 379–387, June 2004.

[6] Jing Xiao-jun, Yu Nong, and Shang Yong, "Image Filtering Based on Mathematical Morphology and Visual Perception Principle," Chinese Journal of Electronics, vol. 13, pp. 612–616, April 2004.

[7] Richard A P, "A New Algorithm for Image Noise Reduction Using Mathematical morphology," IEEE Transaction on Image Processing, vol. 4, pp. 554–568, March (1995).

[8] Chen T., and Wu Q.H., Rahmani-Torkaman R., Hughes J., "A pseudo top-hat mathematical morphological approach to edge detection in dark regions," Pattern Recognition, vol. 35, pp. 199-210, 2002.

Simulation Based Analysis of Routing Protocols in Wireless Ad-Hoc Networks

Neha[1] and Tanu Preet Singh[2]
[1]Lecturer, Department of CSE, SIET, Amritsar, India
[2]Associate Professor, Department of CSE, ACET, Amritsar, India
Email: reetu.gaggel1204@gmail.com, sanjeev_be@indiatimes.com

Abstract—Wireless Ad-hoc networks are basically self organizing, self configuring and peer to peer multi-hop mobile wireless networks in which the information packets are transmitted in a 'store and forward' manner from a source to an arbitrary destination via intermediate nodes. The main objective of this paper is to analyze the performance of various ad-hoc networks routing protocols viz. DSDV (Destination Sequence Distance Vector), DSR (Dynamic Source Routing) and AODV (Ad-hoc On Demand Distance Vector by using the Network Simulator Version 2(NS-2). This paper presents a performance analysis of these three proactive and reactive routing protocols based on metrics such as throughput, packet delivery ratio and average end-to-end delay.

Keywords- Destination Sequence Distance Vector (DSDV), Dynamic Source Routing (DSR), Ad-hoc On Demand Distance Vector (AODV), Network Simulator (NS-2).

1. INTRODUCTION

Wireless networking is an emerging technology that allows user to access information and services electronically, regardless of their geographic position. The Wireless networks are classified as infrastructure networks and infrastructure less (ad-hoc) networks. The need for Internet access through mobile devices, anywhere and anytime, has caused the development of model which is different in comparison to access based on a previously set fixed infrastructure over which wireless devices connect to the Internet nowadays. MANET is a collection of wireless mobile nodes that communicate with each other using multi-hop wireless links without any existing network infrastructure or centralized administration [1]. Each node in the network behaves as a router and forwards packets for other nodes. Routing as an act of moving information from a source to a destination through intermediate nodes, is a fundamental issue for networks. Numerous widely used routing algorithms are proposed for wired networks. Routing is mainly classified into static and dynamic routing. Static routing refers to routing strategies set in the router, manually or statistically. Dynamic routing refers to routing strategies learned by an interior or exterior routing protocol [2] **.** Apart from that with the increase of portable of devices as well as progress in wireless communication, Ad-hoc network gaining importance with the increasing number of widespread application. The following point shows the importance of ad hoc networks [3,4]:

*Instant Infrastructure***:** Unplanned meetings, spontaneous interpersonal communications etc., cannot rely on any infrastructure, it needs planning and administration. It would take too long to set up this kind of infrastructure; therefore ad-hoc connectivity has to setup.

Disaster Relief: Infrastructure typically breakdown in disaster area. Hurricanes cut phone and power lines, floods destroy Base stations, fires burn servers. No forward planning can be done, and the set-up must be externally fast and reliable. The same applies to many military activities, which are, to be honest, one of the major driving forces behind mobile ad-hoc networking research.

Effectiveness: Service provided by existing infrastructure might be too expensive for certain applications. If, for example only connection oriented cellular network exist, but an application sends only small status information every other minute, cheaper ad-hoc packet-oriented network might be a better solution. Registration procedure might take too long and communication overheads might be too high with existing networks. Tailored ad- hoc networks can offer a better solution.

Remote Areas: Even if infrastructure could be planned ahead, it is sometimes too expensive to set up an infrastructure in sparsely populated areas. Depending on the communication pattern, so ad-hoc networks or satellite infrastructure can be a solutions learned by an interior or exterior routing protocol [5].

Many routing protocols have been proposed for MANETs, but none of them has good performances in all scenarios with different network sizes, traffic loads, and node mobility patterns [6][7**].** Each of the proposed protocols is based on different principles and has different characteristics, so their classification is necessary. Usually, classification is made based on characteristics related to the information which is exploited for routing and roles which nodes may take in the routing process. The most popular classification method is based on how routing information is acquired and maintained by mobile nodes. This paper presents a performance evaluation of three prominent routing protocols in MANET based on results analysis obtained by running simulations with different scenarios in Network Simulator version 2 (NS-2) [8]. A description of considered routing protocols is given in Section 2. Simulation based analysis are described in Section 3. Simulation results are presented in Section 4 and finally Section 5 gives the conclusion of this paper.

2. ROUTING PROTOCOLS

The following three routing protocols are considered in this paper:

2.1 Ad Hoc On-demand Distance Vector (AODV)

The AODV [9] routing protocol is an “on demand” routing protocol, which means that routes are established when they are required. This routing protocol is based on transmitting Route Reply (RREP) packets back to the source node and routing data packets to their destination. Used algorithm consists of two steps: route discovery and route maintenance. Route discovery process begins when one of the nodes wants to send packets. That node sends Route Request (RREQ) packets to its neighbors. Neighbors return RREP packets if they have a corresponding route to destination. However, if they don’t have a corresponding route, they forward RREQ packets to their neighbors, except the origin node. Also, they use these packets to build reverse paths to the source node. This process occurs until a route has been found. Routing tables which only have information about the next hop and destination are used for routing information maintenance. When a route link disconnects, for example, a mobile node is out of range, neighbor nodes will notice the absence of this link. If so, neighbor nodes will check whether there is any route in their routing tables which uses a broken link. If it exists, all sources that send traffic over the broken link will be informed with Route Error (RRER) packet. A source node will generate a new RREQ packet, if there is still a need for packet transmission.

2.2 Dynamic Source Routing (DSR)

Routing protocol DSR [10] uses explicit source routing,Which means that each time a data packet is sent, it Contains the list of nodes it will use to be forwarded. In other words, a sent packet contains the route it will use. Routes are stored in memory, and data packets contain the source route in packet header. Mechanism allows nodes on route to cache new routes, and also, allows source to specify the route that will be used, depending on criteria. This mechanism, also, avoids routing loops. If a node has to send a packet to another one, and it has no route, it initiates a route discovery process. This process is similar to AODV route discovery process. In other words, the network is being flooded with RREQ packets. Each node that receives a RREQ packet, broadcasts it, except for destination node or nodes that have a route to destination node in their memory. A route through network is built by RREQ packet, and RREP packet is being routed backward to the source. A route that returns RREP packet is cached on the source node for further use. There can be multiple RREP packets on one RREQ packet. In DSR, when a broken link is detected, a RRER packet is sent backward to the source node. After receiving a RRER packet, a source node initiates another route discovery operation. Additionally, all routes containing the broken link should be removed from the route caches. This protocol aggressively uses source routing and route caches.

2.3Destination Sequenced Distance Vector (DSDV)

The DSDV[11][12] routing protocol is a proactive routing protocol. It is based on the Bellman-Ford routing algorithm. Each node in the network maintains a routing table which contains all available destinations with associated next hop towards destination, metric and destination sequence number. Sequence number presents improvement of DSDV routing protocol compared to distance vector routing, and it is used to distinguish stale routes from fresh ones and avoid formation of route loops. Routing tables are updated by exchanging the information between mobile nodes. Each node periodically broadcasts its routing table to its neighbors. Broadcasting of the information is done in Network Protocol Data Units (NPDU) in two ways: a full dump and an incremental dump. A full dump requires multiple NPDUs, while the incremental requires only one NPDU to fit in all the information. A receiving node updates its table if it has received a better or a new route. When an information packet is received from another node, node compares the sequence number with the available sequence number for that entry. If the sequence number is larger, entry will be updated with the routing information with the new sequence number, whereas if the information arrives with the same sequence number, metric entry will be required. If the number of hops is smaller than the previous entry, new information will be updated. Update is performed periodically or when a significant change in the routing table is detected since the last update. If network topology frequently changes, a full dump will be carried out, since an incremental dump will cause less traffic in a stable network topology. Route selection is performed according to the metric and sequence number criteria. The sequence number is also the time indication that destination node sends, allowing routing table update. If we have two identical routes, the route with a larger sequence number will be saved and used, and the other will be destroyed.

3.SIMULATION BASED ANALYSIS

3.1 Simulation Tool

In this paper, the simulation of AODV, DSDV and DSR routing protocols is done by using network simulator (NS-2) software due to its simplicity and availability. NS is a discrete event Simulator targeted at networking research NS provides substantial support for simulation of TCP, routing and multicast routing protocols over a wired and wireless network. NS-2 is written in C++ and OTCL. C++ for data per event packets and OTCL are used for periodic and triggered event [13]. NS-2 includes a network animator called nam animator which provides visual view of simulation. NS-2 preprocessing provides traffic and topology generation and post processing provide simple trace analysis. AWK programming is used for trace file analysis[14].

3.2 Performance Metrics:

The following performance metrics are used in this paper for the performance evaluation of AODV, DSDV and DSR routing protocols:

a) Throughput

It is the amount of data transferred over the period of time expressed in bits per second or bytes per second.

b) Packet delivery ratio

It is the ratio of the number of data packets received by the destination node to the number of data packets sent by the source mobile node. It can be evaluated in terms of percentage (%)

c) Average End-to-End Delay

The delay of a packet is the time it takes the packet to achieve the destination after it leaves the source. The average packet delay for a network is obtained by averaging over all packets and all source destination pairs.

4. SIMULATION RESULTS

The simulation results are shown in the following section in the form of comparative graphs. In this paper an attempt has been made to compare the performance of three well known routing protocol DSDV, AODV, and DSR according to the simulation results. The simulation results are generated through the simulation tool NS-2 with the following parameters:

Parameter	Value
Radio model	TwoRay Ground
Protocols	DSDV,AODV,DSR
Traffic Source	Constant Bit Rate
Packet size	512 bytes
Max speed	10 m/s
Area	500 x 500
Number of nodes	50
Application	FTP
MAC	Mac/802_11
Simulation time (Sec)	20, 40, 60, 80 & 100

Table 1. Simulation Parameters

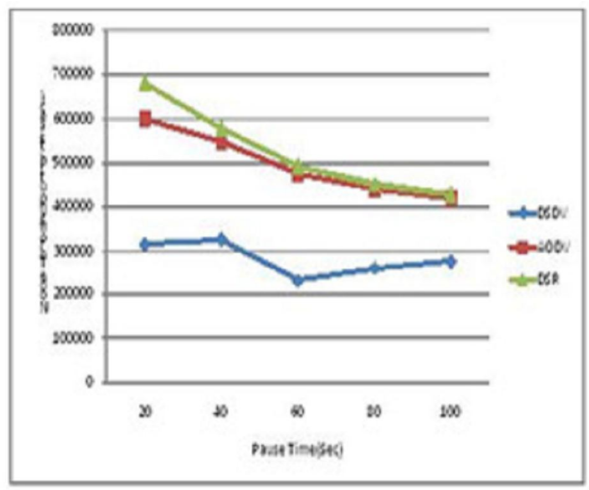


Figure 1. Analysis of Throughput

The Figure 1 depict the throughput for the DSDV, AODV and DSR protocols for the number of nodes 50.Throughput is the ratio of the total amount of data that reaches a receiver from a sender to the time it takes for the receiver to get the last packet. When comparing the routing throughput by each of the protocols, DSR has the high throughput. Based on the above simulation results, the throughput value of DSDV increases initially and reduces when the time increases. The throughput value of AODV slowly increases initially and maintains its value when the time increases. AODV performs well than DSDV since AODV is an on-demand protocol. The throughput value of DSR increases at lower pause time

and grows as the time increases. Hence, DSR shows better performance with respect to throughput among these three protocols.

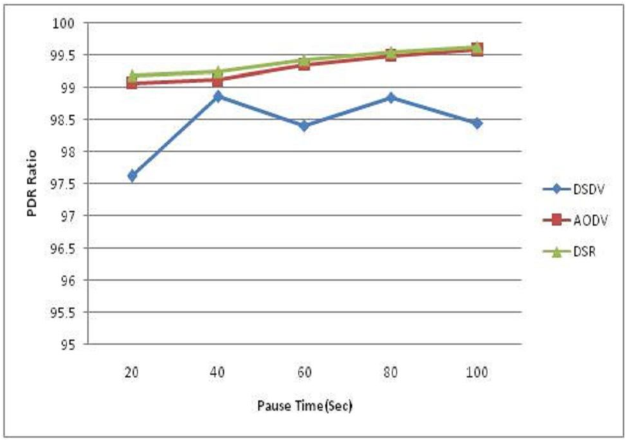


Figure 2. Analysis of PDR

The Figure 2 depict the average Packet Delivery Ratio (PDR) for the DSDV, AODV and DSR protocols for the number of nodes 50.PDR is the ratio between the number of packets transmitted by a traffic source and the number of packets received by a traffic sink. It measures the loss rate as seen by transport protocols and as such, it characterizes both the correctness and efficiency of ad-hoc routing protocols.

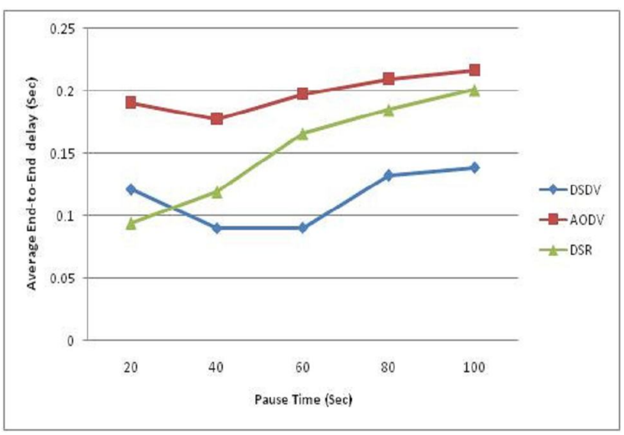


Figure 3. Analysis of Average End-to-End Delay

A high packet delivery ratio is desired in any network. As packet delivery ratio shows both the completeness and correctness of the routing protocol and also measure of efficiency the PDR value of AODV is higher than all other protocols. The PDR values of DSR and AODV are higher than that of DSDV. The PDR value of DSDV is worse in lower pause time and gradually grows in higher pause time. From the above study in view of packet delivery ratio reliability of AODV and DSR protocols is greater than DSDV protocol.

The Figure 3 depict the Average End-to-End delay for the DSDV, AODV and DSR protocols for the number of nodes 50.The packet Average End-to-End delay is the average time that a packet takes to traverse the network. This is the time from the generation of the packet in the sender up to its reception at the destination's application layer and it is measured in seconds. It therefore includes all the delays in the network such as buffer queues, transmission time and delays induced by routing activities and MAC control exchanges. The End-to-End delay is a measure of how well a routing protocol adapts to the various constraints in the network and represents the reliability of the routing protocol. It is clear that DSDV has the shortest End-to-End delay than AODV and DSR, because DSDV is a proactive protocol i.e. all routing informations are already stored in table. Hence, it consumes lesser time than others. On average case, DSR shows better performance than AODV but worse than DSDV. As AODV needs more time in route discovery, it produces more End-to-End delay. From the above study on End-to-End delay, DSDV has high reliability than AODV and DSR.

CONCLUSION

The design of the routing protocols are driven by specific goals and requirements based on respective assumptions about the network properties or application .In this paper, the performance analysis of DSDV, AODV and DSR routing protocols is done through the simulation tool NS-2 which gives the knowledge how to use routing schemes in dynamic network. The comprehensive simulation results of Average End-to-End delay, throughput, and packet delivery ratio over the routing protocols DSDV, DSR and AODV by varying network size, simulation time have been done. DSDV is a proactive routing protocol and suitable for limited number of nodes with low mobility due to the storage of routing information in the routing table at each node. Comparing DSR with DSDV and AODV protocol, byte overhead in each packet will increase whenever network topology changes since DSR protocol uses source routing and route cache. Hence, DSR is preferable for moderate traffic with moderate mobility. As AODV routing protocol needs to find route by on demand, End-to-End delay will be higher than other protocols. DSDV produces low end-to-end delay compared to other protocols. When the network load is low, AODV performs better in case of packet delivery ratio but it performs badly in terms of average End-to-End delay and throughput. Overall in the analyzed scenario, it is found that DSR outperforms AODV because it has less routing overhead when nodes have high mobility considering the above said three metrics. However, there are many other challenges to be faced in routing protocols design. A central challenge is the development of the dynamic routing protocol that can efficiently find routes between two communication nodes.

REFERENCES

[1] S. Shah, et al., “Performance Evaluation of Ad Hoc Routing Protocols Using NS2 Simulation,” *Proceedings of the National Conference on Mobile and Pervasive Computing (CoMPC-2008)*, Chennai, India, August 2008.

[2] K. Gorantala, “Routing Protocols in Mobile Ad Hoc Networks,” Master Thesis, Department of Computing Science, Umeøa University, Sweden, June 2006.

[3] C. Siva Ram Murthy and B. S. Manoj, “Ad Hoc Wireless Networks, Architectures and Protocols”, Second Edition, Low price Edition, Pearson Education, 2007.

[4] Mobile Ad Hoc Networking: An Essential Technology for Pervasive Computing Jun-Zhao *Sun MediaTeam, Machine Vision and Media Processing Unit, 2002.*

[5] Mehran Abolhasan, Tadeusz Wysocki and Eryk Dutkiewicz. “A review of routing protocols for mobile ad hoc networks” *Technical report, Telecommunication and Information Research Institute, University of Wollongong, Wollong, NSW 2522. Motorola Australia Research Centre, 12 Lord St., Botany,* NSW 2525, Australia, 2003.

[6] Azzedine Boukerche “Algorithm and Protocols for wireless and Mobile Ad-hoc networks”

[7] Mehran Abolhasan, Tadeusz Wysocki and Eryk Dutkiewicz ,” A review of routing protocols for mobile
Ad- hoc networks”, Elsevier 2003.

[8] Network Simulator version 2 (NS-2) webpage Available at: http://www.isi.edu/nsnam/ns, accessed at May 1st, 2010.

[9] C. E. Perkins, E. Belding-Royer, and S. Das, “Ad Hoc On-demand Distance Vector (AODV) Routing,” Experimental RFC 3561, Internet Engineering Task Force (IETF), July 2003.

[10] D. Johnson, Y. Hu, and D. Maltz, “The Dynamic Source Routing Protocol (DSR) for Mobile Ad Hoc Networks for IPv4,” Experimental RFC 4728, Internet Engineering Task Force (IETF), February 2007.

[11] C. E. Perkins and P. Begat, “Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers,” *Proceedings of the ACM SIGCOMM‘94 Conference on Communications Architectures, Protocols, and Applications*, London, United Kingdom, August 1994, pp. 234-244.
[12]C.Perkins ,Praving Bhagwat, “Highly dynamic destination sequenced distance vector routing (DSDV) for Mobile computers”.

[13] C.E. Perkins and E.M. Royer, “Ad-hoc on Demand Distance Vector Routing Protocols”, *in proceeded 2nd IEEE work shop mobile computing system and applications,* new ore leans, Loss Angelis, pp 90-100, February 2011.

[14] Md Shohidul Islam, Md Naim Hider, Md. Touhidul Haque, Leton miah “An Extensive Comparison among DSDV, DSR and AODV Protocols in MANET, “*International journal of computer application (0975-8887) Volume 15-no.2*, February 2011.

WiMAX – An Innovative Technology toward Wireless Broadband System : A Review Paper

Amandeep Kaur[1] and Sanjeev Kumar[2]
[1]Lecturer, Department of Electronics & Communication Engineering, RIET, Ropar, India
[2]Associate Professor, Department of Electronics & Communication Engineering, ACET, Amritsar, India
Email: aman2aries@gmail.com, sanjeev_be@indiatimes.com

Abstract: The IEEE 802.16 Working Group is the IEEE group for wireless metropolitan area network. The IEEE 802.16 standard defines the Wireless MAN (metropolitan area network) air interface specification (officially known as the IEEE WirelessMAN* standard). This wireless broadband access standard could supply the missing link for the "last mile" connection in wireless metropolitan area networks. Today there is a strong demand for broadband multimedia services and therefore it is important that a technology provides high bit rates to the end users. These high bit rates need to be provided to the end user in a wireless manner any time, in any location.

Keywords: Broadband Access, WiMAX , IEEE-802.16e,

I. INTRODUCTION

Interoperability for Microwave Access) is poised to become an important player of fixed, portable and mobile data networks. Wimax is an implementation of the IEEE 802.16d 2004 standard now that uses orthogonal frequency division multiplexing (OFDM) for optimization of wireless data services. The OFDM technology transmits the information by assigning small subcarriers (KHz) to end users depending on the radio frequency conditions [1]. This achieves high spectral efficiency and makes WiMAX networks very well suited to high speed data connections for fixed and mobile users. This paper shows experimental results with throughput links of up to 47 Mbps approximately for UDP data transmission, for short distances between the base station (BS) and the end user's terminal (SS) something that decreases as the distance increases. At the physical level, the throughput can be as high as 72 Mbps. Such a high throughput can be used to support a number of applications, including 'last mile' broadband connections, hot spot and cellular backhaul, and high speed enterprise connectivity for businesses. A parameter of a wireless link that needs to betaken care, especially for high bit rate throughputs as that offered by Wimax technology, is the BER parameter. This paper shows methods of setting a wireless link between a BS and an SS in order to achieve low BER. These methods can be applied by adjusting the transmitted and the received power of either the BS or the SS for different modulations (QPSK, 16QAM and 64QAM).

WiMax is supported by the industry itself, including Intel, Dell, Motorola, Fujitsu, AT&T, British Telecom, France Telecom, Reliance Infocomm, Siemens, Sify, PriceWatehouseCoopers and Tata Teleservices – forming an alliance called WiMax Forum. It represents the next generation of wireless networking. WiMAX original release the 802.16 standard addressed applications in licensed bands in the 10 to 66 GHz frequency range. Subsequent amendments have extended the 802.16 air interface standard to cover non-line of sight (NLOS) applications in licensed and unlicensed bands in the sub 11 GHz frequency range. Filling the gap between Wireless LANs and wide area networks, WiMAX-compliant systems will provide a cost-effective fixed wireless alternative to conventional wire-line DSL and cable in areas where those technologies are readily available. And more importantly the WiMAX technology can provide a cost-effective broadband access solution in areas beyond the reach of DSL and cable. The ongoing evolution of IEEE 802.16 will expand the standard to address mobile applications thus enabling broadband access directly to WiMAX-enabled portable devices ranging from smartphones and PDAs to notebook and laptop computers. Fig. 1 below from the WiMAX Forum summarizes the 802.16 standards.

Completion Date	802.16 Dec 2001	802.16a/ 802.16REVd 802.16a: Jan 2003 802.16Revd: Q3 2004	802.16e 2005
Spectrum	10 to 66 GHz	< 11 GHz	< 6 GHz
Channel Conditions	Line-of-Sight only	Non-Line-of-Sight	Non-Line-of-Sight
Bit Rate	32 to134 Mbps	75 Mbps max 20-MHz channelization	15 Mbps max 5-MHz channelization
Modulation	QPSK 16QAM 64QAM	OFDM 256 subcarrier QPSK 16QAM 64QAM	Same as 802.16a
Mobility	Fixed	Fixed	Pedestrian mobility Regional roaming
Channel Bandwidths	20, 25 and 28 MHz	Selectable between 1.25 and 20 MHz	Same as 802.16a with uplink subchannels
Typical Cell Radius	1 to 3 miles	3 to 5 miles (30 miles max based on tower height, antenna gain, and power transmit)	1 to 3 miles

Figure 1. Summary of 802.16 Standards

Telecommunications has grown at a tremendous rate in the last ten to twenty years. Improved semiconductor and electronics manufacturing technology, and the growth of the internet and mobile telecommunications have been some of the factors which have fueled this growth in telecommunications. The deployment of state of the art telecommunications infrastructure and

services has however been restricted to the developed world. The least developed countries have been left in the technological dark ages with few or none of the next generation networks installed. Developed countries now boast high speed connections with a large percentage of homes having access to the internet and broadband services at an affordable fee. The underdeveloped countries are yet to enjoy such facilities. This is referred to as the digital divide [1]. During the first World Summit on the Information Society (WSIS) held in Geneva in December 2003, the Digital Divide was defined as the unequal access to Information and communication Technologies (ICTs), where the least developed countries are separated from the developed countries because of a lack of technology particularly information and communication technology [2].

The digital divide has persisted due to the relatively high cost of putting up modern telecommunications infrastructure. This is compounded by the fact that there are a number of different services available and each service requires its own technology and network [3]. Therefore existing technologies such as Wireless Fidelity (WiFi), Digital Subscriber Line (DSL), Global System for Mobile communications (GSM), Integrated Services Digital Network (ISDN), and the relatively new 3G technologies have not been able to provide a total solution to closing the digital divide. Fig. 2 illustrates the main network types and the prevalent technologies associated with each, mapped against usage models and access modes.

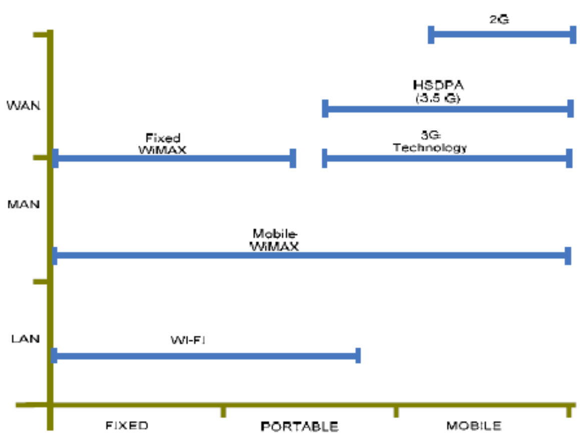


Figure2. Wireless network type and range.

MAN - Metropolitan Area Network (Citywide, Rural Area)
LAN - Local Area Network (Office, Home, Campus)
WAN- Wide Area Network (countrywide, International)

WiMax earned an important seal of approval recently when the Radio communication sector of the International Telecommunication Union (ITU-R) certified it as a 3G (third-generation) mobile data technology. Fig.3 shows the standard history for 802.16.

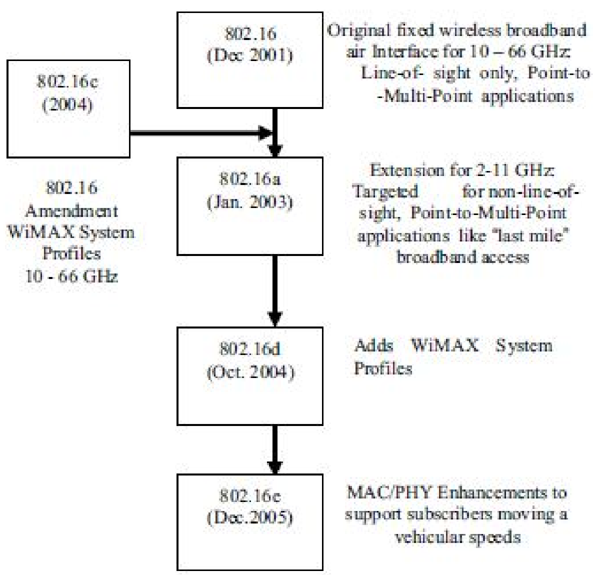


Fig 3: Standard Evolution.

WiMAX will have a larger impact long term than we have seen from cellular phones in the past two decades. Initial rollouts of WiMAX will begin mostly by competitive local phone service carriers and rural Internet service providers. Larger carriers will utilize fixed WiMAX to deliver services to residential customers many of whom are in underserved markets. WiMAX adoption in these underserved markets will be high due to lack of availability of high-speed data access. These deployments will generate capital to be reinvested for future deployments. Larger customer base will begin driving both the cost of carrier and customer equipment down. As the economy of scale makes deployment less expensive mobile platforms will begin to appear. This development will be spread between high population centers and the rural markets that already have fixed platforms deployed. Fixed platform will act as a springboard for mobile deployment. Then interconnections will begin to form between rural markets and metropolitan markets as carriers form cooperative agreements to share network resources. The economy of scale will increase exponentially at this point and we will notice a negative impact on traditional cellular, Internet and voice services. Once the implementation of initial hot underserved rural markets and high-density metro areas are completed, springboard deployments will quickly take WiMAX coverage to the level of coverage offered by traditional wireless today. This process will move much faster than the deployment of cellular networks and devices for the following key reasons:

- The manufacturing process for WiMAX devices will be quite similar to that of wireless devices and mostly the changes will be in components and software.
- Readiness of the current wireless fixed and mobile market and waiting on new technology.
- As carriers built out wireless networks, most of the questions in this field have been answered and can now be applied to the development of a mirror network that provides WiMAX access.

II. NETWORK ARCHITECTURE

WiMAX has a flexible architecture. The Mobile WiMAX End-to-End network architecture is based on an All-IP platform, all packet technology and no circuit switch telephony.

The open IP architecture gives network operators great flexibility when selecting solutions that work with legacy networks or that use the most advanced technologies, and in determining what functionality they want their network to support. They can choose from a vertically integrated vendor that provides a turnkey solution or they can pick and choose from a dense ecosystem of best-of-breed players with a more narrow focus. The architecture allows modularity and flexibility to accommodate a broad range of deployment options such as small scale to large scale, urban, suburban and rural coverage, mesh topologies , flat , hierarchical and their variant, and finally, co-existance of fixed , nomadic portable and mobile usage models [4].

Mobile WiMAX adds both the mobility and Multiple Input Multiple Output (MIMO) functionalities to the IEEE 802.16-2004 standard. It is one of two standards adopted by the WiMAX forum with the other one being the IEEE 802.16 – 2004. Mobile WiMAX network architecture mainly has three components. These include the Access Services Network (ASN), the Core Services Network (CSN) and the Application Services Network(AS). Fig. 4 illustrates the interconnection of these networks. The WiMAX network supports the following key functions:

- All IP Access and core service networks
- Support for fixed, nomadic and mobile access
- Interoperability with existing networks via internetworking functions
- Open interfaces between ASN's and between the ASN and the CSN
- Support for differential quality of service depending on the application
- Unbundling of the Access, core and application
- service networks

A. ASN

The ASN is the access network of WiMAX and it provides the interface between the user and the core service network. Mandatory functions as defined by the WiMAX forum include the following:

- Handover
- Authentication through the proxy authentication, authorization and accounting (AAA) server
- Radio resource management
- Interoperability with other ASN's
- Relay of functionality between CSN and MS,

e.g. IP address allocation.

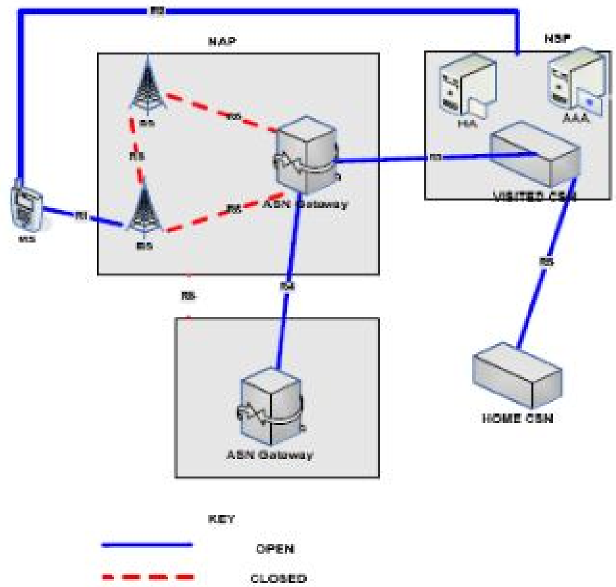


Figure 4. WiMAX Network Architecture [5].

Base Station (BS). The cell equipment comprises the basic base station equipment, radio equipment and a base station link to the backbone network. The base station is what actually provides the interface between the mobile user and the WiMAX network. The coverage radius of a typical base station in urban areas is around 500 to 900 meters [6]. In rural areas the operators are planning cells with a radius of 4 kilometer (Km). This is quite a realistic number now and quite similar to the coverage areas of GSM and UMTS/HSDPA base stations today.

Deployment is driven either by the bandwidth required to meet demand, or by the geographic coverage required to cover the area. Based on the cell planning of other previous technologies, urban and suburban segments cell deployment will likely be driven by capacity. Rural segment deployment will likely be driven by the cell radius.

ASN Gateway. The ASN Gateway performs functions of connection and mobility management and inter-service provider network boundaries through processing of subscriber control and bearer data traffic. It also serves as an Extensible Authentication Protocol (EAP) authenticator for subscriber identity and acts as a Remote Authentication Dial in User Service (RADIUS) client to the operator's AAA servers.

B. Core Services Network

The CSN is the transport, authentication and switching part of the network. It represents the core network in WiMAX. It consists of the home agent (HA) and the AAA system and also contains the IP servers, gateways to other networks i.e. Public Switched Telephone Network (PSTN), and 3G.

WiMAX has five main open interfaces which include; reference points R1, R2, R3, R4 and R5 interface [7].

The R1 interface interconnects the subscriber to the base station in the ASN and is the air interface defined on the physical layer and Medium Access Control (MAC) sub layer. The R2 is the logical interface between the mobile subscriber and the CSN. It is associated with authorization, IP host configuration management, services management, and mobility management. The R3 is the interface between the ASN and CSN and supports AAA, policy enforcement and mobility management capabilities. The R4 is an interface between two ASN's. It is mainly concerned with coordinating mobility of Mobile Stations (MS's) between different ASN's. The R5s an interface between two CSN's and is concerned with internetworking between two CSN's. It is through this interface that activities such as roaming are carried out.

The unbundling of WiMAX divides the network based on functionality. The ASN falls under the Network Access Provider (NAP). The NAP is a business entity that provides WiMAX network access to a Network Service Provider (NSP). The NSP is a business entity that provides core network services to the WiMAX network and consists of the CSN. The Applications services fall under the Applications Services Provider (ASP).

III. APPLICATIONS

The WiMAX standard has been developed to address a wide range of applications. Based on its technical attributes and service classes, WiMAX is suited to supporting a large number of usage scenarios. Table I address a wide range of applications [23].

TABLE I. SUMMARY OF WIMAX APPLICATIONS

CLASS DESCRIPTION	REAL TIME	APPLICATION TYPE	BANDWIDTH
Interactive Gaming	Yes	Interactive Gaming	50-85 Kbps
VoIP, Video Conferencing	Yes	VoIP	4-64 Kbps
		Videophone	32-384 Kbps
Streaming Media	Yes	Music/Speech	5-128 Kbps
		Video clips	20 – 384 Kbps
		Movies streaming	> 2 Mbps
Information Technology	No	Instant Messaging	< 250 byte messages
		Web browsing	> 500 Kbps
		Email (with attachments)	> 500 Kbps
Media Content download (store and forward)	No	Bulk data, Movie download	> 1 Mbps
		Peer to Peer	> 500 Kbps

VOIP & IP

Mobile WiMAX is an all IP network. The use of OFDMA on the physical layer makes it capable of supporting IP applications. It is a wireless solution that not only offers competitive internet access, but it can do the same for telephone service. Voice over Internet Protocol (VoIP) offers a wider range of voice services at reduced cost to subscribers and service providers alike .VoIP is expected to be one of the most popular WiMAX applications.

Its value proposition is immediate to most users. While WiMAX is not designed for switched cellular voice traffic as cellular technologies as are CDMA and WCDMA, it will provide full support for VoIP traffic because of QoS functionality and low latency. IPTV enables a WiMAX service provider to offer the same programming as cable or satellite TV service providers. IPTV, depending on compression algorithms [24], requires at least 1 Mbps of bandwidth between the WiMAX base station and the

subscriber. In addition to IPTV programming, the service provider can also offer a variety of video on demand (VoD) services. IPTV over WiMAX also enables the service provider to offer local programming as well as revenue generating local advertising.

IV BENEFITS OF WIMAX

WiMAX is a global technology. Different countries refer to their systems by different names for example; WiBro is the name of 802.16e standard in South Korea and HIPERMAN(High Performance Radio Metropolitan Are Network) in Europe. The Widely used international broadband spectrum range is 3.5 GHz. The followings are some of the advantages of WiMAX. By using a WiMAX system, companies/ residents no longer have to rip up buildings or streets or lay down expensive cables.

High Bandwidth. WiMAX can provide shared data rates of up to 70 Mb/s. this is enough bandwidth to support more than 60 businesses at once with T1-type connectivity. It can also support over a thousand homes at 1 Mb/s DSL-level connectivity Also, there will be a reduction in latency for all WiMAX communications.

Long Range. The most significant benefit of WiMAX compared to existing wireless technologies is the range. WiMAX has a communication range of up to 40 Km [25].

Multi-Application. WiMAX uses the Internet protocol and is therefore capable of efficiently supporting all multimedia services from VoIP to high speed internet and video transmission. It also supports a differentiated quality of service enabling it offer dynamic bandwidth allocation for different service types. WiMAX has the capacity to deliver services from households to small and medium enterprises, small office, home office (SOHO), Cybercafés, Multimedia Tele-centers, Schools and Hospitals.

Flexible Architecture. WiMAX supports several systems architectures, including Point-to-Point, Point-tomultipoint, and ubiquitous coverage.

High Security. The security of WiMAX is state of the art. WiMAX supports advanced encryption standard triple data encryption standard. WiMAX also has built-in VLAN support, which provides protection for data that is being transmitted by different users on the same base station. Both variants use Privacy Key Management (PKM) for authentication between base station and subscriber station. WiMAX offers strong security measures to thwart a wide variety of security threats.

QoS. WiMAX can be dynamically optimized for a mix of traffic that is being carried.

Multi Level Service. QoS is delivered generally based on the service level agreement between the end user and the service provider.

Interoperability. WiMAX is based on international, vendor-neutral standard. This protects the early investment of an operator since it can select the equipments from different vendors.

Low Cost and Quick deployment. WiMAX requires little or no external plant construction compared with the deployment of wired solutions. Base stations will cost under $20,000 but will still provide customers with T1-class connections [26].

Worldwide Standardization. WiMAX is developed and supported by the WiMAX forum (more than 470 members). The WiMAX forum collaborates with different international standards organizations that are developing broadband wireless standards with the intent to provide interoperability among the standards. Some of the other broadband wireless standards include HiperMAN/HiperLAN (Europe) and WiBRO (South Korea). These standards are compatible with WiMAX at the physical layer. WiMAX will become a truly global technology based standard for broadband and will guaranty interoperability, reliability and evolving technology and will ensure equipment with very low cost.

V. DRAWBACKS OF WIMAX

Broadband wireless in general and WiMAX in particular face a number of challenges that could impede their adoption in the marketplace. The most significant challenge is that WiMAX is a new technology with emerging support.

Hesitancy. Companies are very hesitant of setting up WiMAX base stations today since it has not yet reached widespread use. Intel has made their Centrino laptop processors WiMAX enabled. All laptops are expected to have WiMAX by 2008[27].

Exclusion of Start-Up Companies. Even though cost provides a low barrier to entry, none of the startup companies are projected to be major players in the development of WiMAX. Intel and Cisco seem to have an obvious advantage today, and by the time it reaches widespread use, large operators will find WiMAX to be a very attractive new way of raising revenues.

Research and Development. For WiMAX to succeed, new products must be researched and developed to incorporate WiMAX. Without the help of major companies investing in this R&D, WiMAX could be gravely underutilized.

CONCLUSION

Broadband wireless is a significant growth marketplace for the telecom industry to deliver a variety of applications and services to both mobile and fixed users. The combination of both advanced radio features and flexible end-to-end architecture makes WiMAX attractive solution for diverse operators. It provides many different services on one network, services which required different networks in the past. It also provides convergence of fixed and mobile networks. It provides high speed access to the subscriber at a reasonable cost, thereby enabling the service provider to make a profit from the technology, using economies of scale. It offers the advantage of reduced total cost of ownership during the lifetime of a network deployment.

REFERENCES

[1] Gerard Smyth, “Wireless Technologies and e-Learning: Bridging the Digital Divide”, Intel Corporation, December 2006.

[2] G. Cayla, S. Cohen and D. Guigon, “WiMAX an Efficient Tool to Bridge the Digital Divide,” WiMAX Forum**,** November, 2005, pp. 2.

[3] Saifur Rahman and Manisa Pipattanasomporn,” Alternate Technologies for Telecommunications and Internet Access in Remote Locations”. In Proc. 2002 3rd Mediterranean Conference and Exhibition on Power Generation, Transmission, Distribution and Energy Conversion, Greece, November 2002.

[4] L. Bai,” Analysis of the Market for WiMAX Services”,Thesis, Lyngby, Denmark, May 2007

[5] P. Yegani, Cisco Systems white paper, “WiMAX Overview,” IETF-64 Nov. 7-11, Vancouver, Canada, 2005, pp. 4.

[6] Bruno Puzzolante, Gianni Redaelli and Gianni Grazia,“Nationwide Implementation of a WiMAX Mobile Access Network” at CEFRIEL, STEM.

[7] Loutfi Nuaymi, WiMAX Technology for Broadband and Wireless Access, J. Wiley, New York, 2007.

[8] M. Paolini, “Mobile WiMAX: The Best Personal Broadband Experience!” WiMAX forum, June 2006. pp. 6.

[9] Flarion, “OFDM for Mobile Data Communication”, August 2004.

[10] G. Parsaee, and A. Yarali, “OFDMA for the 4th Generation Cellular networks,” (CCECE 2004), Niagara falls, Canada. May 2-5, 2004.

[11] A. Salvekar, S. Sumeet, L. Qinghua, V. Minh-Anh and Q. Xiaoshu, “Multiple-Antenna Technology in WiMAX Systems,” Intel Technology Journal, Volume 08 Issue 03, August 20, 2004. pp. 235.

[12] G. Nair, J. Chou, T. Madejski, K. Perycz, D. Putzolu, and J. Sydir, “IEEE 802.16 Medium Access Control and Service Provisioning,” Intel Technology Journal, Volume 08 Issue 03, August 20, 2004. pp. 217.

[13] White paper, “A comparative Analysis of Mobile WiMAX Deployment Alternatives in the Access Networks”, May 2007, WiMAX Forum.

[14] White Paper,” IEEE 802.16a Standard and WiMAX Igniting Broadband Wireless Access”, WiMAX Forum.

[15] G. Parsaee, A.Yarali, E. Ebrahimzad, “MMSE-DFE Equalizer Design for OFDM Systems with Insufficient Cyclic Prefix”, Proceeding of IEEE VTC04 conference on Sept. 26-29, 2004, Los Angeles, Ca.

[16] C. Hoymann, M. Dittrich, S. Goebbels and B. Walke, “Dimensioning Cellular WiMAX Part II: Multihop WiMAX Networks”, in Proceedings of IEEE mobile WiMAX, Mar 2007.

[17] A.Yarali, J. Wheatley, J. Ponder, M. Sharifi, H. Behrooz, “Wireless Cellular and PCS Network Cell Count Estimator”, ASEE conference at E Peoria, IL March 26, 2004.

[18] White Paper,” Consideration for Deploying Mobile WiMAX at Various Frequencies”, Nortel.

[19] V. Erceg, “Channel Models for Broadband Fixed Wireless Systems”, IEEE 802.16.3c-2000/47.

[20] White paper, “Understanding WiMAX and 3G for Portable/Mobile Broadband Wireless”, Intel, December 2004.

[21] Abdulrahman Yarali, Bwanga Mbula, Ajay Tumula,” WiMAX: A Key to Bridging the Digital Divide”, Proceedings of IEEE SoutheastCon 2007, Rcihmond, VA.

[22] Christian Hoymann, and Stephan Goebbels,“Dimensioning Cellular WiMAX Part I: Single HopNetworks”, in Proceedings of 13th European Wireless conference, 2007.

[23] D. Gray, “Mobile WiMAX – Part II: A Comparative Analysis,” May, 2006. pp. 14 - 41.

[24] White paper,” WiMAX Deployment Considerations for Fixed Wireless Access in the 2.5 GHZ and 3.5 GHZ Licensed Bands”, June 2005, WiMAX Forum.

[25] A. Yarali and A. Cherry, “Internet Protocol Television (IPTV)”, Proceeding of IEEE TENCON05, Melbourne, Australia, Nov. 21-24, 2005.

[26] M. Sauter, “WiMAX Base Station Prices and Coverage Ranges “, March 31, 2007.
http://mobilesociety.typepad.com/mobile_life/2007/03/wimax_base_stat.html

[27] “WiMAX Weekly News: Weekly Clip Report”, April 26th -May 3rd, 2007, WiMAX Forum.

Review on Different Techniques for Heart Diseases Detection

Sanjeev Kumar[1], Gursimranjeet kaur[2]
[1]AP, Department of ECE, ACET, Amritsar
[1]Lecturer, Department of ECE, GIMET, Amritsar
sanjeev.btech@gmail.com , gursimranvirk@yahoo.com

***Abstract* - The medical field is very versatile field and one of the interested research areas for the scientist. It deals with many medical disease problems starting with the diagnosis of the disease, preventing from the disease and treatment for the disease. There are various types of medical disease and accordingly various types of treatment methods. The recognition of heart disease from diverse features or signs is a multi-layered problem that is not free from false assumptions and is frequently accompanied by impulsive effects. And thus the work is challenging to predict the heart diseases with less number of attributes. There are many methods that are used for analysing the heart diseases. Here we are discussing First is by using neural network in which by using two different methods we can predict the heart analysis one is by using Backpropagation Neural Network (BPNN) and Radial Basis Function Neural Network (RBFNN) for training and testing and second is by using neural network in which data warehouse is pre-processed in order to make it suitable for the mining process. The second method is by using Fuzzy System In which risk, smoke, cholesterol, blood pressure, diabetes, sex and age are main risk factors that affect on heart disease risk. The data set is first made and then it describe the input variables with their membership functions. After that the output variable with its membership functions including rules.**

I. INTRODUCTION

Health care awareness and technology developments have led to huge number of hospitals and health care centers. But still, quality of health care service at affordable cost is still a challenging issue in developing countries[3]. Nowadays the use of computer technology in the fields of medicine area diagnosis, treatment of illnesses and patient pursuit has highly increased. Despite the fact that these fields, in which the computers are used, have very high complexity and uncertainty and the use of intelligent systems such as fuzzy logic, artificial neural network and genetic algorithm have been developed[3]. The past research in the medical field shows the huge requirement of the expert system for the diagnosis of the disease [1]. One problem is the right treatment to the patient is only provided if the diagnosis of disease is done accurately and on the right time In domain of heart disease risk, smoke, cholesterol, blood pressure, diabetes, sex and age are main risk factors that affect on heart disease risk [3]. Because of the many and uncertain risk factors in the heart disease risks, sometimes heart disease diagnosis is hard for experts.

The artificial neural network is one way to diagnosis the disease the multilayer-perceptron model is one of the basic for of neural network and its network architecture forms by the neurons consists in the input, hidden and output layer. The numbers of neural network models are available from which we use in this paper the BPNN and RBFNN[1].

The data mining techniques: clustering and frequent pattern mining. The heart disease data warehouse consists of mixed attributes containing both the numerical and categorical data. These records are cleaned and filtered with the intention that the irrelevant data from the warehouse would be removed before mining process occurs. Then clustering is performed on the preprocessed data warehouse using K-means clustering algorithm with K value so as to extract data relevant to heart attack. Subsequently the frequent patterns significant to heart disease diagnosis are mined from the extracted data using the MAFIA algorithm. The significant weightage is calculated for each frequent pattern using the approach proposed [2].

The other accurate tool that we are discuusing here is the fuzzy logic. Motivated by the need of such an important tool, in this study, we designed an expert system to diagnose the heart disease.

1.By Using Neytral And Modular Network

The artificial neural network is one way to diagnosis the disease the multilayer-perceptron model is one of the basic for of neural network and its network architecture forms by the neurons consists in the input, hidden and output layer. The numbers of neural network models are available from which we use in this paper the BPNN and RBFNN[1]. The BPNN architecture consists of input, hidden and output layer. In RBFNN the weights according the error at the output layer cause by the difference in the actual output and target output. In the RBFNN similar kind of architecture as of the BPNN is formed. It also has the three layers. The ability of the monolithic neural network is less due to problem in handling the large dataset. Due to this we use the modular neural network divide the attributes among the different modules. The modular neural network gives better performance as compare to single neural network model many cases.

Backpropagation Neural Network:

Backpropagation Neural Network: It is the one of the more focused area of research. The basic architecture consists of input, hidden and output layers. In this the weights are updated according to the error at the output layer due to the difference in between the actual output and target output.

Radial Basis Function Network:

RBFNN have similar network architecture as of BPNN is shown in the fig. 1. The radial basis function neural network consists of neuron at the input layer and output layer which shows the input vector and desired output at the output layer. Biological nerves system adjusts to a narrow range of local area of frequencies. The feed forward architecture consists by the neural network with non-linear units and an output layer with linear units An RBFNN consists of single hidden layer whereas the MLP consists of more than one hidden layer. The RBFNN consist non-linear functions at the hidden layer and linear functions at the output layer. This makes the RBFNN functionally different at the hidden layer and output layer.

Modular Neural Network (MNN):

The overall architecture of proposed system is shown in fig. 3.The analysis of the problem is done to make a solution. The whole problem is big and complex to handle out[1].

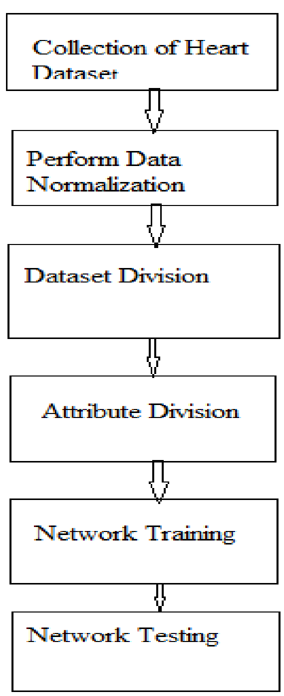


Fig: General Experimental Methodology

The division of the task into number of sub task makes it elementary, simpler and less complex task. There results obtained from different modules are integrated to provide obtain the final results. The conventional neural network is change into the modular neural network if it is divided into number of modules which works independently and results of them are integrated by the integration unit. The task of integration unit is to integrate the results from the different modules and gives the final output result. Some of the methods use for the integration are winner take all, voting, average, and fuzzy inference system .

The dataset is collected from the uci machine learning repository which is divided into 70% training and 30 % testing dataset. The attributes are divided among the four modules. The attribute 1, 2, 3, 4, 5, 6 and7 represents the age, sex, chest pain type, resting blood pressure, serum cholesterol in mg/dl, fasting blood pressure sugar which is more than 120 mg/dl, and resting electrocardiographic results of patient and attributes 8,9,10,11,12 and 13 represents the maximum heart rate achieved, exercise induced angina, old peak = ST depression induced by exercise relative to rest, the slop of the peak exercise ST segment, number of major vessels (0-3) colored by fluoroscopy, thal: 3 = normal; 6 = fixed defect; 7 = reversible defect.

The dataset is divided two parts of 7 attributes and 6 attributes. The feature space is complex when the whole problem is used as compare to the feature space of the sub-problem which having limited number of attributes which is less complex. All the four modules are work independently and produce the results which are further use by the integration unit. The attribute 1 to 6 is given to the module 1 and module 2. The remaining attributes from 8 to 13 are given to the module 3 and module 4. The two types of method is use for the integration one is probabilistic sum method and other is probabilistic product method. First Probabilistic sum method is use as integration method. The output vectors from all the four modules are use in the equation to obtain the final output. The output vectors from the module 1, module 2, module 3 and module 4 are denoted by a, b, c and d. The final output produce by the integration unit is *O* is shown in equation 1.

$$O = aw1 + bw2 + cw3 + dw4$$

Where all the weights w1, w2, w3, w4 are equal to 0.25. The weights are taken fixed due to the error which comes after the training from each module is similar in nature. The normalization equation is use to find out the final output with probabilistic product method is shown in equation 2.

$$A = anw1 * bnw2 * cnw3 * dnw4$$

If the output O is greater than 0.5 than the heart disease is present and if it is less than 0.5 than the heart disease is absent. This is similar for the output A if it is greater than 0.5 than the heart disease is present and if it is less than 0.5 than the heart disease is absent

2. Second Method By Using Neural Network

The other way by using neural network This research work is the extension of our previous work with intelligent and effective heart attack prediction system designed with the aid of neural network[2]. In our previous work, we have presented an efficient approach for extracting patterns, which are significant to heart attack, from the heart disease data warehouses. In that we have utilized the data mining techniques: clustering and frequent pattern mining. The heart disease data warehouse consists of mixed attributes containing both the numerical and categorical data. These records are cleaned and filtered with the intention that the irrelevant data from the warehouse would be removed before mining process occurs. Then clustering is performed on the preprocessed data warehouse using K-means clustering algorithm with K value so as to extract data relevant to heart attack. Subsequently the frequent patterns significant to heart disease diagnosis are mined from the extracted data using the MAFIA algorithm. The significant weightage is calculated for each frequent pattern using the approach proposed. Then the patterns with significant weightage greater than a predefined threshold value are chosen. Afterwards, the neural network is trained with the selected significant patterns in order to predict heart attack in an efficient manner. We have employed the Multi-layer Perceptron neural network for the design of prediction system with Back-propagation as training algorithm.

Extraction of Significant Patterns from Heart Disease Data Warehouse

The extraction of significant patterns from the heart disease data warehouse is presented in this section. The heart disease data warehouse contains the screening clinical data of heart patients. Initially, the data warehouse is preprocessed to make the mining process more efficient. The preprocessed data warehouse is then clustered using the K-means clustering algorithm with K=2. This result in two clusters, one contains the data that are most relevant to heart attack and the other contains the remaining data. The frequent patterns are mined from the data, relevant to heart attack, using the MAFIA algorithm. The significant weightage is calculated for all frequent patterns with the aid of the approach proposed. The frequent patterns with significant weightage greater than a predefined threshold are chosen. These chosen significant patterns can be used in the design and development of heart attack prediction system.

Data Preprocessing

Cleaning and filtering of the data might be necessarily carried out with respect to the data and data mining algorithm employed so as to avoid the creation of deceptive or inappropriate rules or patterns. The actions comprised in the pre-processing of a data set are the removal of duplicate records, normalizing the values used to represent information in the database, accounting for missing data points and removing unneeded data fields. In order for making the data appropriate for the mining process it needs to be transformed. The raw data is changed into data sets with a few appropriate characteristics. Moreover it might be essential to combine the data so as to reduce the number of data sets besides minimizing the memory and processing resources required by the data mining algorithm . In our approach, the heart disease data warehouse is refined by removing duplicate records and supplying missing values. Furthermore it is also transformed to a form appropriate for clustering.

Clustering Using K-Means Algorithm

The categorization of objects into various groups or the partitioning of data set into subsets so that the data in each of the subset share a general feature, frequently the proximity with regard to some defined distance measure is known as Clustering. The clustering problem has been addressed in numerous contexts besides being proven beneficial in many applications. Clustering medical data into small yet meaningful clusters can aid in the discovery of patterns by supporting the extraction of numerous appropriate features from each of the clusters thereby introducing structure into the data and aiding the application of conventional data mining techniques. Numerous methods are available in the literature for clustering. We have employed the renowned K-Means clustering algorithm in our approach. The k-means algorithm is one of the widely recognized clustering tools that are applied in a variety of scientific and industrial applications. K-means groups the data in accordance with their characteristic values into K distinct clusters. Data categorized into the same cluster have identical feature values. K, the positive integer denoting the number of clusters, needs to be provided in advance.

Multi-Layer Perceptron Neural Network (MLPNN)

Literature analysis unveils a persistent application of feed forward neural networks, from amidst the various categories of connections for artificial neurons. In feed-forward neural networks the neurons of the first layer forward their output to the neurons of the second layer, in a unidirectional fashion, which explains that the neurons are not received from the reverse direction. A kind of feedforward neural network mechanism is the Multi-layer Perceptron Neural Networks (MLPNN) or Multilayer feed-forward neural network (MFNN). The structure of MLPNN is shown in Figure 1.

It is an alteration of the typical linear perceptron where in it employs three or more layers of neurons (nodes)with

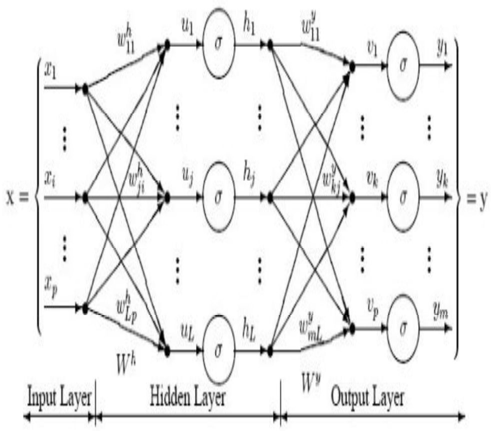


Fig. Muliti – Layer Perceptron Neural Network

nonlinear activation functions. The lone and primary task of the neurons in the input layer is the division of the input signal *i x* among neurons in the hidden layer. Every neuron *j* in the hidden layer adds up its input signals *i x* Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network once it weights them with the strengths of the respective connections ***ji w*** from the input layer and determines its output *j y* as a function *f* of the sum, given as

$$Yj = f\ (\Sigma Wji\ Xi)$$

At this instant it is possible for *f* to be a simple threshold function such as a sigmoid, or a hyperbolic tangent function. The output of neurons in the output layer is determined in an identical fashion.

3. Fuzzy Expert System Designing

The most important application of fuzzy system (fuzzy logic) is in uncertain issues. When a problem has dynamic behavior, fuzzy logic is a suitable tool that deals with this problem. First step of fuzzy expert system designing is determination of input and output variables[3]. There are 11 input variables and 1 output variable. After that, we must design membership functions (MF) of all variables. These membership functions determine the membership of objects to fuzzy sets. At first, we will describe the input variables with their membership functions. In second step, we introduce the output variable with its membership functions. And then we'll show the rules of system.The thirteen attributes are

1. Chest pain
2. Blood Sugar (Diabetes)
3. Blood Pressure
4. Cholesterol
5. Resting Electrocardiography (ECG)
6. Maximum Heart Rate
7. Exercise
8. Old Peak
9. Sex
10. Age
11. Thallium Scan

Output Variable

The "goal" field refers to the presence of heart disease in the patient. It is integer value from 0 (no presence) to 4. By increasing of integer value, heart disease risk increases in patient.After this the fuzzy rile base is designed. Rule base is the main part in fuzzy inference system and quality of results in a fuzzy system depends on the fuzzy rules. This system includes 44 rules.

CONCLUSION

The diagnosis method is proposed to for the diagnosis of the heart disease. The mixture of expert system consists of two neural network models works upon the part of dataset after the attribute division independently. The performance of the overall of modular system performance is better than the monolithic neural network. The integration unit founds the final results. The other benefit of the use mixture of neural network model works upon the different modules independently that lower down the risk involved in working out on the whole problem.

We also discussed artificial neural network in which we have presented an intelligent and effective heart attack prediction system using data mining and artificial neural network techniques. Firstly, we have provided an efficient approach for the extraction of significant patterns from the heart disease data warehouses for the efficient prediction of heart attack. The preprocessed heart disease data warehouse was clustered with the K-means clustering algorithm in order to obtain data most applicable to heart attack.. On basis of the calculated significant weightage, the frequent patterns comprising of value greater than a predefined threshold were selected for the prediction of heart attack.. The experimental results have illustrated the efficacy of the designed prediction system in predicting the heart attack.

In Fuzzy Expert System for Heart Disease Diagnosis designed with follow membership functions, input variables, output variables and rule base. Designed system has been tested with expert-doctor. Designing of this system with fuzzy base in comparison with classic designed improves results. Results have been shown from this system in compression with past time system are logical and more efficient. This system simulates the manner of expert-doctor. This system is designed in way that patient can use it himself. This fuzzy expert system that deals with diagnosis has been implemented. Experimental results showed that this system did quite better than non-expert urologist and about 92 % as a well as the expert did.

REFERENCES

[1] Use of Modular Neural Network for Heart Disease Harsh Vazirani 1, Rahul Kala 2, Anupam Shukla 3, Ritu Tiwari 4 Soft Computing and Expert System Laboratory, Indian Institute of Information & Management, Gwalior, M.P., India

[2] Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network Shantakumar B.Patil Ph.D. Scholar, Dr. MGR Educational and Research Institute, Chennai,Y.S.Kumaraswamy Senior Professor &HOD, Department of MCA (VTU)

[3] A Fuzzy Expert System for Heart Disease Diagnosis, Ali.Adeli, Mehdi.N

Electronics Components Failure Prediction using Artificial Intelligence Techniques: A Review

Cherry Bhargava[1], Dr. Vijay Kumar Banga[2]
[1]Assistant Professor, Department of Electronics and Communication,Lovely Professional University,Phagwara
[2]Professor and Head, Department of Electronics and Communication, ACET, Amritsar
Email:cherry_bhargav@yahoo.co.in, v_banga@rediffmail.com

Abstract - To generate/make an efficient, low cost and early failure detection method of electronics component failure at early stage.Neuro-Fuzy logic techniques are to be employed. Initially failure is decided using fuzzy logic technique only. Then the method/technique will be tested by its making its software model and implementation with NEFCON software.

I. INTRODUCTION

1.1 Fuzzy Logic

The concept of Fuzzy Logic (FL) was conceived by Lotfi Zadeh, a professor at the University of California at Berkley, and presented not as a control methodology, but as a way of processing data by allowing partial set membership rather than crisp set membership or non-membership. FL is a problem-solving control system methodology that lends itself to implementation in systems ranging from simple, small, embedded micro-controllers to large, networked, multichannel PC or workstation-based data acquisition and control systems [1]. 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 upon vague, ambiguous, imprecise, noisy, or missing input information. Fuzzy Logic's approach to control problems mimics how a person would make decisions, only much faster.

Fuzzy logic imitates the logic of human thought, which is much less rigid than the calculations computers generally perform [2].Fuzzy logic has its roots in the work of renegade mathematicians who saw the value of a multivalent logic system. The credit for fuzzy logic's application to the areas of control and in engineering belongs solely to Lotfi Zadeh. In 1965, he formalized fuzzy set theory and in 1973 he brought fuzzy set theory into the context of control systems. According to Lotfi Zadeh, fuzzy logic brings to control systems a ‘higher machine intelligence quotient’.

On a mathematical level, fuzzy logic abandons the strict bivalent logic of TRUE and FALSE, ONE and ZERO, ON and OFF. Fuzzy logic allows for half-truths. On an engineering level, fuzzy logic provides a platform for easily encoding human knowledge into the control of a system. It has been used in an increasing number of applications, especially in Japan. The Sendai railway in Japan is controlled by fuzzy logic controllers [3]. Applications have been developed in tracking problems, tuning, interpolation, classification, handwriting, voice recognition, and image stabilization in video cameras, washing machines, vacuum cleaners, air conditioners, electric fans, hot plates, and Lexus automatic transmissions.

Fuzzy logic is a method of characterizing knowledge in terms of fuzzy sets and a rule base. A fuzzy system has one or more inputs that are fuzzified, a rule base that is evaluated according to the inputs, and one or more outputs that are defuzzified into ‘crisp’ values. Bringing fuzzy logic to control problems is a way to use a human expert’s knowledge about an analog process in a digital computer. Fuzzy logic is not always the best way to solve a control problem, but it offers several advantages. A fuzzy system structure is illustrated in figure 1.

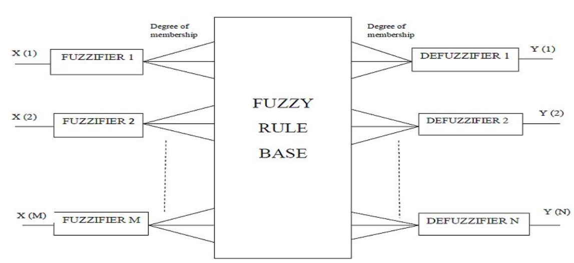


Fig. 1: A Typical Fuzzy System

1.2 Neural Networks

A neural network is a computational model of the brain. Neural network models usually assume that computation is distributed over several simple units called neurons, which are interconnected and operate in parallel (hence, neural networks are also called parallel distributed-processing systems or connectionist systems). The most popular neural network is the multi-layer perceptron, which is a feed forward network. For artificial neural network to give any results it must be trained with series of examples and conditions [4]. During the training neural network "learns" the governing relationships in given data sets e.g. input vectors to produce right solutions e.g. output vectors. For this purpose, back-propagation training algorithm is used. It is an iterative algorithm for minimizing the mean square error between predicted and desired output values.

All signals flow in a single direction from the input to the output of the network. Feed forward networks can perform static mapping between an input space and an output space: the output at a given instant is a function only of the input at that instant.

Recurrent networks, where the outputs of some neurons are fed back to the same neurons or to neurons in layers before them, are said to have a dynamic memory, the output of such networks at a given instant reflects the current input as well as previous inputs and outputs. Implicit 'knowledge' is built into a neural network by training it [5].

Some neural networks can be trained by being presented with typical input patterns and the corresponding expected output patterns. The error between the actual and expected outputs is used to modify the strengths, or weights, of the connections between the neurons. This method of training is known as supervised training. In a multi-layer perceptron, the back-propagation algorithm for supervised training is often adopted to propagate the error from the output neurons and compute the weight modifications for the neurons in the hidden layers.

Some neural networks are trained in an unsupervised mode, where only the input patterns are provided during training and the networks learn automatically to cluster them in groups with similar features. The attitudes of researchers towards neural networks have experienced an evolution since their inception in the early 1940s. According to Leon Cooper, these attitudes progressed from skepticism to what we have at present: general realistic acceptance of neural networks as the preferred -most efficient, most economic- solution to certain classes of problems [6].

Neural networks are composed solely of two elements-processing elements and interconnections. The processing elements are called neurons and the connections are termed synapses. A processing element generally has many inputs and a single output as shown in figure 2. The neural network performance is governed by the architecture of the processing element's interconnection, the transfer functions for the processing elements, and the learning law. There are two popular models of neural networks: the feed-forward model and the feedback model.

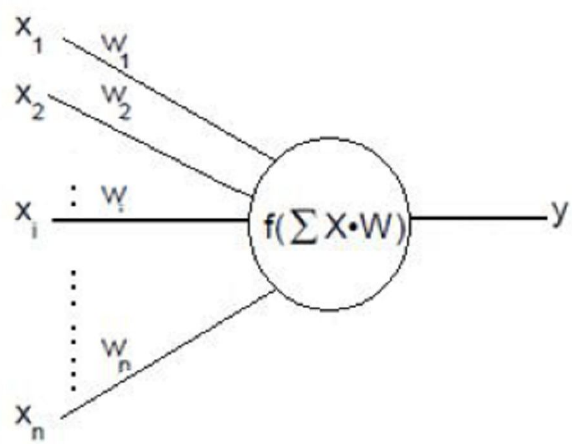


Fig. 2: An Artificial Neuron

The feed-forward model, as illustrated in figure 3, is composed of layers where the output from each level is the input for the next level. Input is applied to the input layer. The signals go to the hidden layers and then out to the output layer. Its operation is similar to the operation of a combinational logic circuit. Feed-forward neural networks work well for 'natural' problems such as pattern recognition.

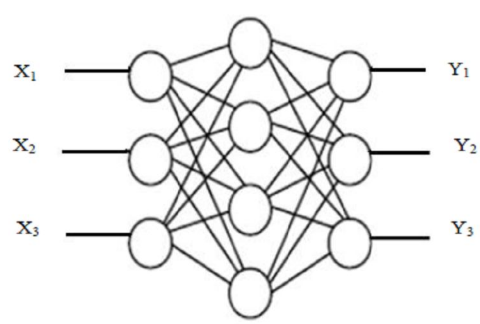


Fig.3 A feedforward neural network

Figure 4 illustrates the feedback model, which has connections between different levels, forward and backward. This is more like an asynchronous logic circuit in which the nodes evolve to a final state. Optimization problems (like the Traveling Salesman problem) are best implemented on feedback neural networks.

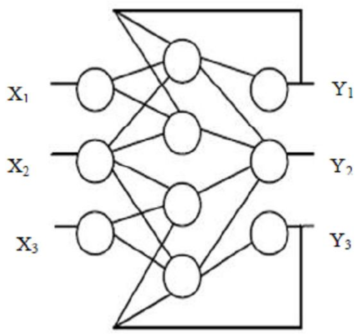


Fig.4 A feedback neural network

1.3 Neuro-Fuzzy system

Neuro-fuzzy approach combines two powerful computing disciplines: Adaptive neural networks and fuzzy set theory. Neural networks are well known for its ability to learn and adapt to unknown or changing environment to achieve better performance. Fuzzy set theory, on the other hand, can by its effectiveness in handling linguistic information, incorporate human knowledge, deal with imprecision and uncertainty, and clarify the relation between input and output variables. A neuro-fuzzy can be used to study both neural as well as fuzzy logic systems. A neural network can approximate a function, but it is impossible to interpret the result in terms of natural language. The fusion of neural networks and fuzzy logic in neuro fuzzy models provide learning as well as readability. Control engineers find this useful, because the models can be interpreted and supplemented by process operators.

Recently, the combination of neural networks and fuzzy logic has received attention. The idea is to lose the disadvantages of the two and gain the advantages of both. Neural networks bring into this union the ability to learn. Fuzzy logic brings into this union a model of the system based on membership functions and a rule base [7].

This field of study is still in its infancy. Universally accepted techniques and a general consensus on the direction of research have not yet been established. Most of the work done in this area is still associated with individual researchers and has not been adopted as standard strategy.

Determining the fuzzy membership functions from sample data using a neural network is the most obvious method of using the two together. The definition of the membership function has a huge impact on the system response. Often, the programmer must use trial and error to find acceptable values [8].

Assuming a certain shape and finding the beginning and endpoints for the fuzzy values in a fuzzy set is a neural network optimization problem. Figure 5 is a diagram of such a system.

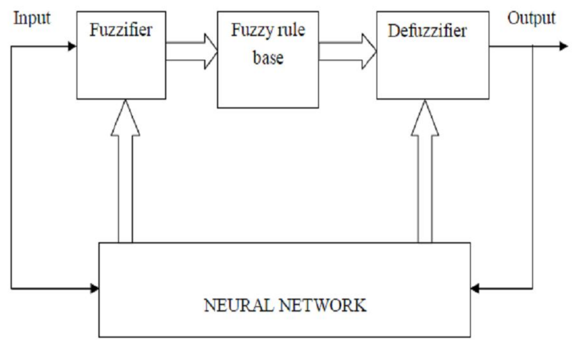


Fig.5 A fuzzy system whose membership function is adjusted by neural network

II CONDITION MONITORING AREA

P. Lall, P. Gupta and K. Goebel presented a method for failure mode classification using a combination of Karhunen Loeve transform with parity-based stepwise supervised training of a perceptrons [9]. Statistical similarity and validation of different classified dominant failure modes is performed by multivariate analysis of variance and Hoteling's T-square. The results of different classified dominant failure modes are also correlated with the experimental cross sections of the failed test assemblies. The methodology adopted by them can perform real-time fault monitoring with identification of specific dominant failure mode and is scalable to system level reliability [10].

A review on condition monitoring for device reliability in power electronic converters was presented by S Yang, D Xiang, A Bryant, P Mawby, L Ran and P Tavner [11].

L N Lu along with H Z Huang, X X Su, B Y Wu and M Cai, presents an investigation on field returned open and short failures related to printed circuit board (PCB), including via hole crack, prepreg crack and insufficient circuit etching [48]. After an experimental study with cross section, time domain reflectometry (TDR), and finite element (FE) modelling, it was found by them that weak plating and corrosion induced via hole crack was a major root cause of interconnect open failures. They concluded that it is important to monitor and control the incoming quality of incoming PCBs so that the risk of high field return rate and high cost for repair can be minimized [12].

III PROBLEM FORMULATION

To generate/make an efficient, low cost and early failure detection method of electronics component failure at early stage,Neuro-Fuzzy logic techniques are to be employed which have vast applications and until now, a very less work has been done on the purposed topic with implementation of these techniques.Five electronics components are to be choosen: Diode , BJT, UJT, SCR, MOSFET.

Each case is considered separately.Inputs parameters are decided as follow:Fabrication, Thermal Stress (Ambient temperature), Mechanical stress

Output parameters:

Health (or Failure)

0 = not failed (or healthy)

0.5 = going to be failed

1 = fail

Then method is tested by NEFCON software

IV CONCLUSION

Initially failure is decided using fuzzy logic technique only. Then the method/technique will be tested by its making its software model and implementation with NEFCON software

REFERENCES

[1] E Cox, “Fuzzy Fundamentals” IEEE Spectrum, (October 1993), pp. 58- 61, 1993.

[2] L A Zadeh, “Fuzzy algorithms”, Info. & Ctl., Vol. 12, pp. 94-102,1968.

[3] H J Zimmermann, Fuzzy Sets, Decision Making and Expert Systems, Boston, Kluwer, 1987.

[4] S Haack, “Do we need fuzzy logic?”, Int. Jrnl. of Man-Mach. Stud., Vol. 11, pp.437-445, 1979.

[5] R Kruse, J Gebhardt, F Klawon, “Foundations of Fuzzy Systems”, Wiley, Chichester 1994.

[6] F Esragh and E H Mamdani, “A general approach to linguistic approximation in Fuzzy Reasoning and Its Applications”, E.H. Mamdani and B.R. Gaines (eds.), London: Academic Press, 1981.

[7] J Engin, J Kuvulmaz and E Omurlu. “Fuzzy control of an ANFIS model representing a nonlinear liquid level system”. Neural Computing and Appl., Vol. 13, n. 3 pp. 202-210, 2004.

[8] M. Denai, F. Palis and A. Zeghbib. “ANFIS based modeling and control of nonlinear systems: A tutorial”, Proceedings of the IEEE Conference on SMC, pp. 3433-3438, 2004.

[9] S Yang, D Xiang, A Bryant, P Mawby, L Ran, P Tavner, “ Condition Monitoring for Device Reliability in Power Electronic Converters: A Review” IEEE Transactions on Power Electronics, Volume: 25 , Issue: 11, November, 2010.

[10] P Lall, P Gupta, K Goebel, “Decorrelated feature space and neural nets based framework for failure modes clustering in electronics subjected to mechanical-shock”, IEEE Conference on Prognostics and Health Management (PHM), 2011, **pp:** 1 – 16, 20-23 June 2011.

[11] D E Verbitsky, “Advanced FTA technique addressing early un-balanced failures of modern commercial electronic devices” Reliability and Maintainability Symposium, 2009, pp. 495 – 501, 26-29 Jan. 2009.

[12] D C Stoudt, R L Gardner, I Kohlberg, “The role of hybrid models in understanding failure mechanisms of infrastructure electronics due to high-power microwave illumination”, International Conference on Electromagnetics in Advanced Applications, 2009, pp.905-908, 14-18 Sept. 2009.

Robotics and Soft Computing

Er.Gurjeet Singh, Dr V.K Banga, Er. Damanbir Singh
Amritsar College of Engineering & Technology, Amritsar
.

***Abstract*— Robots are effectively used in a wide range of application; they must gain the ability to work in unpredictable and changing environments.All over the world, robots are on the move. In order to co-operate with a human, a robot should have a human-like behaviour when moving. To achieve this, is necessary to give the robot human like configuration and human like kinematics. In this paper, we provide a brief review of robotics and its kinematics type. The inverse kinematics of two and three links arm of robotic system have been discussed. Robotics requires a working knowledge of electronics, mechanics and software, and is usually accompanied by a large working knowledge of many subjects. A person working in the field computation using soft computing that learns the inverse kinematics of a robot arm played important role in the robotics.**

***Keywords*—** kinematics, Robotics arm, Neural networks, Fuzzy logic.

I. INTRODUCTION

ROBOTICS is branch of engineering that deal with the construction designing of robot. Although the appearance and capabilities of robots vary vastly, all robots share the features of a mechanical, movable structure under some form of autonomous control[5,6]. The structure of a robot is usually mostly mechanical and can be called a kinematic chain (its functionality being similar to the skeleton of the human body). The chain is formed of links (its bones), actuators (its muscles) and joints which can allow one or more degrees of freedom. Most contemporary robots use open serial chains in which each link connects the one before to the one after it. These robots are called serial robots and often resemble the human arm. Some robots, such as the Stewart platform, use closed parallel kinematic chains. Robots used as manipulators have an end effector mounted on the last link. This end effector can be anything from a welding device to a mechanical hand used to manipulate the environment. The mechanical structure of a robot must be controlled to perform tasks. The control of a robot involves three distinct phases - perception, processing and action (robotic paradigms). Sensors give information about the environment or the robot itself (e.g. the position of its joints or its end effector). Using strategies from the field of control theory, this information is processed to calculate the appropriate signals to the actuators (motors) which move the mechanical structure. The control of a robot involves various

Mobile Robots -The term mobile robot describes a robotic system able to carry out tasks in different places and consisting
of a platform moved by locomotive elements. The choice of the locomotive system depends firstly on the environment in which the robot will operate. This can be aerial, aquatic or terrestrial. In the aquatic and aerial environments, the locomotive systems are usually propellers or screws, although at the seabed legs are also used. The choice of the locomotive system on earth is more complicated due to the variety of terrestrial environments.

Medical Robots -In recent years, the field of medicine has been also invaded by robots, not to replace qualified personnel such as doctors and nurses, but to assist them in routine work and precision tasks. Medical robotics is a promising field that really took off in the 1990s. Since then, a wide variety of medical applications have emerged: laboratory robots, telesurgery, surgical training, remote surgery, telemedicine and teleconsultation, rehabilitation, help for the deaf and the blind, and hospital robots.

Outer Space - Manipulative arms that are controlled by a human are used to unload the docking bay of space shuttles to launch satellites or to construct a space station

The Intelligent Home - Automated systems can now monitor home security, environmental conditions and energy usage. Door and windows can be opened automatically and appliances such as lighting and air conditioning can be pre programmed to activate. This assists occupants irrespective of their state of mobility.

Exploration - Robots can visit environments that are harmful to humans. An example is monitoring the environment inside a volcano or exploring our deepest oceans. NASA has used robotic probes for planetary exploration since the early sixties.

Military Robots - Airborne robot drones are used for surveillance in today's modern army. In the future automated aircraft and vehicles could be used to carry fuel and ammunition or clear minefields

Farms - Automated harvesters can cut and gather crops. Robotic dairies are available allowing operators to feed and milk their cows remotely.

The Car Industry - Robotic arms that are able to perform multiple tasks are used in the car manufacturing process. They perform tasks such as welding, cutting, lifting, sorting and bending. Similar applications but on a smaller scale are now being planned for the food processing industry in particular the trimming, cutting and processing of various meats such as fish, lamb, beef.

Hospitals **-** Under development is a robotic suit that will enable nurses to lift patients without damaging their backs. Scientists in Japan have developed a power-assisted suit which will give nurses the extra muscle they need to lift their patients and avoid back injuries.

II. Robotics Arm

A robotic arm is a robotic manipulator, usually programmable, with similar functions to a human arm. The links of such a manipulator are connected by joints allowing either rotational motion (such as in an articulated robot) or translational (linear) displacement. The links of the manipulator can be considered to form a kinematic chain [7, 11]. The business end of the kinematic chain of the manipulator is called the end effector and it is analogous to the human hand. The end effector can be designed to perform any desired task such as welding, gripping, spinning etc., depending on the application. For example robot arms in perform a variety of tasks such as welding and parts rotation and placement during assembly.

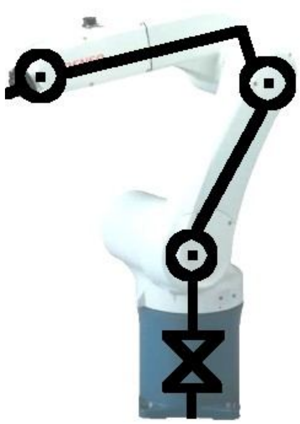


Fig 1: Robotic Arm with translation motion

Robot arms are typically categorized by the number of controlled degree of freedom (DOF) they can execute. This number is equal to the sum of the DOF of each of a robot arm's individual joints. Generally these will be either Hinge joints or Pivot joints, both of which are only capable of rotation about a sing The number of DOF that a manipulator possesses is the number of independent position variables that would have to be specified in order to locate all parts of the mechanism. In other words, it refers to the number of different ways in which a robot arm can move [6]. A normal human arm is redundant (and therefore holonomic) in that it has seven DOF. The shoulder gives pitch, yaw and roll. The elbow allows for pitch. The wrist allows for pitch and yaw. And the elbow and wrist together allow for Roll. Only three DOF are needed to move the hand to any particular point within a given three dimensional space, but having a greater the number of controlled DOF enables human beings to grasp items in that space from a variety of different angles and directions. In a Two Dimensional (2-D) space (like a table-top or the floor) there are three Degrees of Freedom. These include displacement along the X and Y axes, plus rotation. In a Three Dimensional (3-D) space there are six degrees of freedom. These consist of displacement along three perpendicular axes (X, Y and Z), and rotation about those same axes. DOF in 3D space are generally identified using the following nautical terms: Displacements

- Heave: Moving up and down
- Surge: Moving forward and backward
- Sway: Moving left and right Rotations
- Yaw: Turning left and right flight
- Roll: Tilting side to side
- Pitch: Tilting forward and backward

III. Robotic Arm structure

Cartesian Robot / Gantry Robot: Used for pick and place work, application of sealant, assembly operations, handling machine tools and arc welding. It's a robot whose arm has three prismatic joints, whose axes are coincident with a Cartesian coordinator.

Cylindrical Robot: Used for assembly operations, handling at machine tools, spot welding, and handling at die-casting machines. It's a robot whose axes form a cylindrical coordinate system.

*Spherical Robot / Polar Robot (such as the Unimate):*Used for handling at machine tools, spot welding, die-casting, fettling machines, gas welding and arc welding. It's a robot whose axes form a polar coordinate system

SCARA Robot: Used for pick and place work, application of sealant, assembly operations and handling machine tools. It's a robot which has two parallel rotary joints to provide compliance in a plane.

PUMA Robot **-**In 1978, the Puma (Programmable Universal Machine for Assembly) robot is developed by Victor Scheinman at pioneering robot company Unimation with a General Motors design support. These robots are widely used in various organisations such as Nokia Corporation, NASA, Robotics and Welding organization.

Articulated Robot: Used for assembly operations, die-casting, fettling machines, gas welding, arc welding and spray painting. It's a robot whose arm has at least three rotary joints. It is most widely used in the industry.

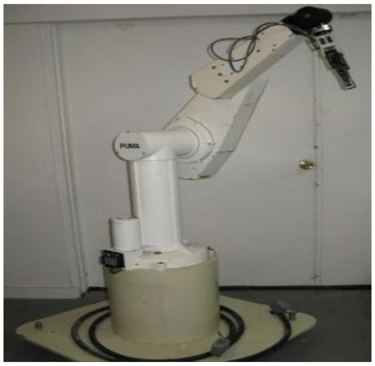


Fig 2: PUMA Robot

An articulated robot is a robot with rotary joints (e.g. a legged robot or an industrial robot). Articulated robots can range from simple two-jointed structures to systems with 10 or more interacting joints. They are powered by a variety of means, including electric motors.

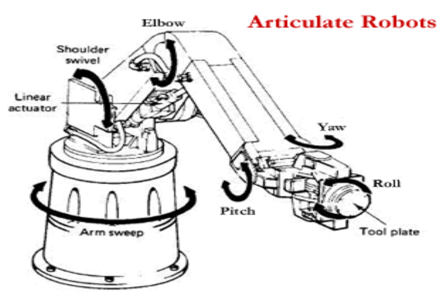


Fig 3: Articulate robot

A rotary joint is a connection between two objects. The connection allows both objects, even though each is connected to another object, the ability to rotate or have movement up to 360 degrees. The two rigid objects that are attached by the joint are sometimes called a kinematic pair while the joint is referred to as a mechanical constraint. Most of the time these two objects that are connected together are cylindrical. The connection gives both objects increased capabilities to perform work functions. Articulated robots usually have several of these connections which gives them a great deal of flexibility in performing work duties. Each joint that a robotic has represents an increase in freedom to perform tasks. There is no limit to the number of rotary joints that articulated robots can have and a robotic may have other types of joints to increase its capability even more [3, 5].

Parallel Robot: One use is a mobile platform handling cockpit flight simulators. It's a robot whose arms have concurrent prismatic or rotary joints.

Anthropomorphic Robot: Shaped in a way that resembles a human hand, i.e. with independent fingers and thumbs.

IV. Kinematics

It is (from Greek κινε□ν, kinein, to move) is the branch of classical mechanics that describes the motion of bodies (objects) and systems (groups of objects) without consideration of the forces that cause the motion. Kinematics is not to be confused with another branch of classical mechanics: analytical dynamics (the study of the relationship between the motion of objects and its causes), sometimes subdivided into kinetics (the study of the relation between external forces and motion) and statics (the study of the relations in a system at equilibrium). Kinematics also differs from dynamics as used in modern-day physics to describe time-evolution of a system [1]. Kinematics is the process of calculating the position in space of the end of a linked structure, given the angles of all the joints. It is easy, and there is only one solution. Inverse Kinematics does the reverse. Given the end point of the structure, what angles do the joints need to be in the achieve that end point. It can be difficult, and there are usually many or infinitely many solutions. This process can be extremely useful in robotics. You may have a robotic arm which needs to grab an object. If the software knows where the object is in relation to the shoulder, it simply needs to calculate the angles of the joints to reach it. The simplest application of kinematics is for particle motion, translational or rotational [8]. The next level of complexity comes from the introduction of rigid bodies, which are collections of particles having time invariant distances between themselves. Rigid bodies might undergo translation and rotation or a combination of both. A more complicated case is the kinematics of a system of rigid bodies, which may be linked together by mechanical joints. It s of two types.

- Forward kinematics
- Inverse kinematics

Forward kinematics: The essential concept of forward kinematic animation is that the positions of particular parts of the model at a specified time are calculated from the position and orientation of the object, together with any information on the joints of an articulated model. So for example if the object to be animated is an arm with the shoulder remaining at a fixed location, the location of the tip of the thumb would be calculated from the angles of the shoulder, elbow, wrist, thumb and knuckle joints. Three of these joints (the shoulder, wrist and the base of the thumb) have more than one degree of freedom, all of which must be taken into account. If the model were an entire human figure, then the location of the shoulder would also have to be calculated from other properties of the model [2, 4].

Inverse kinematics: It will enable us to calculate what each joint variable must be if we desire that the hand be located at particular point and have a particular position. The position and orientation of the end effector relative to the base frame compute all possible sets of joint angles and link geometries which could be used to attain the given position and orientation of the end effector [1, 4].

Forward Kinematics

$$x = l_1 \cos\theta 1 + l_2 \cos(\theta 1+ \theta 2) \quad (1)$$
$$y = l_1 \sin \theta 1 + l_2 \sin (\theta 1+\theta 2) \quad (2)$$

Inverse kinematics

$x^2 + y^2 = l_1^2 \cos^2\theta_1 + l_2^2 \cos^2(\theta_1+ \theta_2) + 2 l_1 l_2 \cos\theta_1 \cos^2 (\theta_1+ \theta_2)+ l_1^2 \sin^2\theta_1 + l_2^2 \sin^2(\theta_1+ \theta_2) + 2 l_1 l_2 \sin\theta_1 \sin^2 (\theta_1+ \theta_2)$
(3)

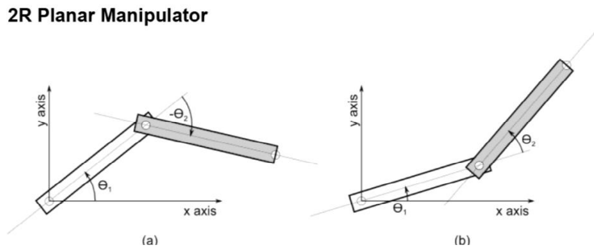


Fig 4: Two link robotic arm

$$= l_1^2+ l_2^2 + 2l_1l_2 \cos\theta_1 \cos (\theta_1+ \theta_2)+ \sin \theta_1 \sin^2(\theta_1+ \theta_2) \quad (4)$$

Next we use the following equalities

$$\mathrm{Sin}(x \pm y) = \sin x \cos y \pm \cos x \sin y \quad (5)$$
$$\mathrm{Cos}(x\pm y)= \cos x \cos y \pm \sin x \sin y \quad (6)$$

Therefore

$$x^2 + y^2 = l_1^2 + l_2^2 + 2 l_1 l_2[\cos\theta_1 \cos\theta_2 - \sin \theta_1 \sin \theta_2 + \sin \theta_1 (\cos\theta_2 \sin \theta_1 + \cos\theta_1 \sin \theta_2)] \quad (7)$$

$$= l_1^2 + l_2^2 + 2 l_1 l_2 [\cos^2\theta_1 \cos\theta_2 + \sin^2\theta_2 \cos\theta_2]$$
$$= l_1^2 + l_2^2 + 2 l_1 l_2 \cos\theta_2 \quad (8)$$

And

$$\cos\theta_2= \frac{x^2 + y^2 - l_1^2 - l_2^2}{2l_1 l_2} \quad (9)$$

From here we could get the angle directly using the arc cose function but this function is very inaccurate for small angle s, . the typical way to avoid this accuracy is to convert further until we can use the atan2 function

$\cos^2\theta_2 + \sin^2\theta_2 =1$ and $\sin\theta_2=\pm\sqrt{1-\cos^2\theta_2}$

the two solutions corresponding to the 'elbow up' and 'elbow down 'configuration as shown in above figure and finally we get .

$$\theta_2 = \mathrm{atan2}(\sin \theta_2 \cos \theta_2) \quad (10)$$

$$= \mathrm{a\ tan2}\ (\pm\sqrt{1-\cos^2\theta_2}.\ \cos \theta_2$$

$$=\mathrm{atan2}(\pm\sqrt{1-(\frac{x^2 + y^2 - l_1^2 - l_2^2}{2l_1 l_2})}, \frac{x^2 + y^2 - l_1^2 - l_2^2}{2l_1 l_2}$$

(11)

for solving θ_1 we rewrite the original nonlinear equations using a change of variables as follow (figure 4)

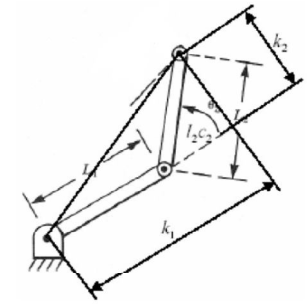


Fig 5: Mathematical expression for two link

$$x=l_1\cos\theta_1+l_2\cos (\theta_1+ \theta_2) \quad (12)$$
$$y=l_1\sin\theta_1+l_2 \sin (\theta_1+ \theta_2) \quad (13)$$
$$x = k_1 \cos\theta_1 + k_2 \sin\theta_1 \quad (14)$$
$$y=k_1\sin\theta_1+ k_2 \cos\theta_1 \quad (15)$$

where

$$k_1 =l_1+ l_2 \cos\theta_2 \quad (16)$$
$$k_2 = l_2 \sin\theta_2 \quad (17)$$

next we change the way we write the constant $k_1\ k_2$ (figure 5)

$$r = \sqrt{k_1^2 + k_2^2}$$

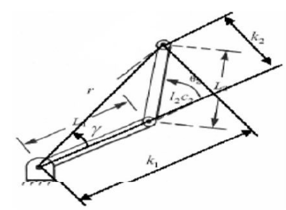


Fig 6: Mathematical expression

$$□= \mathrm{atan}\ 2(k_2,k_1) \quad (18)$$

K_1= r cos□
K_2= r sin□

Inserting into previous transformations of x and y yields.

$$X= r \cos□\ \cos\theta_1+r \sin \theta_1 \sin□ \quad (19)$$
$$y= r \cos□\ \cos\theta_1+r \sin \theta_1 \sin□ \quad (20)$$

or $\frac{y}{r}= \sin (\theta_1+□)$

$\frac{x}{r}= \cos (\theta+□)$

Apply the atan2 function

$$□+ \theta_1\mathrm{atan}\ 2(\frac{y}{r}, \frac{y}{r})=\mathrm{atan2}(y,x) \quad (21)$$
$$\theta_1 = \mathrm{atan}(y,x)-\mathrm{atan2}(k_2, k_1) \quad (22)$$

Mathematical model of three degree-of-freedom (3 DOF) robotic system

To calculate movements in dynamic systems made up of several parts, the main approach is to calculate possible movements with the aid of mathematical models. At the same time it is necessary to understand both the mechanics and the physical aspects. A vertical articulated robotic arm with 3 links (Figure 1) having length l1, l2 and l3 respectively, is considered which has a three degree-of-freedom [2,3]. In three degree-of-freedom robotic arm the inverse kinematics equations are as below:

$x = l1 \cos \theta 1 + l2 \cos (\theta 1 + \theta 2) + l3 \cos (\theta 1 + \theta 2 + \theta 3$ (23)
$y = l1 \sin \theta 1 + l2 \sin (\theta 1 + \theta 2) + l3 \sin (\theta 1 + \theta 2 + \theta 3)$ (24)
$\theta = \theta 1 + \theta 2 + \theta 3$

Knowing the arm link lengths l1, l2 and l3 for position (x, y) we had calculated the values of joint angles θ1,θ2,θ3 [3]

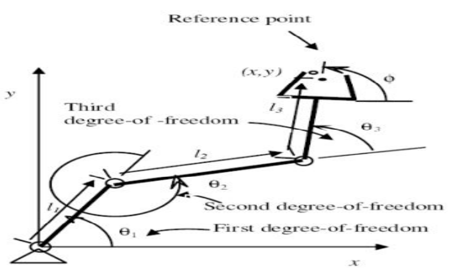


Fig 7: Three link robotic arm

Modeling of robotics arm

The inverse kinematics problem is much more interesting and its solution is more useful. At the position level, the problem is stated as, "Given the desired position of the robot's hand, what must be the angles at all of the robots joints?"Humans solve this problem all the time without even thinking about it. When you are eating your cereal in the morning you just reach out and grab your spoon. You don't think, "my shoulder needs to do this, my elbow needs to do that, etc." Below we will look at how most robots have to solve the problem. We will start with a very simple example.

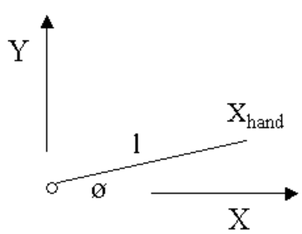


Fig 8: Single link manipulator

The figure above is a schematic of a simple robot lying in the X-Y plane. The robot has one link of length l and one joint with angle Ø. The position of the robot's hand is Xhand. The inverse kinematics problem (at the position level) for this robotis as follows: Given Xhand what is the joint angle Ø? We'll start the solution to this problem by writing down the forward position equation, and then solve for Ø.

Xhand = lcosØ (forward position solution)

CosØ = Xhand / l

Ø = cos-1(Xhand/l)

To finish the solution let's say that this robot's link has a length of 1 foot and we want the robot's hand to be at X = .7071 feet. That gives:

Ø = cos-1(.7071) = +/- 45 degrees

Even for this simple example, there are two solutions to the inverse kinematics problem: one at plus 45 degrees and one at minus 45 degrees! The existence of multiple solutions adds to the challenge of the inverse kinematics problem. Typically we will need to know which of the solutions is correct. All programming languages that I know of supply a trigonometric function called ATan2 that will find the proper quadrant when given both the X and Y arguments: Ø = ATan2(Y/X). There is one more interesting inverse kinematics problem

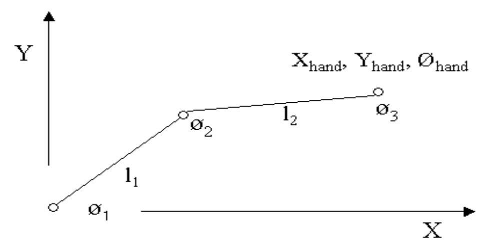


Fig 9: Two link manupulator

You may have to use your imagination a bit, but the schematic above is the planar part of the SCARA robot we discuss in the industrial robots section. Here's the statement of the inverse kinematics problem at the position level for this robot:

Given: Xhand, Yhand, Øhand

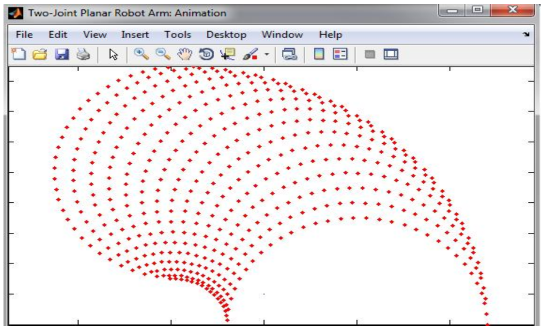
Find:Ø1,Ø2andØ3 To aid in solving this problem, I am going to define an imaginary straight line that extends from the robot's first joint to its last joint as follows:

B: length of imaginary line
q1: angle between X-axis and imaginary line
q2: interior angle between imaginary line and link l1
Then we have:
B2 =Xhand2 + Yhand2 (22)
q1 =ATan2 (Yhand/Xhand) (23)
q2 =a cos [(l12 - l22 + B2)/2l1B] (24)
Ø1 = q1 + q2

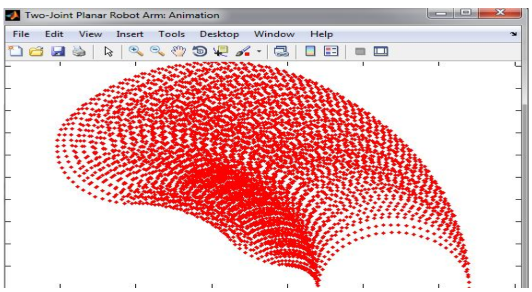
V. Workspace Area

The workspace area is the area where the endeffector can move easily and the resolution is maximum in this area. The work space area of 2 and 3 arm Robot with help of GUI

(Graphic User Interface) is shown below .The red star show the resolution.



(a)



(b)

Fig 10: Workspace area of (a) Two Arm (b) Three Arm

VI. SOFT COMPUTING

This problem is well addressed by neuro-fuzzy techniques because a solution is not easily found by analytical or numerical techniques. While an analytical technique is difficult, moving an arm in the presence of an obstacle can be instinctively performed b a child. Neuro-fuzzy systems excel in using sample data to determine an input-output relationship. Neural networks bring to this solution the ability to learn while fuzzy logic is based on mimicking an expert's thinking. In addition, as hardware technology progresses, more and more value will be placed on solutions that can utilize parallel processing, like neural networks. The field of neuro-fuzzy technology has gone in many directions. The neuro-fuzzy technique replaces the traditional fuzzy logic system with a multilayer back propagation neural network. This type of system is beneficial for several reasons. While it is true that a child is able to move an arm around an obstacle to reach a desired goal, that ability is intuitive. Putting the instructions for performing such a task into a neat, fuzzy logic, IF/THEN rule base is not easy [12]. Thus, there is a necessity for the neural network to learn the rules. The fuzzifiers and defuzzifiers necessary for any fuzzy system provide an interface between an expert's control of a simulated arm and the neural network. GAs are tools on probabilistic and casualty, not necessarily they will have the same type of evolution when applied to the same problem. GAs are slower because they are tools of evolution and not for specific optimizations [9]. They are simpler, easy programming, and demand less mathematics complexity to describe the process to be optimized. ANN and fuzzy logic techniques required more information regarding system and more mathematics as compare to GA. The great advantage observed in the GAs are tool of easy application and in robotics they could be thoroughly used to do several tasks, needing for that only small description of the problem.

VII. CONCLUSION

In this paper we have presented the robotics kinematics for two and three link manipulator problem and discussed the kinematics in the contest of soft computing techniques like fuzzy, neural network and genetic algorithms. It is concluded that in the presence of several optimization attributes for a physical system of higher order manipulator, soft computing techniques are alternatives to find the solutions of kinematics problem. Various soft computing techniques have their own advantages as neural networks required complete information of the system and required training where as fuzzy and genetic algorithms required less information of the system and easy to implement.

REFERENCES

[1] Ahuactzin J. M., Gupta K. K., 1999. The kinematic roadmap: a motion planning based global approach for inverse kinematics of redundant robots. IEEE Transactions on Robotics and Automation, 15(4), 653-669.
[2] Banga V K, Kumar R, Singh Y “Fuzzy genetic optimal control for robotics system”, International journal of Physical Sciences. 6(2), 204-212, 2011.
[3] Banga V K, Kumar R, Singh Y “Modling and simulation of robotics arm movement using soft computing ”world academy of science ,75,614-620 2011.
[4] Cameron S., 1994. Obstacle avoidance and path planning. International Journal of Industrial Robot 21(5), 9-14.
[5] Craig J.J., *Introduction to Robotics: Mechanics and control*, Pearson Education Asia, 2004.
[6] Deb S.R., *Robotics Technology and Flexible Automation*. Tata Mc-Graw Hill, 2002.
[7] Fu K.S., Gonzalez R.C., Lee C.S.G., *Robotics: Control, Sensing, Vision and Intelligence*. Mc-Graw Hill International Editions, 1987.
[8] Galicki M., 2000. Time-optimal controls of kinematically redundant manipulators with geometric constraints. IEEE Transactions on Robotics and Automation 16(1), 89-93.
[9] Gemeinder M. and Gerke M., 2003. GA-based path planning for mobile robot systems employing an active search algorithm. Elsevier Applied Soft Computing 3(2), 149-158.
[10] Lozano P. and Erez T., 1987. A simple motion-planning algorithm for general robot manipulators. IEEE Journal of Robotics and Automation 3(3), 224–238.
[11] Schilling R.J., *Fundamental of Robotics: Analysis and Control*, Prentice Hall if India Pvt. Ltd., 2002.
[12] Tang Y. and Velez-Diaz D., 2003. Robust fuzzy control of mechanical systems. IEEE Transactions on Fuzzy Systems 11(3), 411- 418.

Ant Colony Optimization for Load Balancing in Grid Computing

Heena Sharma[1], Shital Sharma[2] , Harpreet kaur[3] , Ramanjyot kaur [4]
Sant Baba Bhag Singh Institute of Engineering and Technology, Jalandhar
Email:heena2691sharma@yahoo.com,Shitalsharma1991@yahoo.in,
erharpreetarora@gmail.com,[4]er.ramanjyot@gmail.com

Abstract: **Ant colony optimization (ACO) is a Swarm Intelligence technique which inspired from the foraging behaviour of real ant colonies. The ant deposit pheromone on the ground in order to mark the rout for identifacation of their routes from the nest to food that should we followed by other members of the colony. This ACO exploits an optimazation mechanism for solving discrete optimization problems in various engineering domain. From the early nineties, when the first ant colony optimization algorithm was proposed. ACO attracted the attention of increasing number of researchers and many successful applications are now available. ACO place the vital role in discrete optimization problems the ACO solves many engineering problems and provides optimal result which includes travelling salesman problem ,network rounting and scheduling.ACO for performance improvement of the grid scheduling is also discussed .it is found that the scheduling system using an ant colony optimization algorithm can efficiently and effectively allocate jobs to proper resources.**

Keyword: Ant colony optimization, swarm intelligence technique, resource sharing scheduling

I. INTRODUCTION

Grid architecture involves the efficient management of distributed, heterogeneous and dynamically available resources. There are various issues in grid resource management such as resource discovery, resource scheduling, resource monitoring, resources inventories, resource provisioning, load balancing, fault isolation, autonomic capabilities and service level management system (Sharma & Bawa, 2008).However, grid scheduling and grid load balancing are the main issues that are often discussed by many Enhancement of Ant Colony Optimization for Grid Load Balancing .Scheduling the jobs to the resources in grid computing is also complicated due to the distributed and heterogeneous nature of the resources (Li, 2006). Submitted jobs must be matched between the available resources in terms of their job characteristics and resources capacity. Current algorithms (Xu et al., 2003, Chang et al., 2007) have considered the jobs characteristics and resources capacity in scheduling of jobs but did not take into considerations of Million Instruction per Second (MIPS) and CPU time needed of each job but did not study the load balancing problem.

Li (2006) proposed a bio-inspired adaptive job scheduling mechanism with a system. In this architecture, there is no direct communication among these agents. The only indirect communication is via the pheromone values stored in a grid resource table. Lorpunmanee et al. (2007) proposed an ant colony optimization (ACO) technique for dynamic job scheduling in grid computing with the aim to develop an effective grid scheduling algorithm that could minimize the total tardiness time of each submitted job. However, both studies did not consider the load balancing and the current conditions of each resource during scheduling process.

The study to improve ant algorithm for dynamic job scheduling in grid computing which is based on the basic idea of ant colony system (ACS) was proposed by Yan et al. (2005). The pheromone update function in this research was performed by adding encouragement, punishment coefficient and load balancing factor. The initial pheromone value of each resource was based on its status where job was assigned to the resource with the maximum pheromone value. The strength of pheromone of each resource was updated after completion of the job. The encouragement, punishment and local balancing factor coefficient were defined by users and were used to update pheromone values of resources.

The dynamic grid scheduling algorithm based on adaptive ant colony algorithm was proposed by Liu & Wang (2008). In the algorithm, the evaporation rate value was adaptively changed and a minimum value of zero was fixed. The local and global pheromone updates were used in order to control the pheromone value of each resource. The performance of the proposed algorithm was compared with the basic ant colony algorithm.

An enhanced ant algorithm for dynamic task scheduling in grid computing was proposed by Sathish & Reddy (2008) which gave better throughput with a controlled cost. The proposed scheduling algorithm increased the performance in terms of low processing time and low processing cost when applied to a grid application with a large number of jobs such as parameter sweeps application. This algorithm worked effectively in minimizing the processing time and processing cost of the jobs.

II . ANT COLONY OPTIMIZATION FOR GRID LOAD BALANCING

Balanced job assignment based on ant algorithm for computing grids called balanced ant colony optimization (BACO) was proposed by Chang et al. (2007) with the aim to minimize the computation time of job executing in Taiwan Uni Grid environment which also focused on load balancing factors of each resource. By considering the resource status and the size of the given job, BACO algorithm chose optimal resources to process the submitted jobs. The local and global pheromone update techniques were used to balance the system load. Local pheromone update function updated the status of the selected resource after a job had been assigned and the job scheduler depends on the latest information of the selected resource for the next job submission. Global pheromone update function updated the status of each resource for all jobs after the completion of the jobs. By using these two update techniques, the job scheduler obtained the latest information of all resources for the next job submission. From the experimental results, BACO was capable of balancing the entire system load regardless of the size of the jobs in the static scheduling benchmark. The study did not consider the capacity of each resource during the scheduling process.

In Moallem & Ludwig (2009), two distributed artificial life-inspired algorithms were introduced, which are ACO and particle swarm optimization (PSO) in solving the static grid load balancing problem. Distributed load balancing are categorized as a robust algorithm that can adapt to any topology changes in a network. In the study, an ant acted as a broker to find the best node in terms of the pheromone value stored in the pheromone table. The node with the lightest load was selected as the best node. The position of each node in the flock could be determined by its load in PSO. The particle compared the load of nodes with its neighbours and moved towards the best neighbour by sending assigned jobs to it. The proposed algorithm performed better than ACO for job scheduling where jobs were submitted from different sources and different time intervals. PSO showed better results than ACO in terms of the make span. However, PSO used more bandwidth and communication compared to ACO.
The main drawback of Ant Colony was that jobs are not scheduled efficiently and therefore load among the resources were not balanced. This problem was fixed by increasing the number of ants that can explore the entire grid system to find resources with the lightest load.

ACO algorithm for dynamic load balancing in distributed systems through the use of multiple ant colonies was proposed by Ali et al. (2010). In this algorithm, information on resources was dynamically updated at each ant movement. Load balancing system was based on multiple ant colonies information. Multiple ant colonies were adopted such that each node sent a coloured colony throughout the network. Colored ant colonies were used to prevent ants of the same nest from following the same route and also force them to be distributed all over the nodes in the system. Each ant acted like a mobile agent which carried newly updated load balancing information to the next node. The algorithm was compared to the work-stealing approach for load balancing in grid computing. Experimental results showed that multiple ant colonies worked better than work-stealing algorithm in terms of their efficiency. However, the multiple ant colonies did not consider resources capacity and jobs characteristics. This can make matching the jobs with the best resources a difficult task for the scheduling algorithm. From the above research, ACS is the most popular variant of ACO that has been successfully used in grid load balancing.

III. EACO ALGORITHM

The mechanisms are initial pheromone value mechanism, resource selection mechanism and pheromone updating mechanism that are used to organize the work of the ant colony. The initial pheromone value mechanism indicates the capacity of each resource to process a particular job while resource selection mechanism will solve the scheduling problem by selecting the best resource to process the job. Pheromone updating mechanism will solve the load balancing problem of the resources.

The algorithm called enhanced ant colony optimization (EACO) consists of 5 steps Which are obtain job requirements, create an ant for a job, calculate and store the initial pheromone for all the resources, assign job to resource with highest pheromone value, and perform global pheromone update after complete processing the job. It is assumed that jobs have different requirements are independently submitted by different users. For each job, the scheduler will record the size of the job and CPU time needed by the job. The jobs that are submitted consist of different computational time, so that each job also requires different data transmission time and computation time for completion. An ant to represent a job in the grid system is created for every job and the ant will move from one resource to another with the aim to evaluate the initial pheromone value.

The initial pheromone value is calculated based on the estimated transmission time and execution time of a given job when assigned to this resource defined by:

$$PV_{rj} = \left[\frac{S_j}{bandwidth_r} + \frac{C_j}{MIPS_r * (1 - load_r)}\right]^{-1}$$

where *rj PV* is the pheromone value for job *j* assigned to resource *r*, *j S* is the size of a given job *j* and *r bandwidth* is the bandwidth available between the scheduler and resource, *j C* is the CPU time needed of job *j*, *r MIPS* is the processor speed of resource *r* and 1-*load* is the current load of resource *r*.

IV. ACO ALGORITHMS FOR THE TSP

In ACO algorithms ants are simple agents which, in the TSP case, constructors by moving from city to city on the problem graph. The ants' solution construction is guided by (artifical) pheromone trails and a priori available heuristic information. When applying ACO algorithm to the TSP, a pheromone strength is associated to each arc (i,j) where T_{ij}(t)is a numerical information which modified during the run of the algorithm and *t* is the iteration counter. If an ACO algorithm is applied to symmetric TSP instances, we always have $T_{ij}(t)=T_{ji}(t)$in applications to asymmetric TSPs (ATSPs), we will possibly have $T_{ij}(t) =/T_{ji}(t)$

The best performing ACO algorithms for the TSP improve the tours constructed by the ants applying a local search algorithm. Hence, these algorithms are in fact hybrid algorithms combining probabilistic solution construction by a colony of ants with standard local search algorithms. Such a combination may be very useful since constructive algorithms for the TSP often result in a relatively poor solution quality compared to local search algorithms. Yet, it has been noted that repeating local search from randomly generated initial solutions results in a considerable gap to the optimal solution. Consequently, on the one hand local search algorithms may improve the tours constructed by the ants, while on the other hand the ants may guide the local search by constructing promising initial solutions. The initial solutions generated by the ants are promising because the ants use with higher probability those arcs which, according to the search experience, have more often been contained in shorter tours. For example, the later does not capture the application of ACO Algorithms to network routing problems.

Procedure *ACO algorithm for TSPs*
Set parameters, initialize pheromone trails
While (termination condition not met) do
Construct Solutions
Apply Local Search % optional
Update Trails
End
End *ACO algorithm for TSPs*

Tour construction. Initially, each ant is put on some randomly chosen city.
At each construction step, ant *k* applies a probabilistic action choice rule. In particular, the probability with which ant *k*, currently at city *i*, chooses to go to city *j* at the *t*th iteration of the algorithm is:

$$p_{ij}^k(t) = \frac{[\tau_{ij}(t)]^\alpha \cdot [\eta_{ij}]^\beta}{\sum_{l \in \mathcal{N}_i^k} [\tau_{il}(t)]^\alpha \cdot [\eta_{il}]^\beta} \qquad \text{if } j \in \mathcal{N}_i^k$$

where ´*ij* = 1=*dij* is an a priori available heuristic value, ® and ¯ are two parameters which determine the relative impudence of the pheromone trail and the heuristic information, and *Nk i* is the feasible neighbourhood of ant *k*, that is, the set of cities which ant *k* has not yet visited. The role of the parameters ® and ¯ is the following. If ® = 0, the closest cities are more likely to be selected.

V. ACO ALGORITHM FOR JOB SCHEDULING

The Ant Colony Optimization (ACO) has been shown to be an effective strategy for several problems closely related to scheduling jobs in data intensive application, and it seems to be an effective strategy in this domain. The ants build their solution using both information encoded in the pheromone trail and also problem-specific information in the form of a heuristic.

1. Initialization of algorithm

All the pheromone values and parameters are initialized at the beginning of the algorithm.

2. Solution construction

The N artificial ants are used in the algorithm. The N ants set out to build N solutions to the problem based on pheromone and heuristic values using the selection rule.

3. Pheromone updating

After all ants have completed their solution at the end of the each iteration, the pheromone values are updated.

3.1 Defining pheromone and heuristic Information

The following notions are used for formulate the mathematical model
Ti – deadline given by the user i
Bi - budget of the grid user i
Pj – price unit of resource j
Wj – total workload of resource j
Cj – current workload of j
TRi – time required completing the job at resource j

When a resource j join the grid system, it should submit his quality factors in set S

$$\tau_j(0) = \sum_{i=1}^{n} s_i \cdot f_i, \quad \sum_{i=1}^{n} f_i = 1 \qquad (1)$$

e
τ j (O) – Innate performance of resource j
fi – Intensive weight age factors
si = system centric parameters

In any ACO algorithm we must first determine what Information we will encode in the pheromone trail, which will Allow the ants to share useful information about good solutions. The p pheromone value τ j (t) represent the favourability of Scheduling a particular job i onto a particular resource j at time t.

VI. CONCLUSION

EACO algorithm has successfully scheduled the jobs among resource in all conditions which leads to a balanced load network. EACO was able to solve the load balancing problem. The proposed ACO method provided the optimality in grid environment. The scheduling of the grid is smoothened using ACO. As a consequence, the load on various grids are shared among all available grids and it offers load balancing in the grid environment.

REFERENCES

[1] Casanova, H., F. Berman, G. Obertelli and R. Wolski, 2003. The apples parameter sweep template: User level middleware for the grid. Proceedings of the Proceedings of the 2000 ACM/IEEE conference on Supercomputing, Nov. 04-10, IEEE Computer Society, Dallas, Texas, pp: 60-60. DOI:10.1109/SC.2000.10061.

[2] Sadhasivam, S., & Meenakshi, K. (2009). Load balanced, efficient scheduling with parallel job submission in computational grids using parallel particle swarm optimization. *World Congress on Nature* & *Biologically Inspired Computing*,175- 180.

[3] Chandra Mohan, B. and Baskaran, R. "Survey on Recent Research and Implementation of Ant Colony Optimization in Various Engineering Applications", International Journal in Computational Intelligent Systems, Vol. 4, No. 4, pp. 556-582, 2011(e)

[4] Chandra Mohan, B. and Baskaran, R. "Priority and Compound Rule Based Routing using Ant Colony Optimization", International Journal of Hybrid Intelligent System, IOS Press ,Netherland, Vol. 8, No. 2, pp.93-97, 2011(c).

[5] Chandra Mohan, B. and Baskaran, R. "Improving network performance by optimal load balancing using ACO based Redundant Link Avoidance algorithm", International Journal Computer Science Issues, Vol. 7, Issue. 3, No. 6, pp. 27-35, 2010.

[6] Suguna, N., 2011. An independent rough set approach hybrid with artificial bee colony algorithm for dimensionality reduction. Am. J. Applied Sci., 8: 261-266. DOI: 10.3844/ajassp.2011.261.266

[7] Ka-Cheong Leung, and Victor O.K. Li, "Generalized Load Sharing for Packet-Switching Networks I: Theory and Packet-Based Algorithm", IEEE Transactions on Parallel and Distributed Systems, Vol. 17, No. 7, July 2006.

[8] Ali, A., Belal, M., A., & Al-Zoubi, M., B. (2010). Load balancing of distributed systems based on multiple ant colonies optimization. *American Journal of Applied Sciences*, *7*(3), 433-438.

[9] Chang, R., Chang, J., & Lin, P. (2007). Balanced job assignment based on ant algorithm for computing grids. *Proceedings of the 2nd IEEE Asia-Pacific Service Computing Conference*, 291-295.

[10] Nasir, H.J.A. & Ku-Mahamud, K.R. (2010). Grid load balancing using ant colony optimization. *Proceeding of The* 2nd *International Conference on Computer and Network Technology*, 23-25 April 2010, Bangkok, Thailand, 207-211.

[11] Israel Cidon, Raphael Rom, and Yuval Shavitt, "Analysis of Multi-Path Routing", IEEE/ACM Transactions On Networking, Vol. 7, No. 6, December-1999

Fault Diagnosis and Implementation in MANET

Kanwarpal Singh[1], Gagandeep Singh[2], Er. Rishma Chawla[3]
[1]Ramgharia Institute of Eng. & Technology, Phagwara, Er.kanwarpal@gmail.com
[2]Assistant Professor, GIMET,Amritsar, Gagandeep.engineer@gmail.com
[3]Assistant Professor, Ramgharia Institute of Eng. & Technology, Phagwara, er.rismachawla@gmail.com

***Abstract*— In this paper we describe here the diagnosis the dynamic topology in a faulty environment. In this paper we have optimize that type of network in which less nodes are used to reach source to destination so that data loss are reduced. We have compare the data loss after optimize the new path to reach the source to destination. In dynamic topology nodes are free to move randomly. Mobile stations from an arbitrary topology.**

Keywords- VANET, INVANET, IMANET

I. INTRODUCTION

Wireless network is a emerging as a significant aspect of Internet networking[1]. It presents a set o unique issues based on the fact that the only limit to a wireless network is the radio signal strength .There is no wiring to define membership in a network .Wireless technology has helped to simplify networking by enabling multiple computer user to simultaneously share resources in a home or business without additional or intrusive wiring. These resources might include a broadband Internet connection, network printers, data files, and even streaming audio and video. This kind of resource sharing has become more prevalent as computer users have changed their habits from using single, stand-alone computers to working on networks with multiple computers, each with potentially different operating systems and varying peripheral hardware. U.S. Robotics wireless networking products offer a variety of solutions to seamlessly integrate computers, peripherals, and data. A wireless ad-hoc network is a decentralized type of wireless network. The network is ad hoc because it does not rely on a pre-existing infrastructure, such as routers in wired networks or access points in managed infrastructure wireless networks. Instead, each node participates in routing by forwarding data for other nodes, and so the determination of which nodes forward data is made dynamically based on the network connectivity. In addition to the classic routing, ad hoc networks can use flooding for forwarding the data. An ad hoc network typically refers to any set of networks where all devices have equal status on a network and are free to associate with any other ad hoc network devices in link range[2]. Very often, ad hoc network refers to a mode of operation of IEEE 802.11 wireless networks.It also refers to a network device's ability to maintain link status information for any number of devices in a 1 link range, and thus this is most often a *Layer 2* activity. Because this is only a Layer 2 activity, ad hoc networks alone may not support a routable IP network environment without additional *Layer 2* or *Layer 3* capabilities.

II. TYPE OF WIRELESS AD HOC NETWORK

Wireless ad hoc networks can be further classified by their application:

2.1 mobile ad-hoc networks (MANET)
2.2 wireless mesh networks (WMN)
2.2 wireless sensor networks (WSN)

2.1 MANET

MANETs are a kind of wireless ad-hoc networks that usually has a routable networking environment on top of a Link Layer ad hoc network. A mobile ad-hoc network (MANET) is a self-configuring infrastructure less network of mobile devices connected by wireless links. ad hoc is Latin and means "for this purpose". Each device in a MANET is free to move independently in any direction, and will therefore change its links to other devices frequently. Each must forward traffic unrelated to its own use, and therefore be a router. The primary challenge in building a MANET is equipping each device to continuously maintain the information required to properly route traffic [5]. Such networks may operate by themselves or may be connected to the larger Internet. In mobile ad-hoc networks where there is no infrastructure support as is the case with wireless networks, and since a destination node might be out of range of a source node transmitting packets; a routing procedure is always needed to find a path so as to forward the packets appropriately between the source and the destination. Within cell, a base station can reach all mobile nodes without routing via broadcast in common wireless networks. In the case of ad-hoc networks, each node must be able to forward data for other nodes. This creates additional problems along with the problems of dynamic topology which is unpredictable connectivity changes [6].

I.2 WIRELESS MESH NETWORK

A wireless mesh network can be seen as a special type of wireless ad-hoc network. A wireless mesh network often has a more planned configuration, and may be deployed to provide dynamic and cost effective connectivity over a certain geographic area.[7] An ad-hoc network, on the other hand, is formed ad hoc when wireless devices come within communication range of each other. The mesh routers may be mobile, and be moved according to specific demands arising in the network. Often the mesh routers are not limited in terms of resources compared to other nodes in the network and thus can be exploited to perform more resource intensive functions. In this way, the wireless mesh network differs from an ad-hoc network, since these nodes are often constrained by resources. Mesh networking (topology) is a type of networking where each node must not only capture and disseminate its own data, but also serve as a relay for other nodes, that is, it must collaborate to propagate the data in the network.

I.3 WIRELESS SENSOR NETWORK

A wireless sensor network (WSN) consists of spatially distributed autonomous sensors to monitor physical or environmental conditions, such as temperature, sound, vibration, pressure, motion or pollutants 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. The development of wireless sensor networks was motivated by military applications such as battlefield surveillance; today such networks are used in many industrial and consumer applications, such as industrial process monitoring and control, machine health monitoring, and so on [4]. The WSN is built of "nodes" – from a few to several hundreds or even thousands, where each node is connected to one sensors.

III. TYPES OF MANET

MANETs are of following types

*3.1 Vehicular Ad-Hoc Networks (VANETs)***:** This[7] type of MANET is mainly used to communicate between the vehicles and the roadside equipments or just to communicate among the vehicles. A Vehicular Ad-Hoc Network or VANET is a technology that uses moving cars as nodes in a network to create a mobile network. VANET turns every participating car into a wireless router or node, allowing cars approximately 100 to 300 metres of each other to connect and, in turn, create a network with a wide range. As cars fall out of the signal range and drop out of the network, other cars can join in, connecting vehicles to one another so that a mobile Internet is created. It is estimated that the first systems that will integrate this technology are police and fire vehicles to communicate with each other for safety purposes. We can understand VANETs as subset of MANET and best example of VANET is Bus System of any University which are connected. These buses are moving in different parts of city to pick or drop students if they are connected, make a Ad hoc Network. With the Internet becoming an increasingly significant part of our lives, the dream of a WiFi-enabled city is becoming closer and closer to reality.

3.2 Intelligent vehicular ad hoc networks (InVANETs): It includes artificial intelligence that aids the vehicles to behave in intelligent manner during drunken driving, collision etc. Intelligent vehicular ad-hoc networks (InVANETs) use WiFi IEEE 802.11p (WAVE standard)and WiMAX IEEE 802.16 for easy and effective communication between vehicles with dynamic mobility. Effective measures such as media communication between vehicles can be enabled as well methods to track automotive vehicles. InVANET is not foreseen to replace current mobile (cellular phone) communication standards [5]. In VANET can be used as part of automotive electronics, which has to identify an optimally minimal path for navigation with minimal traffic intensity. The system can also be used as a city guide to locate and identify landmarks in a new city. Communication capabilities in vehicles are the basis of an envisioned In VANET or intelligent transportation systems (ITS). Vehicles are enabled to communicate among themselves (vehicle-to-vehicle, V2V) and via roadside access points (vehicle-to-roadside, V2R). Vehicular communication is expected to contribute to safer and more efficient roads by providing timely information to drivers, and also to make travel more convenient.

3.3 Internet Based Mobile Ad hoc Networks (iMANET): This type of ad-hoc network connects mobile nodes with the internet gateway node. Here the ad-hoc routing algorithms cannot be applied directly.

IV. PROBLEM FORMULATION

A node becomes faulty because of battery discharge, crash and limitation in age. An important problem in designing hosts MANET is handling failure of nodes is the distributed self diagnosis problem. In distributed self-diagnosis system each mobile node is able to diagnose the status of all nodes and knows the correct status of other nodes in the network.[7]

Each node in the system can be in one of two states faulty or fault-free. Faults can be categorized based on their duration, how it behaves after failure and occurrence of fault during diagnosis session.[9]

A. Based on the Duration

Based on duration faults can be of three types:

*4.1Transient fault***:** A transient fault can disappear without any visible event it appears in a network for short time. The recovery of transient faults from system is addressed using repeated-round techniques. A probabilistic model used for the action of faulty periods, and a fault analysis is used to obtain the optimum retry period. A transient fault is a fault that is no longer present if power is disconnected for a short time. Many faults in overhead power lines are transient in nature. At the occurrence of a fault power system protection operates to isolate area of the fault. A transient fault will then clear and the power line can be returned to service.In electricity transmission and distribution systems an automatic reclose function is commonly used on overhead lines to attempt to restore power in the event of a transient fault. This functionality is not as common on underground systems as faults there are typically of a persistent nature. Transient faults may still cause damage both at the site of the original fault or elsewhere in the network as fault current is generated.

4.2. Intermittent fault: It is problematic type of transient fault; we can't predict its appearance and disappearance in the network. An intermittent fault is occurred by several factors, some may be erect randomly, which occur simultaneously. These factors can only be identified when malfunction is occurred. Intermittent faults are difficult to identify and repair. An intermittent fault, often called simply an "intermittent", is a malfunction of a device or system that occurs at intervals, usually irregular, in a device or system that functions normally at other times. Intermittent faults are common to all branches of technology, including computer software. An intermittent fault is caused by several contributing factors, some of which may be effectively random, which occur simultaneously. The more complex the system or mechanism involved, the greater the likelihood of an intermittent fault.Intermittent faults are notoriously difficult to identify and repair "troubleshoot" because each individual factor does not create the problem alone, so the factors can only be identified while the malfunction is actually occurring. The person capable of identifying and solving the problem is seldom the usual operator. Because the timing of the malfunction is unpredictable, and both device or system downtime and engineers' time incur cost, the fault is often simply tolerated if not too frequent unless it causes unacceptable problems or dangers. For example, some intermittent faults in medical life support equipment can kill a patient. If an intermittent fault occurs for long enough during troubleshooting, it can be identified and mresolved in the usual way.

4.3 Permanent fault: Once it appears in network it remains until it removed and repaired by some external administrator. Permanent faults are simpler to deal.

B. Based on the Behaviour

Based on behaviour faults can be of two types:
4.4 Soft Fault: Soft faulted units can communicate with its neighbours but with unexpected behaviours and always give undesirable response.

4.5 Hard fault: Hard faulted units cannot communicate with its neighbours. It neither sends nor receives any information from the network.

C. Based on the Occurrence

Based on occurrence faults can be of two types:

4.6 Static fault: All faulty nodes be faulty from the starting of diagnosis session. The fault-free node can't be faulty during diagnosis session.

4.7. Dynamic fault: Fault-free node may become faulty during diagnosis session. It is hard 5to diagnosis because any node may fail after it diagnosed fault-free by any fault-free node.

D . Other Faults

Another type of fault is Byzantine fault which fail the components of a system in arbitrary ways by processing requests incorrectly. It is of two types:

4.8 Omission failures: This type of failure doesn't response for a request, e.g., crash, failing to receive a request, or failing to send a response.

4.9.Commission failures: This type of failure may respond in any unpredictable way, e.g., processing a request incorrectly, corrupting local state, and/or sending an incorrect or inconsistent response to a request.
Hardware Failures: Hardware failures can be described as failures that occur in mechanisms like disks or storage media. Hardware failures tend to offset other failures. It is recommended to utilize platforms that can monitor internal temperatures, as well as trigger alarms accordingly. With random access memories, error correcting codes (ECCs) can be utilized to identify and correct single errors and to identify two-bit errors. Software Failures: Determining the reason of a system outage can be quite intricate. Virus protection defects can cause system outages. Often, incorrect system configuration can also lead to system failures.

4.10 Network Failures: Any changes to the network design or topology of a layer of the protocol stack can have an impact on the entire network. It is therefore better to assess each layer when making any network changes.

E. Fault diagnosis

Fault identification is one of the important part in many protocols. When the actual behaviour is deviated by system or nodes of the system, a diagnosis function started to determine which node performed abnormal behaviour that is called diagnosis. Diagnosis is classified based on the occurrence of fault. It simply can be classified as static diagnosis and dynamic diagnosis. In static diagnosis, the fault does not occur during the diagnosis session; they already appeared in the networks. In dynamic diagnosis, the faults can occur during the diagnosis session, it is difficult to handle because node can be faulty after it has been diagnosed as fault-free by other node. We considered the problem of dynamic failures of node and remove those nodes from the network. Previously all work has dealt with the static fault situation where node cannot be faulty during diagnosis period.

V. ANALYSIS OF THE PROBLEM AND IMPLEMENTATION

In this figure we can see that the node moves anywhere in lxl area .the source node send the request .All node generates the hello message to start up the time.

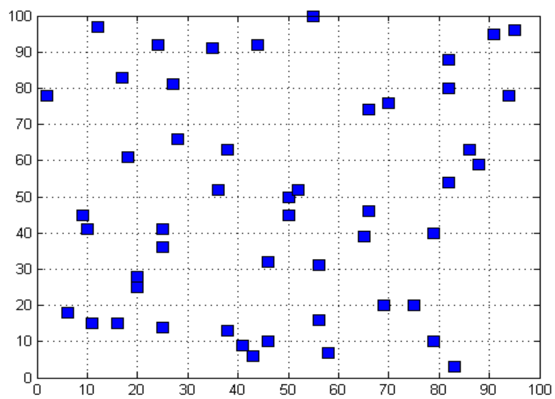


Fig.1

In this figure we can see that the actual path is generated between the source and the destination .The minimum number of nodes is 1and the maximum number of nodes is 50 is used when the actual path is generated. For measure the distance between the nodes we can use the formula mentioned in the proposed idea .in this sceanario nodes moves anywhere in lxl area. The nodes picked randomly to generate the path. The number of nodes effected on the transmission range of the data packet .when the actual path is generated between source and destination we calculated the data loss. The data loss is 0.7943 .The number of nodes effected the transmission of the data.

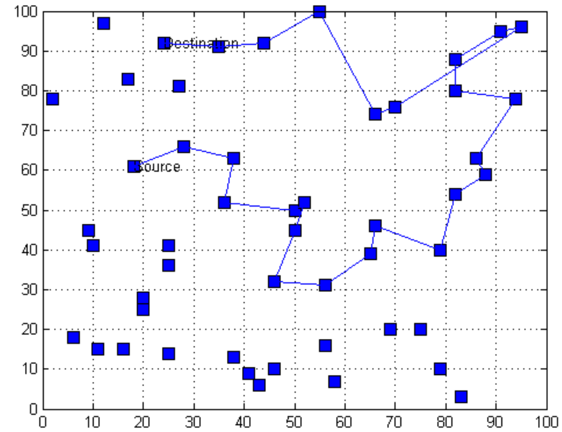


Fig.2

In the figure we can see that new path is optimized between the source and the destination .the new optimize path the less number of nodes than the actual path establish between the source and destination. After reduction the number of nodes the data loss is 0.3112.This data loss is known as a new data loss .the data loss which is computed when the actual path is establish is known as a old data loss. If we compare the new data loss with the old data loss .then we can see that the new data loss is less than the old data loss .we can say that the number of nodes effect on the transmission of the packets.

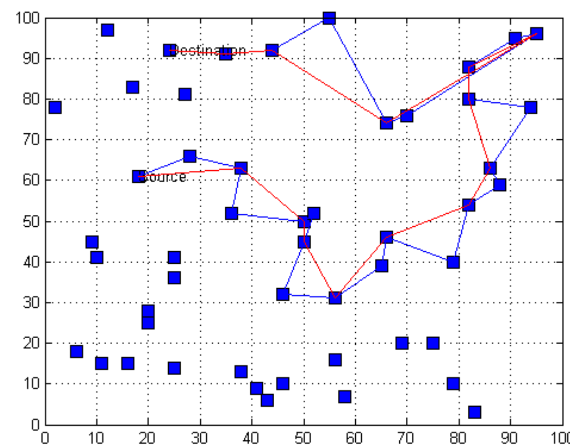


Fig . 3

VI. CONCLUSION

In this paper ,we discussed the diagnosis a dynamic topology is more complex than the static topology. We can optimize the reliable path between the source and the destination. In our research we can compare the old data loss with the new data loss .The old data loss is come from the actual path is establish source and the node. We can conclude from our research is that the number of nodes affect on the transmission of the packets. We can optimize that type of path in which less number of nodes are used that gives good throughout.

REFERENCES

[1] Clare ,loren p, GregoryJ pottie and jonathan agre (1999)" self organizing distributed sensor networks " proc accessories 99

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

[3] Imrich Chlamtac , Marco Conti , Jennifer J.-N. Liu" Mobile ad hoc networking: imperatives and challenges" Elsevier, Ad Hoc Networks 1 (2003) 13–64.

[4]] Fujian Qin and Youyuan Liu, "Multipath Based QoS Routing in MANET", Journal of Networks, vol. 4, no. 8, October 2009

[5] Jinyang Li, Charles Blake, Douglas S. J. De Couto, Hu Imm Lee, and Robert Morris Capacity of Ad Hoc Wireless Networks, in the proceedings of the 7th ACM International

[6] Laura Marie Feency .Ataxomony for routing protocols in mobile ad hoc networks technical report, institute of computer surden 1999

[7] M. Elhadef, A. Boukerche, H. Elkadiki, Diagnosing mobile ad hoc networks: two distributed comparison-based self-diagnosis protocols, in: Proceedings of the 4th ACM International Workshop on Mobility Management and Wireless Access Protocols, Terre[3] molinos, Spain, October 2006.\

[8] R. Ahlswede, N. Cai, S. Y. R. Li, and R. W. Yeung. Network information flow. *Information Theory, IEEE Transactions on*, 46(4):1204–1216, 2000.

[9] Ramasubramanian, Z. J. Haas, and E. G. Sirer. SHARP: A Hybrid Adaptive Routing Protocol for Mobile Ad Hoc Networks. In ACM MobiHoc

.

Recent Trends in Hand Gesture Recognition for PC Applications

Sahib Singh[1], Vijay Kumar Banga
M-Tech Student[1], Professor[2]
Department of Electronics and Communication Engineering, ACET, Amritsar, Punjab, India[1&2]
E-mail:dhillon_sahib85@yahoo.com[1], vijaykumar.banga@gmail.com[2]

***Abstract* - In today's digitized world, processing speeds have increased dramatically, with computers being advanced to the levels where they can assist humans in complex tasks. Yet, input technologies seem to cause a major bottleneck in performing some of the tasks, under-utilizing the available resources and restricting the expressiveness of application use. Hand Gesture recognition comes to rescue here. Interpretation of human gestures by a computer is used for human-machine interaction in the area of computer vision. The main purpose of gesture recognition research is to identify a particular human gesture and convey information to the user pertaining to individual gesture. Hand gesture recognition can be used to enhance human–computer interaction without depending on traditional input devices such as keyboard and mouse. Gesture recognition has the potential to be a natural and powerful tool supporting efficient and intuitive interaction between computers and humans.**

Keywords: Hand Gesture Recognition(HGR), Human-Computer Interaction (HCI), Input Devices (keyboard, mouse), Intuitive Interaction.

I. INTRODUCTION

Body language forms a crucial component of face-to-face conversation. However, it has been conspicuously missing from interactions in networked virtual environments. While voice communication is quickly gaining popularity in virtual worlds, non-verbal communication is usually performed with manual keyboard control or not at all. Automatic generation of body language for avatars in virtual Worlds would make nonverbal communication a natural part of Social interaction in these environments[1].

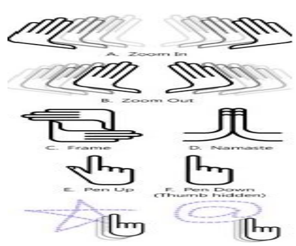


Fig.1: Some Gesture Examples

Interactive tabletop systems aim to provide a large and natural interface for supporting direct manipulation of visual content for human-computer interactions. They present a system that tracks the 2D position of the user's hands using color markers on a table surface, allowing the user drawing, writing and manipulating virtual objects over this surface[4].

There is one more application, which uses motion detection as its first step, and then does some interesting routines with the detected object – HGR. Let's suppose there is a camera which monitors an area[9]. When somebody gets into the area and makes some hands gestures in front of the camera, the application should detect the type of the gesture, and raise an event, for example. When a hands gesture is detected, the application could perform different actions depending on the type of the gesture. For example, a gesture recognition application could control some sort of device, or another application sending different commands to it depending on the recognized gesture. Interpretation of human gestures by a computer is used for human-machine interaction in the area of computer vision. The main purpose of gesture recognition research is to identify a particular human gesture and convey information to the user pertaining to individual gesture. From the corpus of gestures, specific gesture of interest can be identified, and on the basis of that, specific command for execution of action can be given to the machine[3]. Overall aim is to make the computer to understand human body language, thereby bridging the gap between machine and human. Hand gesture recognition can be used to enhance human–computer interaction without depending on traditional input devices such as keyboard and mouse.

Gesture recognition is a topic in computer science and language technology with the goal of interpreting human gestures via mathematical algorithms. Gestures can originate from any bodily motion or state but commonly originate from the hand. Current focuses in the field include hand gesture recognition. Many approaches have been made using cameras and computer vision algorithms to interpret sign language. However, the identification and recognition of posture, gait and human behaviours is also the subject of gesture recognition techniques. Visual interpretation of hand gestures can help in achieving the ease and naturalness desired for Human Computer Interaction (HCI). This has motivated many researchers in computer vision-based analysis and interpretation of hand gestures as a very active research area [10].

This article examines the recent existing techniques for the Hand Gesture Recognition by defining the underlying technologies and how they work.
The rest of the paper is organised as follows: In section II, some of the related works are presented along with their basic methodology. In section III, some existing problems and challenges are discussed and in section IV we finally present our conclusion and discuss the future work scope.

II. RELATED WORK

Hand Gesture Recognition is a very active area of research. In the past researchers have explored a variety of methods to Gesture Recognition applications. We present some of them in this section.

The work given in [7] describes research on HCI through real time gesture recognition and three dimensional hand-tracking. Using two cameras that are focused at the user's hand, the system recognizes three gestures and tracks the hand in three dimensions.The system can simultaneously track two fingers (thumb and pointing finger)and output their poses. The pose (X,Y,Z,a: , c)for each finger consists of three positional coordinates and two angles (azimuth and elevation). By moving the thumb and the pointing finger in 3D, the user can control 10 degrees of freedom in a smooth and natural fashion. This system is used as a multidimensional input-interface to computer games, terrain navigation software and graphical editors. The two fingers can also be used as a virtual robot arm allowing the user to grasp and move virtual objects in 3D (see Figure 2).

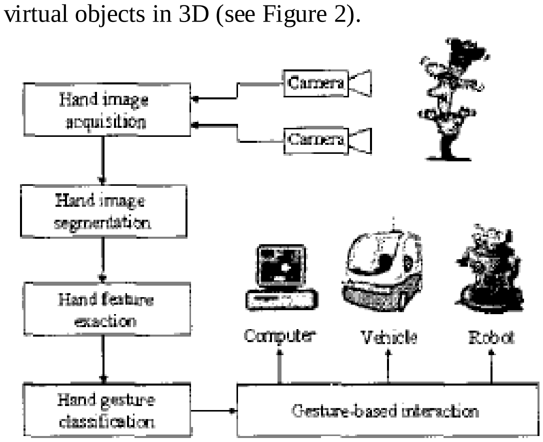


Fig.2: Process of Hand Gesture Recognition

This system is much more natural and intuitive to use compared to traditional input devices.The system is user independent, and operates at the rate of 60 Hz.Our system is meant to augment (and, in some cases, even replace) devices such as the computer mouse, joystick, or track-ball.

The work presented in [9] identifies hand gestures for use in controlling computer programs, such as browsers, Powerpoint, or any other applications. The goal is two fold. First, gesture recognition can complement other forms of human computer interaction (such as keyboard, mouse, and voice), but provides such control at a distance without touch. For example, this gesture technology will allow a person to easily interact with a virtual vehicle in an intuitive and natural manner, greatly increasing the utility of the virtual reality design system, and enabling systems to be designed and developed in a more efficient and effective manner. Second, it can be used as a convenient computer interface to the millions of people who are unable to adequately use typical computer interaction techniques. The system uses a standard windows PC and an inexpensive commercially available USB camera. This arrangement does not require any connection to the user (either by tags or active beacons worn on the hand).

The technique given in [3] is a new sub-gesture modeling approach which represents each gesture as a sequence of fixed sub-gestures (a group of consecutive frames with locally coherent context) and provides a robust modeling of the visual features. We further extend this approach to the task of gesture spotting where the gesture boundaries are identified using a filler model and gesture completion model. Experimental results show that the proposed method [ ] outperforms state-of-the-art Hidden Conditional Random Fields (HCRF) based methods and baseline gesture spotting techniques. Recent advances in computer vision and machine learning have led to a number of techniques for modeling gestures in real-time environment. Modeling each video sequence using a single gesture model usually leads to lower performance owing to a high temporal variability. To address this issue, this method proposes to model each gesture as a sequence of smaller sub-gesture units. This is inspired by recent study in speech and handwriting recognition. In speech recognition, a phoneme is defined as smallest segmental unit employed to form an utterance (speech vector). Similarly in handwriting recognition, a word can be represented as a sequence of strokes, where each stroke is the smallest segmental unit. This inspires our definition of gesture in a natural and intuitive manner, where each gesture can be represented as a sequence of contiguous sub-gestures denoting a consistent action performed by the user. Typical examples of such sub gestures would be moving both hands from rest to horizontal positions, moving both hands away from each other and resting both hands which if performed in sequence would model a "zooming in" gesture (see fig 3).

Fig.3: Example of Zoom in gestures and corresponding sub-gestures

Most of the earlier techniques on gesture recognition were based on one-vs. all models, where a separate model is trained for each gesture e.g. HMMs and its variants. Recent advances in gesture recognition research suggest that multiclass (one model for all gesture classes) models like HCRF which are considered to be state of art for gesture recognition, outperforms the one-vs.-all models since they jointly learn the best discriminative structure among gestures. HCRF provides us an excellent framework for modeling each gesture as a combination of sub-gestures by sharing hidden states among multiple gesture classes. However, this sharing is usually implicit and there is no way to explicitly define a sub-gesture sequence for a given gesture which might be useful for continuous gesture recognition. Moreover, HCRF training algorithms are computationally very expensive as compared to HMMs. To address these issues, this method propose a novel variant of HMM for gesture recognition which combines the advantages of HCRF and HMM. Our proposed model explicitly takes the sequence of sub-gestures for a given gesture (gesture grammar) and uses it to learn a model for each gesture. This provides us greater flexibility in modeling gestures as the observations (image features) are used to model a smaller sequence of actions (sub-gestures) as compared to a larger sequence (gesture) in the previous methods. Moreover, it also provides a simple framework for explicit sub-gesture modeling and is computationally efficient. Also, the proposed approach does not require exact segmentation of actions in a gesture. The second advantage of this approach is that it can be easily extended to other gesture related tasks such as gesture spotting. Gesture spotting is an alternative technique to gesture recognition in real time scenarios, where the goal is to locate the gesture boundaries (start and end of a gesture) as well as gesture label. Most of the spotting frameworks rely on learning a threshold or using heuristic rules to detect the start and end frames of a gesture. This framework addresses this limitation by using a combination of filler model and gesture-completion model to locate gesture boundaries in a probabilistic framework.

The work given in [15] gives a solution to the first stage of a Hand Gesture Recognition system i.e. hand detection and movement tracking. Recognizing gestures is a complex task which involves many aspects such as motion modeling, motion analysis, pattern recognition and machine learning. Meaningful hand gestures can be classified as static hand postures and temporal gestures (motion patterns). Hand postures appear statically as a combination of different finger states, orientations, and angles of finger joint [3]. Hand modeling is the main and also the most difficult task in recognizing postures, as fingers are usually articulated, self-occlusion makes the detection and tracking local hand configuration challenging. Temporal gestures, on the other hand, represent the meaning of gesture by hand movement rather than the details of finger appearance. In computer vision (CV) based solutions, hand gestures are captured by web cameras which used as input. Due to the resolution of the videos, only a general sense of the figure state can be detected, we focus on the gesture represented by hand movement in this paper. Procedures in a general framework of gesture recognition system are show in Fig.4. The recognition rate of the system will highly depend on the accuracy of this feature extraction stage, and the processing frame rate of the system is also dominated by the efficiency of hand detection. There are two main approaches for detecting the hand in a video frame. The first one is skin color based hand detection (SCHD). The boundaries of skin clusters can be trained from a set of sample frames/images to form the skin color model .The

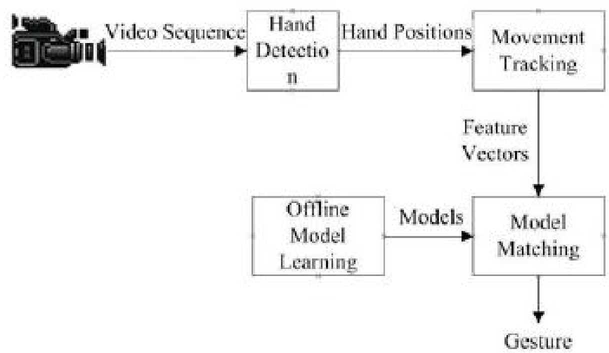


Fig.4: Framework of Hand Gesture Recognition System

statistical skin color model is learned by manually label pixels in the training set as skin/non-skin pixels, this is a trivial work and the model may not be robust on a different test data set. A skin classifier can explicitly define the boundaries of skin cluster in *RGB* color space as a set of rules:

P(R, G, B) is classified as a skin pixel if:

$R > 95$ and $G > 40$ and $B > 20$ and
$Max\{R, G, B\} - min\{R, G, B\} > 15$ and
$|R-G| > 15$ and $R > G$ and $R > B$

There are some drawbacks of SCHD solutions:

1. Its computationally expensive as it's a pixel-wise classification approach.
2. Skin-liked objects may also be classified as part of hand.
3. Lighting reflection may fail the skin detection. A few works considered this problem.
4. The range of skin color varies among different human species.

The second approach is background subtraction based hand detection (BSHD). When the camera is fixed and a video frame contains only pure background is available, foreground objects in other frames can be extracted by subtracting them with the background frame. BSHD overcomes SCHD both on performance and efficiency aspects since it takes correlation between frames into consideration, objects are detected by the measurement of inter-frame difference.

The matter presented in [14] describes a whole new technique –The Sixth Sense - Sixth Sense is a wearable, gestural interface that augments our physical world with digital information, and lets us use natural hand gestures to interact with that information. Sixth Sense brings intangible, digital information into the tangible world, and allows us to interact with this information via natural hand gestures. SixthSense frees information from its confines, seamlessly integrating it with reality, thus making the entire world your computer. The SixthSense prototype comprises a pocket projector, mirror, and camera worn in a pendant-like mobile device. Both the projector and the camera are connected to a mobile computing device in the user's pocket. The system projects information onto the surfaces and physical objects around us, making any surface into a digital interface; the camera recognizes and tracks both the user's hand gestures and physical objects using computer-vision-based techniques. SixthSense uses simple computer-vision techniques to process the video-stream data captured by the camera and follows the locations of colored markers on the user's fingertips (which are used for visual tracking). In addition, the software interprets the data into gestures to use for interacting with the projected application interfaces. The current SixthSense prototype supports several types of gesture-based interactions, demonstrating the usefulness, viability, and flexibility of the system.

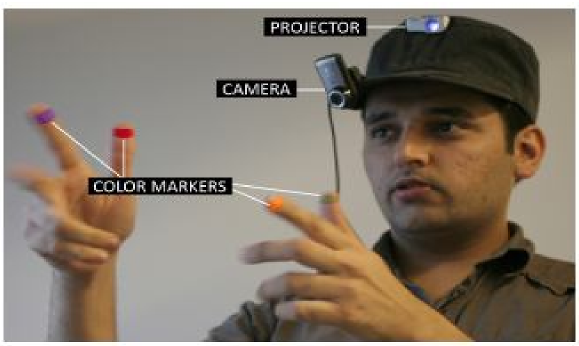


Fig.5: WUW Prototype System

III. EXISTING PROBLEMS & CHALLENGES

Appreciable work has been done on gesture recognition in the past using different methods. But there have been a few problems around.

1. One common problem was that in most of the previous work, system was trained (through ANNs or HMMs) to identify only a limited set of hand gestures and hence only allowing the partial control of the system limiting the full control.

2. The methods used to track the movement of the moving hand of user were either based on Skin Colour Based Hand Detection (SCHD) or Background Subtraction Based Hand Detection (BSHD). These methods are difficult to implement as one has to set skin colour threshold values for different persons and background subtraction method is heavily affected by noise levels in the background.

3. Most of the earlier techniques on gesture recognition were based on one-vs-all models, where a separate model is trained for each gesture e.g. Hidden Macrov Models(HMMs) and its variants.

4. Recent developments in gesture recognition introduced multiclass models (one model for all gesture classes) like Hidden Conditional Random Fields (HCRF) which are usually very expensive.

5. The prototypes used in some techniques are very expensive e.g. the current prototype system of WUW costs approximately $350 to build.

IV. FUTURE WORK SCOPE AND CONCLUSION

In this paper we have reviewed a few methods which are currently in use "IN HAND GESTURE RECOGNITION FOR PC APPLICATIONS" It has been observed that these methods and techniques are alresdy making a very good contribution for the society and with the ever-increasing diffusion of computers into the society, it is widely believed that present popular mode of interactions with computers (mouse and keyboard) will become a bottleneck in the effective utilization of information flow between the computers and the human. These traditional devices are on their way to obsolete in the time to come. Vision based Gesture recognition has the potential to be a natural and powerful tool supporting efficient and intuitive interaction between the human and the computer. With the development of information technology in our society, we can expect that computer systems to a larger extent will be embedded into our environment. These environments will impose needs for new types of human computer-interaction, with interfaces that are natural and easy to use.

REFERENCES

[1] J. Davis ,M. Shah "*Visual gesture recognition*" IEEE Vol.141,NO.2, University of Florida, USA , pp 101-106, 1994

[2] Vladimir I. Pavlovic, Rajeev Sharma, and Thomas S. Huang "*Gestural Interface to a Visual Computing Environment for Molecular" IEEE 2nd Int. Conf. on Automatic Face and Gesture Recognition ,Killington ,VT USA ,pp 30-35, 14-16 October 1996*

[3] Manavender R. Malgireddy, Jason J. Corso "*A Framework for Hand Gesture Recognition and Spotting Using Sub-gesture Modeling"* IEEE 20th Int. Conf. On Pattern Recognition , Istanbul , Turkey , pp 3780-3783 , 23-26 August 2010

[4] Norman I. Badler "*Real-Time Virtual Humans*" IEEE 5th Pacific Conf. on Computer Graphics and Applications, Philadelphia, USA, pp 4-13, 13-16 October 1997

[5] Dan Ionescu,Cristian Gadea and Shahidul Islam "*An Intelligent Gesture Interface for Controlling TV Sets and Set-Top Boxes*" IEEE 6th Int. Symposium On Applied Computational Intelligence and Informatics , Ottawa, ON, Canada, pp 159-164,19-21 May 2011

[6] Takahiro Watanabe, Masahiko Yachida "*Real-Time Gesture Recognition Using KL Expansion of Image Sequence"* IEEE Vol. 2, Int. Conf. On Proceedings of Intelligent Robots and Systems ,Grenoble , France , pp 973-979 , 7-11 September 1997

[7] Jakub Segen and Senthil Kumar "*Human-Computer Interaction using Gesture Recognition and 3D Hand Tkacking"* IEEE Vol. 3, Int. Conf. On Image Processing , Chicago , USA , pp 188-192 , 4-7 October 1998

[8] Senthil Kumar and Jakub Segen "*Gesture based 3D Man-Machine Interaction using a Single Camera"* IEEE Vol. 1, Int. Conf. On Multimedia Computing and System ,Florence , USA , pp 630-635 , July 1999

[9] Charles J.Cohen, Glenn Beach,Gene Foulk " *A Basic Hand Gesture control System for PC Applications"* IEEE 30th Int. Conf. On Applied Imagery Pattern Recognition ,Washington DC , USA , pp 74-79 , 1-12 October 2001

[10] Kye Kyung Kim, Keun Chang Kwak and Su Young Chi "*Gesture Analysis for Human Robot Interaction*" IEEE Vol. 3, 8th Int. Conf. On Advanced Communication Technology , Phoenix Park USA , pp 1824-1827, 20-22 February2006

[11] Yingen Xiong, Bing Fang, and Francis Quek "*Extraction of Hand Gestures with Adaptive Skin Color Models and its Applications to Meeting Analysis*" IEEE 8th Int. Symposium On Multimedia , San diego USA , pp 647-651, December 2006

[12] Jonathan Alon, Quan Yuan "*A Unified Framework for Gesture Recognition and Spatiotemporal Gesture Segmentation*" IEEE Vol. 31, No. 9,Transactions on Pattern Analysis and Machine Intelligence ,Chicago , USA , pp 1-18 , September 2009

[13] Pranav Mistry and Pattie Maes " *Sixth Sense: A Wearable Gestural Inrerface*" MIT Media Laboratory , New York ,USA ,2009

[14] Pranav Mistry, Pattie Maes, Liyan Chang " *WUW - Wear Ur World - A Wearable Gestural Interface"* 27th international conference extended abstracts on Human factors in computing systems - CHI EA, Boston USA, 2009

[15] Hong-Min Zhu and Chi-Man Pun *"Movement Tracking in Real-time Hand Gesture Recognition"* IEEE 8th Int. Conference On Computer and Information Science, Yamagata Japan ,pp 240-245 , 18-20 August 2010

[16] Shuai Jin, Jianxun Luo et al "*Interaction and Control with the Auxiliary of Hand Gesture*" Int. Conference On Information Science and Technology, Nanjing, Jiangsu , China, pp 1362-1366 , 26-28 March 2011

Performance Optimization in MANETs by Fault Measurement and Load Balancing

Jayant Vats[1], R.K. Singh[2], Rakesh Jaitly[3] and Takshi Gupta[4]
[1, 3,4]Amritsar College of Engineering & Technology, Amritsar, Punjab, INDIA
[2]Uttrakhand Technical University, Dehradun, Uttrakhand, INDIA
Email: jayantvasu@gmail.com, rksingkec12@rediffmail.com, rakesh_jaitly@acetedu.in, takshi_gupta12@yahoo.com

***Abstract*-Wireless sensor networks are a set of small, low cost, low power sensing devices with wireless communication capabilities. A wireless sensor networks is totally based on the limiting factor i.e. energy consumption. A wireless sensor network consists of large number of sensor nodes distributed or scattered in particular network region. MANETs consist of node that is highly mobile, so in particular the in range of the nodes is very important. Each device in a MANETs is free to move independently in any direction, and will therefore change its links to other devices frequently. Each must forward traffic unrelated to its own use, and therefore be a router. The primary challenge in building a MANETs is regarding the equipping of each device to continuously maintain the information required to properly route traffic. In this paper, we introduce the algorithm which is a security based power awareness algorithm that deals with the security of data transmitting nodes in MANETs. The security has been the great issue regarding the data transmission in MANETs along with the performance optimization of the MANETs. The FTA is an enhanced form of AODV protocol that has the ability of self recovering regarding the security issues of the network structure. The simulations are performed using the NS2 simulator [21] and the results obtained shows that the consideration of security and the mobility factors enhance the performance of the network model and thus increases the throughput of the ad hoc networks.**

Keywords: wireless sensor networks, security based power awareness, energy efficiency, and security.

I. INTRODUCTION

MANET are a kind of wireless ad hoc networks that usually has a route able networking environment on top of a Link Layer ad hoc network. They are also a type of mesh network, but many mesh networks are not mobile or not wireless. The growth of laptops and 802.11/Wi-Fi wireless networking have made MANET a popular research topic since the mid- to late 1990s [6] [7] [23]. Many academic papers evaluate protocols and the abilities assuming varying degrees of security within a bounded space, usually with all nodes within a few hops of each other and usually with nodes sending data at a constant rate. Different protocols are then evaluated based on the packet drop rate, the overhead introduced by the routing protocol, and other measures. The concept of our model is based on CPACL-AODV protocol that has been given on basis of cross layer design [6] [22]. The FTA algorithm given in this paper is the enhancement of the above written algorithm. In this paper, we define the efficiency of the MANETs and include the factor of mobility and security in it. The important factor of the mobility and security explains the behavior of the network model when the mobility and security of the node is considered, this means that instead of taking the readings by considering the nodes constant at particular instance of time, the varying behavior is considered. The rest of paper includes related work, system model, and energy efficiency considering the security and mobility, and numerical result along with the performance evaluation and finally the future aspects and conclusion.

II. RELATED WORK

The related work, includes the cost based power aware cross layer routing protocol that has been done on the basis of AODV protocol [23] [22] [6], this increases the throughput and the performance of the ad hoc networks but on the cost of battery life of the nodes, the major concern regarding this is the getting the node into dead state and also the mobility factor is neglected that has adverse effect on making a node dead. In another related work [25] PAMAS protocol has been proposed, PAMAS uses two different channels, for data transfers and separate channel for signal transferring. The signaling channel tells the nodes when to power off their RF devices if packet is not being transmitted nor received. But the problem is regarding the issues that arise when a node gets into dead node, is the false transfer that gets into wastage of the energy left out and thus not an efficient method of transferring, so for this we give FTA Algorithm.

III. SYSTEM MODEL

The network model we considered comprises of k number of hops, hops here are the nodes, and the nodes here considered are to be single channel node. This means for k number of nodes there is k number of channels. Thus, if two nodes are communicating at a time, then we have k-1 number of relaying nodes in the network model. The distance between the source and the destination is denoted by d. the distance between the relaying nodes can be decided on basis of the dynamic routing considered or it can be given on mathematical computations, this means that the distance between the relaying nodes will be less than the actual distance between the source and destination. Thus, if we consider a constant, let this constant be α_n then, from the theoretical analysis [7], we obtain that this value is multiplied with the total distance to obtain the actual distance between the relaying nodes then this value should be positive and less than one, if on computations the de that is the distance between the relaying nodes come out to be greater than d on summation, it means these nodes have the probability of being arranged in straight line. Thus, the distance between the relaying nodes will be:

$$de = \sum_{n=0}^{k} \alpha d \text{……….[7].}$$

Another factor considered is the §, the factor included by us for mobility based analysis of the MANETs. As the attenuation loss for the structure is given by in one of the papers is:

$$Pr = \beta \frac{Pout}{d\eta} \text{…………….[7]}$$

The mobility introduces another simple concept. If the mobility of the structure nodes is more, the attenuation has greater effect but if the nodes are considered to be at rest then, the attenuation comes out to be so small that it can be neglected. Thus, the modified formula for the attenuation loss in a network model (Fig. A) will be:

$$Pr = \beta \frac{Pout}{d\eta} \text{ ʍ}$$

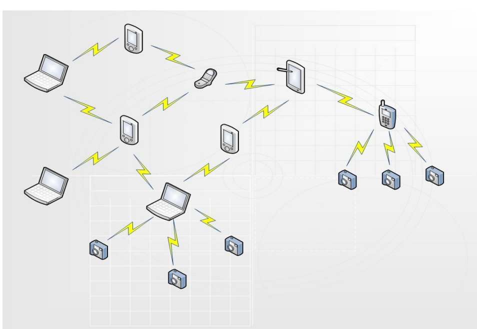


Fig: A ….[4]

Here, Pr is the attenuation loss in the MANETs, β is the antenna constant, d is the end to end distance between source and destination, η is the path loss constant such that $2 < \eta < 4$ and ʍ is the mobility factor. The mobility can be computed by analyzing the movement in terms of number of bits transferred per second per meter of the network model. Here, *Pout* = *fo* (*Pin*), which is based on the working power amplifier present in each of the node. The concept says that *Pout i*s the functional dependent on the input power (*Pin*) supplied to the power amplifier [7] [10]. The result of our computations shows that in case of the static nodes, the attenuation loss is negligible as the term gets very much small for numerical computations. The security of the nodes is dependent on the arrival of the data that is the form on which the data is processed at the receiving node. The previous concept of implementing of RSA algorithm has been used already that allows the security in transmission to some extent but what if the there is some error in the data, in that case, the retransmission takes place. But this increases the time of communication and the delays caused causes the inefficient way of delivery of packets. So, for efficient and recovery of data the FTA algorithm has been proposed in this paper.

IV. FAULT TOLERANCE ALGORITHM (FTA)

In this section of paper, we give the security based energy efficient algorithm that has the ability of significantly solving the issue of lost data at the node and also do not allow the node to get into the dead state. Before describing our algorithm, the CPACL-AODV algorithm basics are that this is a reactive routing algorithm that maintains the path as long as it is required by the sender toward the destination [6].

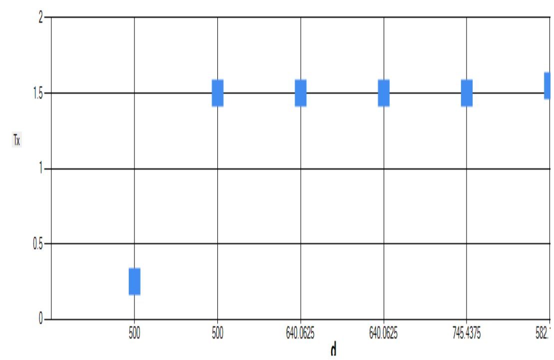
In FTA algorithm, the node on receiving the data from the previous node sends the special signal giving the details of the energy left with it after processing the obtained data, thus, the node transferring and the node to which the data is being transferred maintains a table, that holds the dynamic values regarding the transmitters transmitting energy Tx and receivers processing energy Ep. These signals also include the special security signals regarding the security of the data, these signals contain the signals in form of true and false, the standard RSA algorithm is applied and the transmission is checked, if the data is delivered correctly, then the true is included in the security column of the table maintained at each transmitting node. Thus, this helps to calculate the level of correctness in transmission of data. In case, the data transmitted is incorrect or has some errors then, there is always a chance of transmission that can lead to wastage of energy of the node and the node might become dead. Thus, the node before transmitting the data checks if a node has its energy greater than the threshold energy, the threshold energy is the minimum amount of energy required by a node to process the data obtained from previous node and further transmit towards the node nearer to the destination, thus, if the node holds the above condition can participate in transmission process and in opposite case, a dynamic routing will be performed that checks for another node that holds the condition and is nearer toward the destination. This will prevent a node from getting into the dead state and thus the energy can be increased without affecting the transmission process and the performance of the MANETs. Thus, for this the mobility of the MANETs node has tremendous affect on the performance and efficiency evaluation of the system, if the mobility of nodes is more than the dynamic routing cannot be performed easily, the reason for this is that the routing can be performed efficiently only if mobility is less or in other words more bytes of data is transferred per second per meter of the network model. Therefore, the various possibilities of the FTA algorithms output is:

- Either the node has to perform routing at start of the transmission process.
- Other condition can be intermediate routing that is performed if node has energy less than the threshold energy.
- Increasing the node energy, either by battery backups or by sending the energy packet to the node.

From the above pseudo code, it is clear that for enhanced and better transmission of data, the mobility should be less and thus, it should be maintained by transferring more number of bits in given metric area of network model.

*Example of Table maintenance by a node:

Head Node	Linked Node	Ep (in microns)	Security	Energy Left
A	F	0.25	True	1.24
	H	0.25	True	1.03
	K	0.25	True	1.11



*Nodal structure.

V. PERFORMANCE EVALUATION AND NUMERICAL RESULTS

The evaluation of our algorithm is evaluated using NS2 Simulator. The number of nodes considered is 50 in 1500 * 1500 m^2 Network Area. The packet size was considered to be 1024. In the beginning of the simulations the battery consumptions were about 0.25 for processing and 0.34 units for transmission purpose thus giving the overall consumption of 0.59 for single node at transfer of one packet. The result shows that the when the node has energy below threshold energy, the dynamic route is adopted on basis of least distance and packet is transmitted to destination without failure. Also, with the advent of mobility factor, it is achieved that when the mobility is more in terms of transfer of packets w.r.t. 1 m^2 area, the performance increases. Graphical analysis of the FTA algorithm is done with AODV protocol as it forms the basis of our algorithm and the results shows that the performance is greatly influenced by the mobility factor and thus, gives more concerned data that can be practically adopted. Following, are the graphical results obtained from the simulations performed with NS2 simulator.

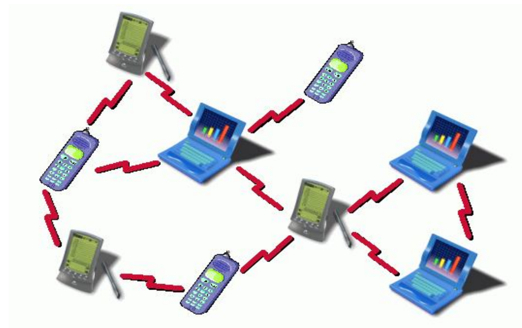


Fig: 1

The above graph shows the relation between the transmitter energy of node and the distance dependence. The values in the graph are varied due to factor of relative transmission between nodes. The graphs below are the resultants of the simulations showing the dependence of the distance, bandwidth, mobility, hop count for network model. The graphical interpretation of the simulations shows that the efficiency of the system is increased by the FTA algorithm over the conditions of dead node.

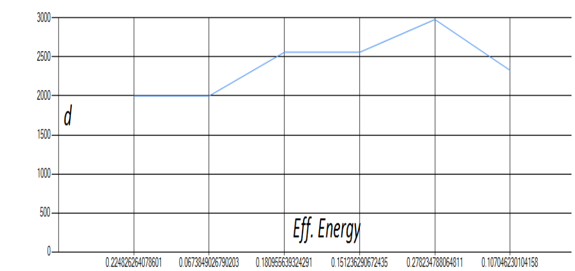


Fig: 2

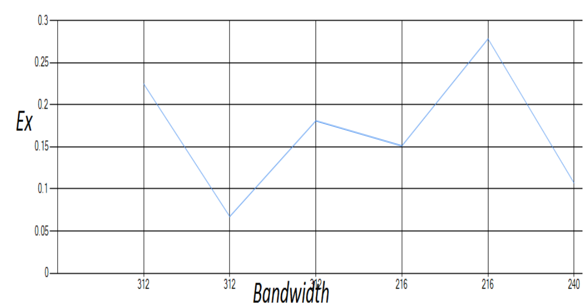


Fig: 3

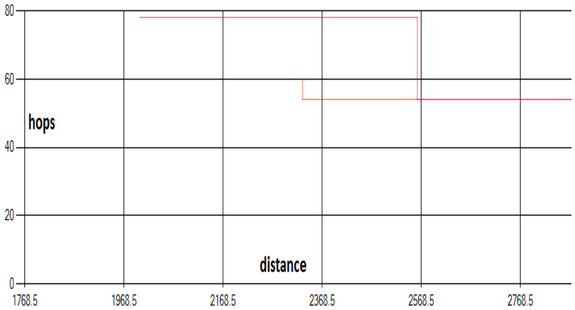


Fig: 4

VI. CONCLUSION

The paper proposed security based power awareness algorithm that solved the problem of security issues and the dead node with energy efficient transmission. Also, this reduced the various overheads of the network layer. The technique is practically adoptable. Also, the algorithm prevents the termination of the transmission process and loss of energy. In future, the advancement can be made in terms of selecting the different ways of providing energy to node with energy lower than the threshold energy and the security backups method can also be imposed in future.

REFERENCES

[1] Sunil Taneja and Ashwani Kush:" A Survey of Routing Protocols in Mobile Ad Hoc abstract". International Journal of Innovation, Management and Technology, Vol. 1, No. 3, August 2010.
[2] M.Saravana karthikeyan, M.Murali, Dr.S.Sujatha:"Identifying performance metrics to maximize Manet's throughput"; 2010 International Conference on Advances in Computer Engineering.
[3] Manika Vinay Rali, Min Song, Sachin Shetty:"Virtual wired transmission scheme using directional antennas to improve energy efficiency in Wireless Mobile Ad Hoc Network"; 978-1-4244-2677-5, IEEE 2008.
[4] http://www. csi.uoregon.edu.
[5] Sehoon Kim, Jinkyu Lee and Ikjun Yeom:" Modeling and Performance Analysis of Address Allocation Schemes for Mobile Ad Hoc Networks", IEEE transactions on vehicular technology, vol. 57, NO. 1, JANUARY 2008.
[6] Rekha Patil, Dr. A. Damodaram:"cost basedd power aware cross layer routing protocol for Manet"; 2008 IJCSNS.
[7] Changchun Bae and Wayne E. Stark:"A Tradeoff between Energy and Bandwidth Efficiency in Wireless Networks"; 2007 IEEE.
[8] V. Rodoplu and T. H. Meng: "Bits-per-Joule capacity of energy-limited wireless networks," IEEE Transaction Wireless Communications, vol.6(3), pp.857-865, March 2007.
[9] B. Rankov and A. Wittneben: "Spectral efficient protocols for half-duplex fading relay channels," IEEE Journal on Selected Areas in Communications, vol. 25, pp.379-389: Feb. 2007.
[10] O. Oyman and S. Sandhu: "Non-ergodic power-bandwidth tradeoff in linear multihop wireless networks," in Proc. IEEE International Symposium on Information Theory (ISIT'06): 2006.
[11] C. Bae and W. E. Stark: "Energy and bandwidth efficiency in wireless networks," in Proc. International Conference on Communications Circuits and Systems (ICCCAS'06): vol. 2, pp.1297-1302, June 2006.
[12] M. Sikora, J. N. Laneman, M. Haenggi, D. J. Costello, and T. E. Fuja: "Bandwidth and power efficient routing in linear wireless networks," Joint Special Issue of IEEE Transaction Information: Theory and IEEE Transaction Networking, vol. 52, pp.2624-2633, June 2006.
[13] Sunho Lim, Chansu Yu, and Chita R. Das: "Rcast: A Randomized Communication Scheme for Improving Energy Efficiency in MANETs". Proceedings of the 25th IEEE International Conference on Distributed Computing Systems (ICSCS'05).
[14] Lan Wang, Yang Xiao: "Energy Saving Mechanisms in Sensor Networks"IEEE 2005.
[15] Wei Liu, Yanchao Zhang, and Yuguang Fang: "Conserve Energy through multiple- packets transmission in heterogeneous Mobile Ad Hoc Networks": 2005.
[16] H. Wang and W. E. Stark: "Transport efficiency for wireless network," IEEE Transaction Communications, 2004.
[17] R. Zheng and R. Kravets:" On-Demand Power Management for Ad Hoc Networks". In *Proc. IEEE INFOCOM*, pages 481–491, 2003.
[18] Wing Ho Yuen, Chi Wan Sung:"On energy efficiency and network connectivity of mobile ad hoc network". Proceedings of the 23rd International Conference on Distributed Computing Systems (ICDCS'03).
[19] W. E. Stark, H. Wang, A. Worthen, S. Lafortune, and D. Teneketzis: "Low-energy wireless communication network design," IEEE, Wireless Communications pp.60-72, Aug. 2002.
[20] Wayne Stark, Hua Wang, Andrew Worthen, Stephane Lafortune, and Demosthenis Teneketzis:" low – energy wireless communication Network Design": IEEE 2002.
[21] Network simulator-2 www.isi.edu/nanam/ns/.
[22] C. K. Toh:"maximum battery life routing to support ubiquitious mobile computing in wireless ad hoc networks"; 2001 IEEE.
[23] C. E. Perkins, E.M. belding-Royer and S.R. Das:" Ad hoc on demand distance vector (AODV) routing"; IETF.
[24] P. Gupta and P. R. Kumar, "The capacity of wireless networks:" IEEE Transaction Information Theory, and vol. 46 (2), pp.388-404, Mar 2000.
[25] S. Singh and C.S. Raghavendra:" PAMAS – power aware multi-access protocol with signaling for ad hoc networks"; ACM 1998.

Application of Open UPQC for Improving the Power Quality in a Power System Network with Induction Motor Load

Simarjeet Kaur[#]

[#]Assistant Professor,EEE Department, ACET Amritsar
simu30sidhu@gmail.com

***Abstract*—POWER QUALITY (PQ) is an important issue to certain customers. For this reason, many utilities could sell electrical energy at different prices to their customers, depending on the quality of the delivered electric power and customers are ready to pay for it.[1] The device which can improve PQ is OPEN unified power quality conditioner (UPQC), composed of a series converter as main unit installed in the medium-voltage/low-voltage (LV) substation, along with several shunt converter units connected close to the end users. The series and parallel units do not have a common dc link, so their control strategies are different than traditional UPQC control techniques. The UPQC is a power conditioning device able to compensate all kinds of power quality faults. In this paper the UPQC is used to filter both current and voltage harmonics and to compensate the voltage dip. The proposed solution has been analyzed and described using a model having PI Controller and field oriented induction motor drive as load. The simulation results based on MATLAB/Simulink are discussed to support the concept developed in the paper.**

***Index Terms*- Series converter, shunt converter, open unified power-quality conditioner (UPQC), power quality (PQ), adjustable speed drive (ASD), field oriented control induction motor.**

I. INTRODUCTION

The quality of the power is having disability by several reasons. This degradation of the quality of the power is seen e.g. as supply interruptions, transient overvoltage dips, harmonics and voltage unbalance. The use of non-linear and unbalanced loads such as field oriented induction motor drive in distribution systems is increasing drastically, due to which the currents in the network become unbalanced and distorted. The THD (total harmonic distortion) basically depends on type of load for example adjustable speed drives when used produce a waveform which is not a pure sinewave and

includes harmonic distortion which is supplied to the motor. The harmonics are multiples of a fundamental frequency with a current component and the current component will create 5 to 8 percent extra heat in the motor. As it is a solid state electronic load, it will also cause distortion to be induced on the input electrical power supply. This can severely distort the electrical power supply within the facility and if not properly protected, can hinder the operation of other devices As a result of this, there is a need of custom power devices like dvr, dstatcom, upqc [10]. . Each of these custom power devices has its own benefits and limitations. The most modern power conditioning device is open unified power quality conditioner (UPQC).The open unified power quality conditioner consist of a power-electronic series unit near the installation and several power-electronic shunt units connected close to the users end . The series and parallel units do not have a common dc link, so their control strategies are independent of each other. By using this device there is improvement in PQ, reducing the most common disturbances for all customers that are supplied by the mains (PQ) by using only the series unit power) by the shunt units. Therefore, this new technique simultaneously improve the PQ and reduce the cost for those who need high quality power.

II. UNIFIED POWER QUALITY CONDITIONER

The Unified Power Quality Conditioner is a custom power device that is employed in the distribution system to mitigate the disturbances that affect the performance of sensitive and/or critical load .It is a type of hybrid APF and is the only versatile device which can mitigate several power quality problems related with voltage and current simultaneously therefore is a multi functioning device that compensate various voltage disturbances of the power supply, to correct voltage fluctuations and to prevent harmonic load current from entering the power system. Fig. 1 shows the system configuration of a single-phase UPQC. Unified Power Quality Conditioner (UPQC) consists of two IGBT based Voltage source converters (VSC), one shunt and one series cascaded by a common DC bus. The shunt converter is connected in parallel to the load. It provides VAR support to the load and supply harmonic currents. Whenever the supply voltage undergoes sag then series converter injects suitable voltage with supply [2]. Thus UPQC improves the power quality by preventing load current harmonics and by correcting the input power factor

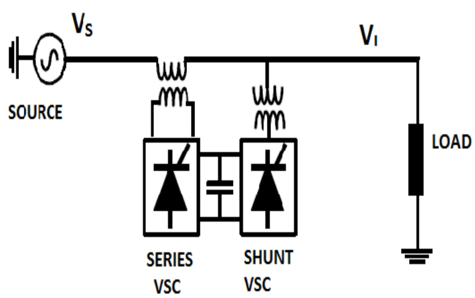


Fig.1 Block diagram of UPQC

The main components of a UPQC are series and shunt power converters, DC capacitors, low-pass and high-pass passive filters, and series and shunt transformers:

1.*Series converter:* It is a voltage-source converter connected in series with the AC line and acts as a voltage source to mitigate voltage distortions. It eliminates supply voltage flickers or imbalance from the load terminal voltage and forces the shunt branch to absorb current harmonics generated by the nonlinear load. Control of the series converter output voltage is usually performed using sinusoidal pulse-width modulation (SPWM). The gate pulses are generated by the comparison of a fundamental voltage reference signal with a high-frequency triangular waveform.

2. *Shunt converter:* It is a voltage-source converter connected in shunt with the same AC line and acts as a current source to cancel current distortions, to compensate reactive current of the load, and to improve the power factor. It also performs the DC-link voltage regulation, resulting in a significant reduction of the DC capacitor rating. The output current of the shunt converter is adjusted (e.g., using a dynamic hysteresis band) by controlling the status of semiconductor switches such that output current follows the reference signal and remains in a predetermined hysteresis band

3. *Midpoint-to-ground DC capacitor bank:* It is divided into two groups, which are connected in series. The neutrals of the secondary transformers are directly connected to the DC link midpoint. As the connection of both three-phase transformers is Y/Yo, the zero-sequence voltage appears in the primary winding of the series-connected transformer in order to compensate for the zero-sequence voltage of the supply system. No zero-sequence current flows in the primary side of both transformers. It ensures the system current to be balanced even when the voltage disturbance occurs.

4. *Low-pass filter:* It is used to attenuate high frequency components at the output of the series converter that are generated by high-frequency switching.

5. *High-pass filter:* It is installed at the output of shunt converter to absorb current switching ripples.

6. *Series and shunt transformers:* These are implemented to inject the compensation voltages and currents, and for the purpose of electrical isolation of UPQC converters. The UPQC is capable of steady-state and dynamic series and/or shunt active and reactive power compensations at fundamental and harmonic frequencies. However, the UPQC is only concerned about the quality of the load voltage and the line current at the point of its installation, and it does not improve the power quality of the entire system. Equivalent circuit for UPQC

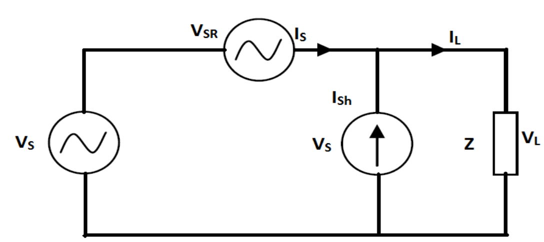


Fig. 2 Equivalent circuit for UPQC

In this circuit,
V_S represents the voltage at power supply
V_{SR} is the series-APF for voltage compensation,
V_L represents the load voltage and
I_{Sh} is the shunt-APF for current and V_{SR} compensation.

Due to the voltage Distortion, the system may contain negative phase sequence and harmonic components.

In general, the source voltage in Figure 2 can be expressed as:

$$V_s + V_{sr} = V_L \qquad (1)$$

To obtain a balance sinusoidal load voltage with fixed amplitude V, the output voltages of the series-APF should be given by;

$$V_{sr} = (V - V_{1p})\sin(\omega t + \theta_{1P}) - V_{Ln}(t) - \sum_{K=2}^{\infty} V_K(t) \qquad (2)$$

where, V_{1P}: Positive sequence voltage amplitude fundamental frequency
θ_{1P}: Initial phase of voltage for positive sequence
V_{1n} : Negative sequence component

The shunt-APF acts as a controlled current source and its output components should include harmonic, reactive and negative-sequence components in order to compensate these quantities in the load current, when the output current of shunt-APF i_{sh} is kept to be equal to the component of the load as given in the following equation:

$$i_L = I_{1p}\cos(\omega t + \theta_{1P})\sin\varphi_{1P} + i_{Ln} + \sum_{K=2}^{\infty} i_{LK} \qquad (3)$$

$$\square_{1P} = \varphi_{1P} - \theta_{1P} \qquad (4)$$

where, φ_{1P}: initial phase of current for positive sequence

As seen from the above equations that the harmonic, reactive and negative sequence current is not flowing into the power source. Therefore, the terminal source current is harmonic-free sinusoid and has the same phase angle as the phase voltage at the load terminal,

$$i_S = i_L - i_{Sh}$$
$$= I_{1p}\sin(\omega t - \theta_{1P})\cos\varphi_{1P} \qquad (5)$$

UPQC Configurations

The right shunt UPQC compensation configuration

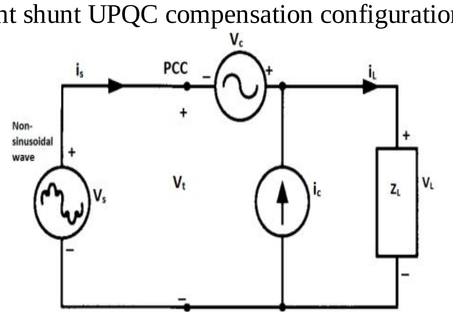


Fig. 3 The right shunt UPQC compensation configuration

The left shunt UPQC compensation configuration

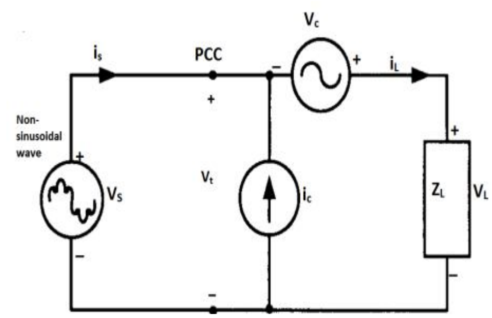


Fig. 4 The left shunt UPQC compensation configuration

There are two possible ways of connecting the unit to the terminal voltage (V_t) at PCC:

- Right-shunt UPQC (figure), where the shunt compensator (i_c) ia placed at the right side of the series compensator (V_c).
- Left –shunt UPQC (figure), where the shunt compensator (i_c) is placed at the left side of V_c.

These two structures have similar features; however the overall characteristics of the right shunt UPQC are superior (e.g. operation at zero) power injection/absorption mode, achieving unity power factor at load terminals, and full reactive power compensation).[1]

Functions performed by UPQC:

- convert the feeder (system) current is to balanced sinusoids through the shunt compensator
- convert the load voltage V_L to balanced sinusoids through the series compensator
- ensure zero real power injection (and/or absorption) by the compensators
- supply reactive power to the load (Q compensation).

III. CONTROL PHILOSOPHY

When ASD is used as load the voltage and current harmonics will be present in the system as shown in Fig.3 .For this firstly load voltage is sensed and passed through a sequence analyzer. The magnitude of load voltage is compared with reference voltage. Pulse width modulated (PWM) control technique is applied for inverter switching so as to produce a three phase 50 Hz sinusoidal voltage at the load terminals [1], [4]. Chopping frequency is kept in the range of a few KHz. PI controller is used with the IGBT inverter to maintain 1 p.u. voltage at the load terminals. PI controller input is an actuating signal which is the difference between the *Vref* and *Vin.* An advantage of a proportional plus integral controller is that its integral term causes the steady-state error additional phase-lag/lead in the three-phase voltages.

IV. PARAMETERS OF TEST SYSTEM

Simulation model of UPQC using pi control and field oriented induction motor drive as load is shown in Fig.3. System parameters of test system are listed in Table

Table-1 System Parameters

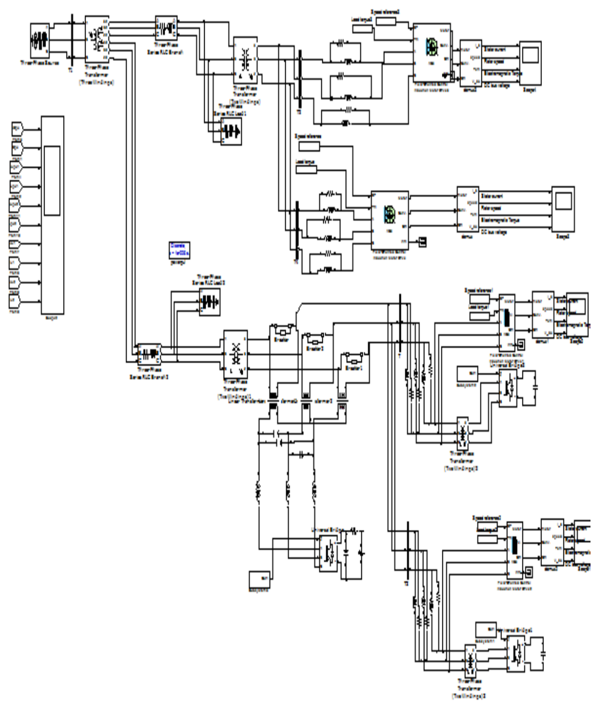


Fig. 5 Simulation model of OPEN UPQC with induction motor load

V. SIMULATION AND RESULTS

An ideal three-phase sinusoidal supply voltage is applied to the non-linear load (Field oriented control Induction motor drive) injecting current and voltage harmonics into the system. Fig. 6(a) shows load 1 voltage in three-phase before compensation. Fig. 6(b) shows THD level for uncompensated load voltage. Fig. 6(c) shows load 1 voltage in three-phase after compensation. Fig. 6(d) shows THD level for compensated load voltage. Fig. 7(a) shows load 1 current in three-phase before compensation. Fig. 7(b) shows THD level for uncompensated load current. Fig. 7(c) shows load 1 current in three-phase after compensation. Fig. 7(d) shows THD level for compensated load current. Fig. 8(a) shows load 2 voltage in three-phase before compensation. Fig. 8(b) shows THD level for uncompensated load voltage. Fig. 8(c) shows load 2 voltage in three-phase after compensation. Fig. 8(d) shows THD level for compensated load voltage. Fig. 9(a) shows load 2 current in three-phase before compensation. Fig. 9(b) shows THD level for uncompensated load cu. Fig. 9(c)shows load 2 current in three-phase after compensation. Fig. 9(d) shows THD level for compensated load current. The Total Harmonic Distortion (THD) for load 1 voltage which was 11.77% as shown in Fig.6(b) before compensation is effectively reduces to 7.31 % as shown in Fig. 6(d) after compensation using PI controller. Shunt inverter is

S. No	System Quantities	Standards
1	Source	3-phase, 13 kV, 50 Hz
2	Inverter parameters	IGBT based, 3-arm, 6-pulse, Carrier Frequency=1080 Hz , Sample Time= 5 μs
3	PI controller	Kp=0.5,Ki=100 for series control, Kp=1000,Ki=1000 for first shunt control, Kp=0.5,Ki=100 for second shunt control Sample time=50 μs
4	RL load	Active power = 1 kW Inductive Reactive Power=400 VAR
5	Motor load	Voltage Vrms=220 V Frequency 50 Hz
6	Transformer1	Y/Δ/Δ 13/115/115 kV
7	Transformer2	Δ/Y 115/11 kV

able

to reduce the harmonics entering into the system. The Total Harmonic Distortion (THD) for load 1 current which was 22.77% as shown in Fig.7(b) before compensation and effectively reduces to 7.28 % as shown in Fig. 7(d) after compensation using PI controller. The Total Harmonic Distortion (THD) for load 2 voltage which was 11.77% as shown in Fig.8(b) before compensation is effectively reduces to 7.31 % as shown in Fig.8(d) after compensation using PI controller. The Total Harmonic Distortion (THD) for load 2 current which was 9.94% as shown in Fig.9(b) before compensation and effectively reduces to 4.98 % as shown in Fig. 7(d) after compensation using PI controller. The voltage compensation is small because system consist of transformers which are already doing compensation for voltage.

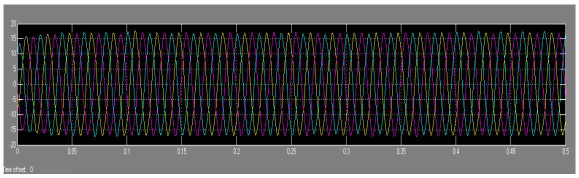


Fig. 6(a) Load 1, voltage waveform without open upqc

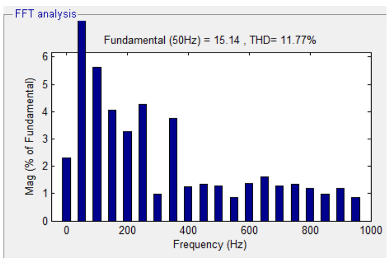


Fig. 6(b) Load 1, THD without open UPQC for voltage

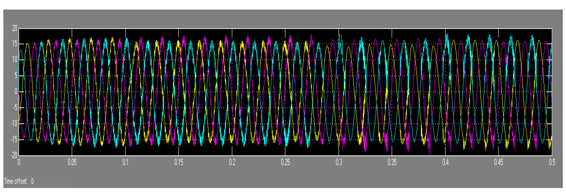


Fig. 6(c) Load 1,voltage waveform with open upqc

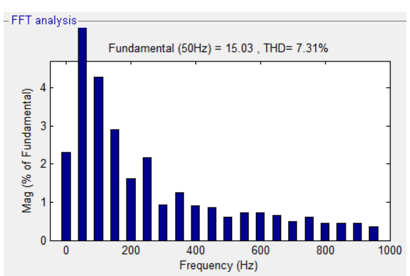


Fig. 6(d) Load 1, THD with open UPQC for voltage

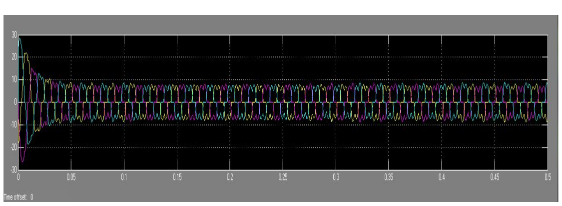


Fig. 7(a) Load 1, current waveform without open upqc

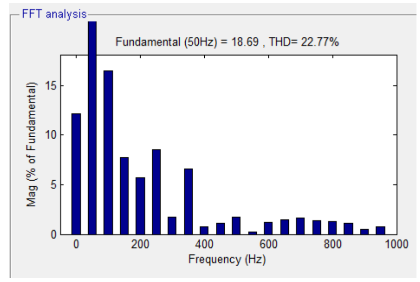


Fig. 7(b) Load 1, THD without open UPQC for current

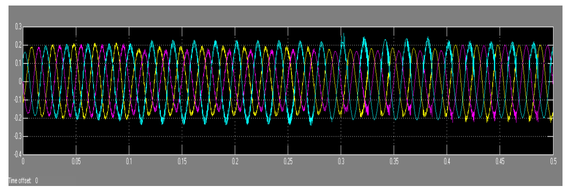


Fig. 7(c) Load 1, current waveform with open upqc

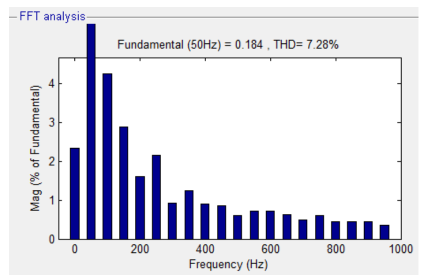


Fig. 7(d) Load 1, THD with open UPQC for current

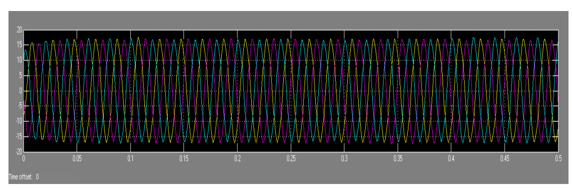


Fig. 8(a) Load 2, voltage waveform without open upqc

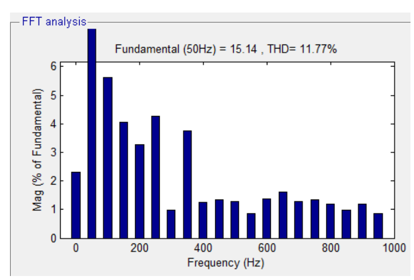


Fig. 8(b) Load 2, THD without open UPQC for voltage

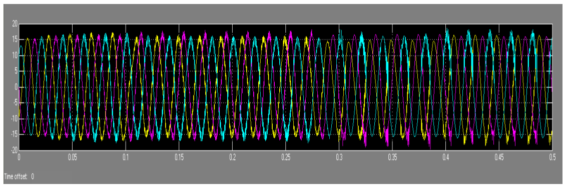


Fig. 8(c) Load 2, voltage waveform with open upqc

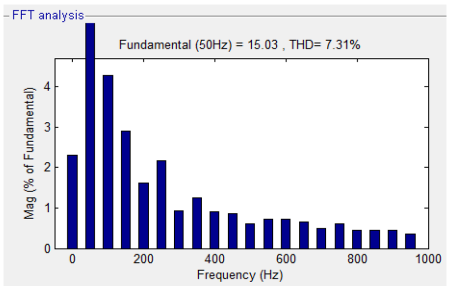


Fig. 8(d) Load 2, THD with open UPQC for voltage

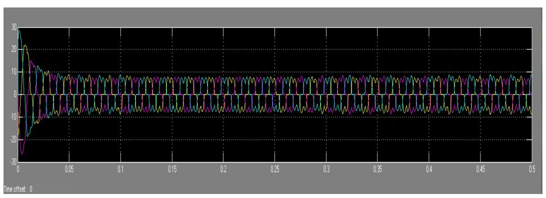


Fig. 9(a) Load 2, current waveform without open upqc

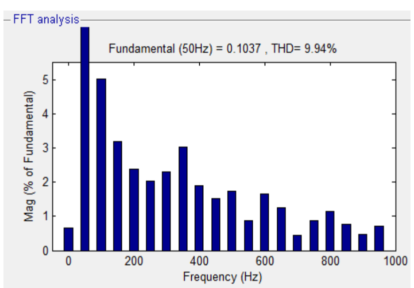


Fig. 9(b) Load 2, THD without open UPQC for current

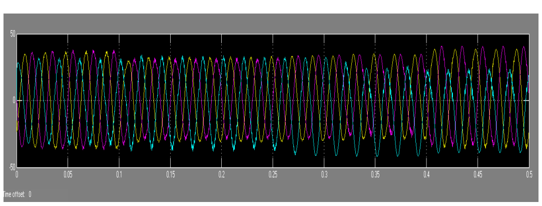


Fig. 9(c) Load 2, current waveform with open upqc

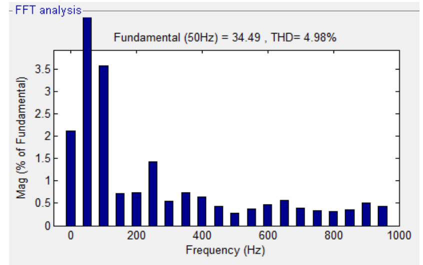


Fig. 9(d) Load 2, THD with open UPQC for current

REFERENCES

[1] Alexander Kusko and Marc T.Thompson, “Power Quality in Electrical Systems”, McGraw-Hill, 2007.

[2] Ahmed M. A. Haidar, Chellali Benachaiba, Faisal A. F. Ibrahim and Kamarul Hawari “Parameters Evaluation of Unified Power Quality Conditioner”, Conference on Electro/Information Technology (EIT) IEEE International, pp.1-6, 2011.

[3] M.Arun, Bhaskar, Dr.S.S.Dash, C.Subramani, M.Jagdeesh Kumar, P.R.Giresh and M.Varun Kumar, “Voltage Quality Improvement Using DVR”, Conference on Recent trends in Information Telecommunication and Computing (ITC), pp.378-380, 2010.

[4] M. Brenna, R. Faranda, E. Tironi, “A New Proposal for Power Quality and CustomPower Improvement: OPEN UPQC” Vol. 24, pp.2107-2116, 2009.

[5] Arindam Ghosh and Amit Kumar Jindal, “A Unified Power Quality Conditioner for Voltage Regulation of Critical Load Bus” ,conference on Power Delivery Transactions Vol. 22, pp.364-372, 2007.

[6] V. Khadkikar and A. Chandra, “A new control philosophy for a unified power quality conditioner (UPQC) to coordinate load-reactive power demand between shunt and series inverters,” Conference on Trans. Power Del., vol. 23, pp.2522–2534, 2008.

[7] M. Basu, S.P. Das and G.K. Dubey, "Performance study of UPQC-Q for load compensation and voltage sag mitigation", Conference of the Industrial Electronics Society Vol: 1, pp.698–703, 2002.

[8] Sai Shankar , Ashwani Kumar and W. Gao, “Operation of Unified Power Quality Conditioner under Different Situations” .Power and Energy Society General Meeting, 2011 IEEE, pp.1-10, 2011.

[9] A.Kazemi, A. Mokhtarpour and M.Tarafdar Haque, “A New Control Strategy for Unified Power Quality Conditioner (UPQC) In Distribution Systems”, Confrence on Power System Technology , pp.1–5, 2006.

[10] A. Mokhtatpour and H.A. Shayanfar, “Power Quality Compensation as Well as Power Flow Control Using of Unified Power Quality Conditioner” Confrence on Power and Energy Engineering Conference (APPEEC), Asia- Pacific ,pp:1-4, 2011

Mobile Computing

Meenakshi Mahajan

Department Of Electronics and Communication
DAV INSTITUTE OF ENGG AND TECHNOLOGY, Jalandhar
er_mania88@ymail.com

***Abstract*— Over the past few years a large number of advances in computing and communications technology have made it possible for computing to occur virtually anywhere. Battery powered laptops were one of the first of these, beginning in the mid 1980s, and further advances in miniaturization and battery technology have reduced the size of a full powered version. Mobile voice communication is widely established throughout the world and has had a very rapid increase in the number of subscribers to the various cellular networks over the last few years. An extension of this technology is the ability to send and receive data across these cellular networks. This is the principle of mobile computing. Mostly GPRS service is best example of mobile computing.**

General Packet Radio Services (GPRS) is a packet-based wireless communication service that promises data rates from 56 up to 114 Kbps and continuous connection to the Internet for mobile phone and computer users. The higher data rates allow users to take part in video conferences and interact with multimedia Web sites and similar applications using mobile handheld devices as well as notebook computers. GPRS is based on Global System for Mobile GSM communication and complements. General packet radio service (GPRS) is a packet oriented data oriented service on the 2G and 3G cellular communication systems.

Keywords— ***Cellular network ,GPRS ,2G,3G***

I. Introduction

The Introduction of mobility in data communications required a move from the Public Switched Data Network (PSDN) to other networks like the ones used by mobile phones. PCSI has come up with an idea called CDPD (Cellular Digital Packet Data) technology which uses the existing mobile network

Mobile technology has already creating hype throughout the world. We are using so many features in our mobile these days which most of us have never dreamed off. We are using SMS, MMS, sending pictures and video files in amazingly quick time, finding locations, accessing high speed internet in your mobiles are the features which were not possible few years back. There are some technologies which are actually responsible for such facilities on our mobiles.

There is numerous mobile communication technologies which makes growth of cellular industry to this extend one of the technology which is by far the best and most widely used in mobile communication is GSM, CDMA, GPRS, etc.

Examples of mobile IT devices include:

- Laptop and notebook computers
- Palmtop computers or personal digital assistants
- Mobile phones and 'smart phones'
- Global positioning system (GPS) devices

II. GPRS (General Packet Radio Services)

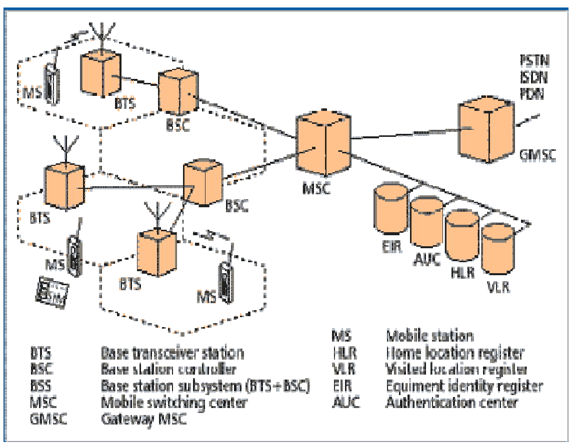
General Packet Radio Services (GPRS) is a packet-based wireless communication service that promises data rates from 56 up to 114 Kbps and continuous connection to the Internet for mobile phone and computer users. The higher data rates allow users to take part in video conferences and interact with multimedia Web sites and similar applications using mobile handheld devices as well as notebook computers. GPRS is based on Global System for Mobile GSM communication and complements General packet radio service (GPRS) is a packet oriented data oriented service on the 2G and 3G cellular communication systems GPRS is a best-effort service, implying variable throughput and latency that depend on the number of other users sharing the service concurrently, as opposed to circuit switching, where a certain quality of service is guaranteed during the connection. In 2G systems, GPRS provides data rates of 56–114 bit/second. 2G cellular technology combined with GPRS is sometimes described as 2.5G that is, a technology between the second 2G and third 3G generations of mobile telephony. It provides moderate-speed data transfer, by using unused time division multiple access (TDMA) channels in, for example, the GSM system. GPRS is integrated into GSM.

GPRS also complements Bluetooth a standard for replacing wired connections between devices with wireless radio connections. In addition to the Internet Protocol (IP), GPRS supports X.25, a packet-based protocol that is used mainly in Europe. GPRS is an evolutionary step toward Enhanced Data GSM Environment (EDGE) and Universal Mobile Telephone Service (UMTS)

III. GPRS Architecture

GPRS is a data network that overlays a second-generation GSM network. This data overlay network provides packet data transport at rates from 9.6 to 171 kbps. Additionally, multiple users can share the same air-interface resources simultaneously. GPRS attempts to reuse the existing GSM network elements as much as possible, but to effectively build a packet-based mobile cellular network, some new network elements, interfaces, and protocols for handling packet traffic are required.In addition to the entities several databases are used for the purpose of call control and network management. These databases are Home Location Register (HLR), Visitor Location Register (VLR), the Authentication centre (AUC), and Equipment Identity Register (EIR). Home Location Register (HLR) stores the permanent (such as user profile) as well as temporary (such as current location) information about all the users registered with the network. A VLR stores the data about the users who are being serviced currently. It includes the data stored in HLR for faster access as well as the temporary data like location of the user. The AUC stores the authentication information of the user such as the keys for encryption. The EIR stores data about the equipments and can be used to prevent calls from a stolen equipments.

GPRS ARCHITECTURE



A. GPRS Mobile Stations

New Mobile Station are required to use GPRS services because existing GSM phones do not handle the enhanced air interface or packet data. A variety of MS can exist, including a high-speed version of current phones to support high-speed data access, a new PDA device with an embedded GSM phone, and PC cards for laptop computers. These mobile stations are backward compatible for making voice calls using GSM.

B. GPRS Base Station Subsystem:

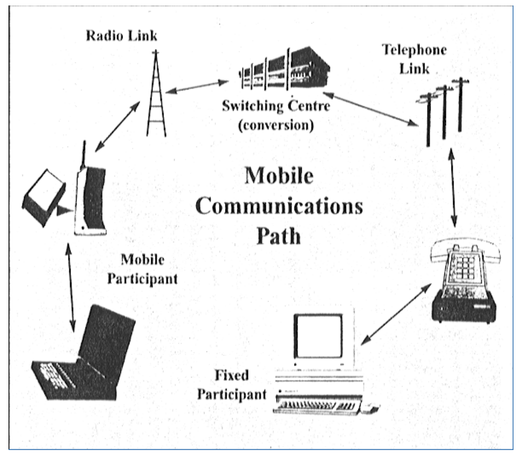
Each BSC requires the installation of one or more Packet Control Units (PCU) and a software upgrade. The PCU provides a physical and logical data interface to the base station subsystem (BSS) for packet data traffic. The BTS can also require a software upgrade but typically does not require hardware enhancements.

When either voice or data traffic is originated at the subscriber mobile, it is transported over the air interface to the BTS, and from the BTS to the BSC in the same way as a standard GSM call. However, at the output of the BSC, the traffic is separated; voice is sent to the mobile switching centre (MSC) per standard GSM, and data is sent to a new device called the SGSN via the PCU over a Frame Relay interface.

IV. DATA COMMUNICATIONS

Data Communications is the exchange of data using existing communication networks. The term data covers a wide range of applications including File Transfer (FT), interconnection between Wide-Area-Networks (WAN), facsimile (fax), electronic mail, access to the internet and the World Wide Web (WWW).

DATA COMMUNICATION IN GPRS



A.Circuit switching

Circuit switching implies that data from one user (sender) to another (receiver) has to follow a pre specified path. If a link to be used is busy the message cannot be redirected a property which causes many delays.

B.Packet switching

Packet switching is an attempt to make better utilisation of the existing network by splitting the message to be sent into packets. Each packet contains information about the sender, the receiver, the position of the packet in the message as well as part of the actual message. There are many protocols defining the way packets can be send from receiver to sender

V. Mobile Communication Technologies

There are number of mobile communication technologies which have been developed ever since from analogue days and then new technology is released to cope with telephonic industry demands. Cellular companies use AMPS, D-AMPS, CMMA2000, UMTS, GSM, EVDO etc. AMPS however pretty much vanished from the scene, AMPS network system was based on analogue communication technology, latest features were not supported by AMPS therefore all cellular networks world wide have adopted digital communication methodologies to meet the need of consumers. Latest mobile handsets offers features which one had never thought off, ultimately it forces mobile network companies to bring these features in practice use to take commercial stages. Currently we have following mobile and internet communication technologies adopted by different mobile companies in different parts of the world.

- GSM
- CDMA
- EDGE
- GPRS
- VOIP

VI. SERVICES OFFERED

If SMS over GPRS is used, an SMS transmission speed of about 30 SMS messages per minute may be achieved. This is much faster than using the ordinary SMS over GSM, whose SMS transmission speed is about 6 to 10 SMS messages per minute.

- SMS messaging and broadcasting
- "Always on" internet access
- Multimedia messaging service(MMS)
- Push to talk over cellular
- Instant messaging
- Point-to-point service: inter-networking with the Internet (IP)
- Point-to-Multipoint service

If SMS over GPRS is used, an SMS transmission speed of about 30 SMS messages per minute may be achieved. This is much faster than using the ordinary SMS over GSM, whose SMS transmission speed is about 6 to 10 SMS messages per minute.

VII. CHARACTERSTICS OF GPRS

GPRS enables a variety of new and unique services to the mobile wireless subscriber. These mobile services have unique characteristics that provide enhanced value to customers. These characteristics include the following:

Mobility: The ability to maintain constant voice and data communications while on the move

Immediacy Allows subscribers to obtain connectivity when needed, regardless of location and without a lengthy login session.

Localization Allows subscribers to obtain information relevant to their current location. The combination of these characteristics provides a wide spectrum of possible applications that can be offered to mobile subscribers. In general, applications can be separated into two high-level categories: corporate and consumer. These include:

Communications: E-mail, fax, unified messaging and intranet/Internet access etc.

Value-added services: Information services and games etc.

E-commerce: Retail, ticket purchasing, banking and financial trading etc.

Location-based applications: Navigation, traffic conditions, airline/rail schedules and location finder etc.

Vertical applications: Freight delivery, fleet management and sales-force automation.

Advertising: Advertising may be location sensitive. For example, a user entering a mall can receive advertisements specific to the stores in that mall.

VIII. APPLICATIONS OF GPRS

Still Images: Still images such as photographs, pictures, postcards, greeting cards and presentations, static web pages can be sent and received over the mobile net

Web Browsing: Using Circuit Switched Data for web browsing has never been an enduring application for mobile users. Because of the slow speed of Circuit Switched Data, it takes a long time for data to arrive from the Internet server to the browser. Alternatively, users switch off the images and just access the text on the web, and end up with difficult to read text layouts on screens that are difficult to read from. As such, mobile Internet browsing is better suited to GPRS.

File Transfer: As this generic term suggests, file transfer applications encompass any form of downloading sizeable data across the mobile network. This data could be a presentation document for a travelling salesperson, an appliance manual for a service engineer or a software application such as Adobe Acrobat Reader to read documents.

IX. LIMITATIONS OF GPRS

Limited Cell Capacity for All Users: GPRS does impact a network's existing cell capacity. There are only limited radio resources that can be deployed for different uses - use for one purpose precludes simultaneous use for another. For example, voice and GPRS calls both use the same network resources. The extent of the impact depends upon the number of timeslots, if any, that are reserved for exclusive use of GPRS. However, GPRS does dynamically manage channel allocation and allow a reduction in peak time signalling channel loading by sending short messages over GPRS channels instead.

No Store and Forward: whereas the Store and Forward Engine in the Short Message Service is the heart of the SMS Centre and key feature of the SMS service, there is any storage mechanism incorporated into the GPRS standard, apart from the incorporation of interconnection links between SMS and GPRS.

Speeds Much Lower in Reality: It is unlikely that a network operator will allow all timeslots to be used by a single GPRS user. Additionally, the initial GPRS terminals are expected be severely limited - supporting only one, two or three timeslots. The bandwidth available to a GPRS user will therefore be severely limited. As such, the theoretical maximum GPRS speeds should be checked against the reality of constraints in the networks and terminals. The reality is that mobile networks are always likely to have lower data transmission speeds than fixed networks.

CONCLUSION

The technology of the Global Positioning System is allowing for huge changes in society. The applications using GPS are constantly growing. The cost of the receivers is dropping while at the same time the accuracy of the system is improving. This affects everyone with things such as faster Internet speed and safer plane landings.

SMS, because of its very nature has unique advantages that other non voice services do not have. It provides a very convenient method of exchanging small bits of information between mobile users. The reasons for the enormous popularity of SMS have been the fact that this mechanism of sending and receiving messages not only saves time but costs less as well. In many situations one is relatively much more comfortable sending a message via SMS than talking over phone. With new information services and unique value added services being used by the operators the popularity of SMS is increasing further. SMS is also uniquely positioned as a very attractive advertisement medium. SMS should no longer be treated as a value added service in mobile networks. SMS is not only providing a useful mechanism for a host of innovative services over mobile networks but it acting as a point of entry for new data services like WAP in mobile networks. Even though the system was originally developed for military purposes, civil sales now exceed military sales.

REFERENCES

[1] S. M. Metev and V. P. Veiko, *Laser Assisted Microtechnology*, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.

[2] J. Breckling, Ed., *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

[3] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, “A novel ultrathin elevated channel low-temperature poly-Si TFT,” *IEEE Electron Device Lett.*, vol. 20, pp. 569–571, Nov. 1999.

[4] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, “High resolution fiber distributed measurements with coherent OFDR,” in *Proc. ECOC’00*, 2000, paper 11.3.4, p. 109.

Techniques to Discover Black Hole nodes in Mobile AdHoc Networks using Efficient AODV Protocol

Ms.Pooja vij , Dr. V. K. Banga
Amritsar College of Engineering & Technology, Amritsar

Abstract-A wireless Adhoc network is a collection of mobile nodes with no pre-established infrastructure, forming a temporary network. In the absence of a fixed infrastructure, nodes have to cooperate in order to provide the necessary network functionality. One of the principal routing protocols used in Ad hoc networks is AODV (Ad hoc On-Demand Distance Vector) protocol. The security of the AODV protocol is compromised by a particular type of attack called 'Black Hole' attack [1]. In this attack a malicious node advertises itself as having the shortest path to the node whose packets it wants to intercept. It is proposed to wait and check the replies from all the neighboring nodes to find the Black hole nodes. In this paper, we detect the Black hole nodes or malicious nodes and after detecting it we will remove those nodes and also find the shortest path from source to destination by using ns2 simulator. We propose that our protocol is increase the throughput, security and life time of the network by reducing the delay than the other conventional AODV protocols.

Keywords—MANETS, black hole attack, malicious node, routing protocols

INTRODUCTION

An ad hoc network is a collection of nodes that do not rely on a predefined infrastructure to keep the network connected. So the functioning of Ad-hoc networks is dependent on the trust and co-operation between nodes. Nodes help each other in conveying Information about the topology of the network and share the responsibility of managing the network. Hence in addition to acting as hosts, each mobile node. *In Infrastructure less or Ad Hoc wireless network*, the mobile node can move while communicating, there are no fixed base stations and all the nodes in the network act as routers. This type of network can be shown as in fig. 1. Most important networking operations include routing and network management [2].

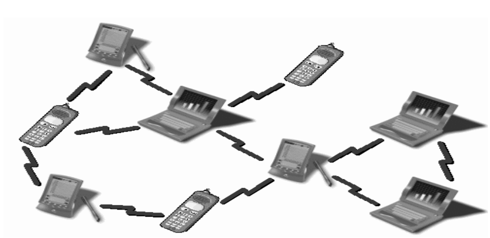


Fig 1: Mobile Ad Hoc Networks

Types of Black Holes

A Black Hole attack [1],[4] is a kind of denial of service attack where a malicious node can attract all packets by falsely claiming a fresh route to the destination and then absorb them without forwarding them to the destination.

Single Black Hole Attack

In single black hole attack only one malicious node attack on the route.

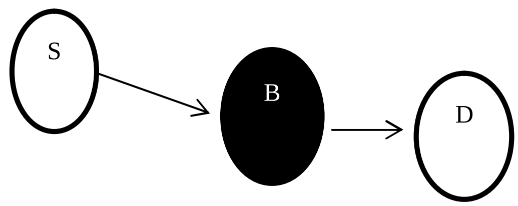


Fig 2: Single Black hole attack

Co-operative Black Hole Attack

Co-operative Black Hole means the malicious nodes act in a group.

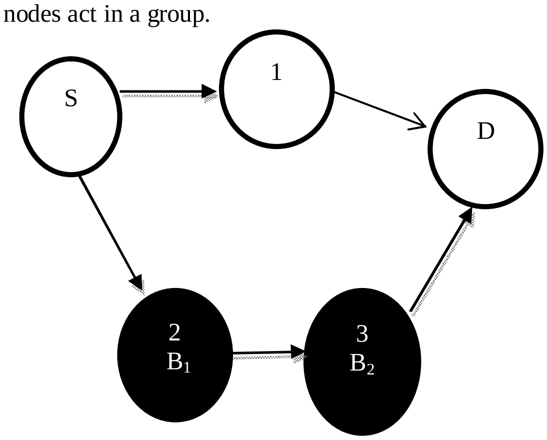


Fig 3: Co-operative Black hole node

Routing protocols can be divided into proactive, reactive and hybrid protocols, depending on the routing topology.

Classification of Routing Protocols

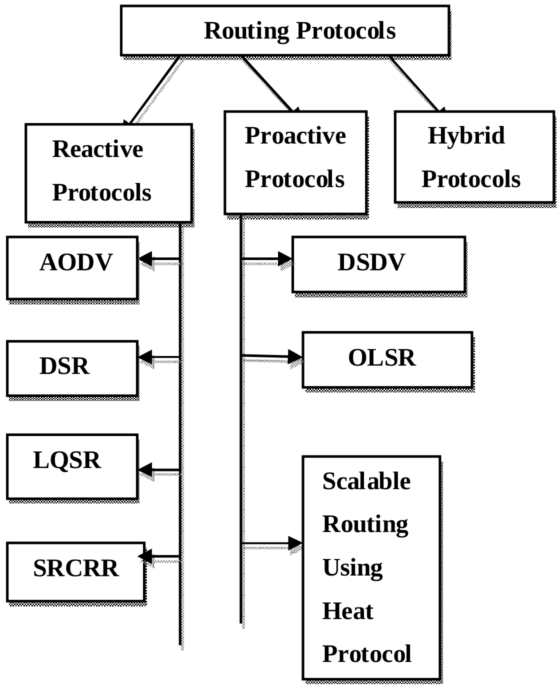


Fig: 4 Classification of Routing protocols

There exists a large number of wireless mesh network routing protocols. They can be broadly classified into three categories as shown in Figure 4. In this study, we focus on two types of protocols: Proactive and Reactive Routing Protocols.

1. Proactive Protocols

Proactive protocols are typically table-driven. Examples of this type include DSDV, WRP. In these types of routing protocols, each node maintains a table of routes to all destination nodes in the network at all times[3]. This requires periodic exchange of control messages between nodes. Since the route to every destination already exists, there is little or no initial delay when first sending data. However, periodic control traffic competes with data transfer to gain access to the channel. The Proactive protocols are classified into Destination Sequence Distance Vector, Optimized Link State Routing, Scalable routing using heat protocol

a) Destination Sequence Distance Vector (DSDV)

DSDV is a proactive type of routing protocol. DSDV table-driven DV routing scheme for MANET, DSDV based on the Bellman-Ford algorithm with adaptations that are specifically targeted for mobile networks. The Bellman-Ford algorithm uses the distance vector approach, where every node maintains a routing table that records the ―next hop for every reachable destination along the shortest route and the minimum distance (number of hops). Whenever there is any change in this minimum distance, the information is reported to neighboring nodes and the tables are updated as required [9] To make this algorithm adequate for mobile ad hoc networks, DSDV added a sequence number with each distance entry to indicate the freshness of that entry. A sequence number is originated at the destination node and is incremented by each node that sends an update to its neighbours. Thus, a newer routing table update for the same destination will have a higher sequence number. Routing table updates are periodically transmitted throughout the network, with each node updating its routing table entries based on the latest sequence number corresponding to that entry. If two updates for the same destination have identical sequence numbers but different distances, then the shorter distance is recorded. The addition of sequence numbers removes the possibility of long-lived loops and also th“counting-to-infinity” problem, where it takes a large number of update message to ascertain that a node is not reachable [9].

b) Optimized Link State Routing(OLSR)

OLSR protocol is a proactive routing protocol. The Optimized Link State Routing (OLSR) protocol was first introduced in [10]. The current OLSR Version 11 is the definitive RFC 3626. It provides optimization of a pure link state algorithm tailored to the requirements of a mobile wireless LAN (OLSR protocol optimized for MANET but can also be used in WMNs). The concept used in the protocol is that of multipoint relays (MPRs). MPRs are selected nodes which forward broadcast messages during the flooding process. This technique provides two key optimizations [10]. First, it reduces the size of the control packets, that is, instead of all links, it declares only a subset of neighbouring links designated as the MPRs. Secondly, flooding of the control traffic is minimized by using only the selected nodes to propagate its messages in the network. Only the MPRs of a node retransmit its broadcast messages. Such procedures substantially reduce the message overhead as compared to pure flooding mechanisms where every node re-transmits each message when it receives the first copy of the packet.

2. Reactive Protocols

Reactive or source-initiated on-demand protocols, in contrary, do not periodically update the routing information. It is propagated to the nodes only when necessary. Example of this type includes DSR, AODV and ABR. In reactive routing protocols, the route is calculated only when a node needs to send data to an unknown destination. Thus, route discovery is initiated only when needed. This saves overhead in maintaining unused routes. However, this may lead to larger initial delays. During route discovery, the query is flooded into the entire network and the reply from the destination (or intermediate nodes) sets up the path between the source and destination. The Reactive protocols are classified into Ad-hoc on Demand Distance Vector, Dynamic Source Routing, SRCRR, Link Quality Source Routing, and Multi radio Link Quality Source Routing.

a) Dynamic Source Routing Protocol

Dynamic Source Routing (DSR) is a routing protocol for wireless mesh networks. It is similar to AODV in that it establishes a route on-demand when a transmitting mobile node requests one. However, it uses source routing instead of relying on the routing table at each intermediate device. Dynamic source routing protocol (DSR) is an on-demand, source routing protocol, whereby all the routing information is maintained (continually updated) at mobile nodes. DSR allows the network to be completely self-organizing and self-configuring, without the need for any existing network infrastructure or administration[8]. The protocol is composed of the two main mechanisms of "Route Discovery “which work together to allow nodes to discover and maintain routes to arbitrary destinations in the ad hoc network. An optimum path for a communication between a source node and target node is determined by Route Discovery process. Route Maintenance ensures that the communication path remains optimum and loop-free according the change in network conditions, even if this requires altering the route during a transmission. Route Reply would only be generated if the message has reached the projected destination node (route record which is firstly contained in Route Request would be inserted into the Route Reply).To return the Route Reply, the destination node must have a route to the source node. If the route is in the route cache of target node,the route would be used. Otherwise, the node will reverse the route based on the route record in the Route Reply message header (symmetric links). In the event of fatal transmission, the Route Maintenance Phase is initiated whereby the Route Error packets are generated at a node. The incorrect hop will be detached from the node's route cache; all routes containing the hop are reduced at that point. Again, the Route Discovery Phase is initiated to determine the most viable route.The major dissimilarity between this and the other on-demand routing protocols is that it is beacon-less and hence it does not have need of periodic hello packet (beacon) transmissions, which are used by a node to inform its neighbors of its presence. The fundamental approach of this protocol during the route creation phase is to launch a route by flooding Route Request packets in the network. The destination node, on getting a Route Request packet, responds by transferring a Route Reply packet back to the source, which carries the route traversed by the Route Request packet received.

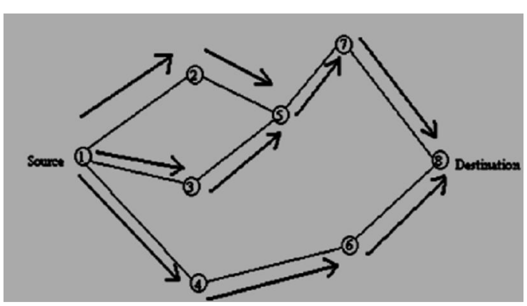


Fig 5: Propagation of Request (PREQ) packet

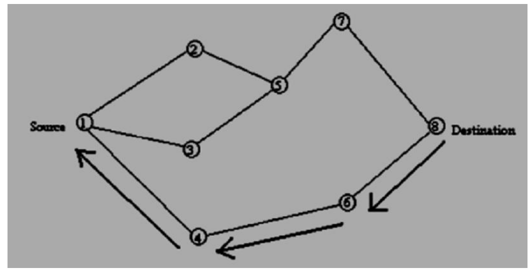


Fig 6: Creation of Route in DSR

A destination node, after receiving the first Route Request packet, replies to the source node through the reverse path the Route Request packet had traversed. Nodes can also be trained about the neighboring routes traversed by data packets if operated in the promiscuous mode. This route cache is also used during the route construction phase. If an intermediary node receiving a Route Request has a route to the destination node in its route cache, then it replies to the source node by sending a Route Reply with the entire route information from the source node to the destination node.

b) Ad Hoc On-Demand Distance routing Protocol (ADOV)

The Ad Hoc On-Demand Distance Vector (AODV) routing protocol is an adaptation of the DSDV protocol for dynamic link conditions [5][6][7]. Every node in an Ad-hoc network maintains a routing table, which contains information about the route to a particular destination. Whenever a packet is to be sent by a node, it first checks with its routing table to determine whether a route to the destination is already available. If so, it uses that route to send the packets to the destination. If a route is not available or the previously entered route is inactivated, then the node initiates a route discovery process. ARREQ (Route REQuest) packet is broadcasted by the node. Every node that receives the RREQ packet first checks if it is the destination for that packet and if so, it sends back an RREP (Route Reply) packet. If it is not the destination, then it checks with its routing table to determine if it has got a route to the destination. If not, it relays the RREQ packet by broadcasting it to its neighbors. If its routing table does contain an entry to the destination, then the next step is the comparison of the ‘Destination Sequence’ number in its routing table to that present in the RREQ packet. This Destination Sequence number is the sequence number of the last sent packet from the destination to the source. If the destination sequence number present in the routing table is lesser than or equal to the one contained in the RREQ packet, then the node relays the request further to its neighbors. If the number in the routing table is higher than the number in the packet, it denotes that the route is a ‘fresh route’ and packets can be sent through this route. This intermediate node then sends a RREP packet to the node through which it received the RREQ packet. The RREP packet gets relayed back to the source through the reverse route. The source node then updates its routing table and sends its packet through this route. During the operation, if any node identifies a link failure it sends a RERR (Route ERRor) packet to all other nodes that uses this link for their communication to other nodes. In the following illustrated figure 7, imagine a malicious node 'M'. When node 'A' broadcasts a RREQ packet, nodes 'B' 'D'and'M' receive it. Node'M', being a malicious node, does not check up with its routing table for the requested route to node 'E'.Hence, it immediately sends back a RREP packet, claiming a route to the destination. Node 'A' receives the RREP from 'M' ahead of the RREP from 'B' and 'D'. Since AODV has no security mechanisms, malicious nodes can perform many attacks just by not behaving according to the AODV rules. A malicious node *M* can carry out many attacks against AODV.This Solutions provides routing security to the AODV routing protocol by eliminating the threat of 'Black Hole' attacks.

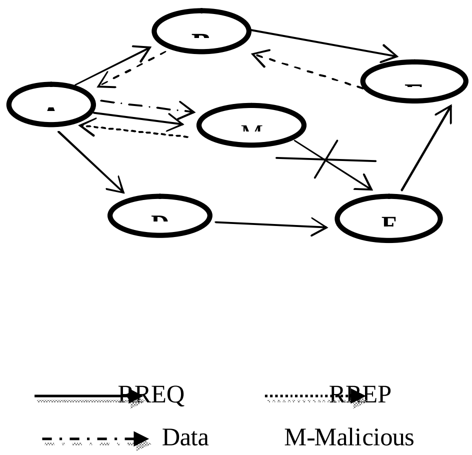


Fig 7: Black hole Attack in ADOV

A Comparison of the Characterstics of the Adhoc routing protocols is given in Table.1

Table:1 Adhoc Network Protocol Comparisions

Protocol Property	*DSR*	*DSDV*	*AODV*
Loop free	Yes	Yes	Yes
Multicaste	Yes	No	No
Distributed	Yes	Yes	Yes
Unidirectional Link Support	Yes	No	No
Multicaste	No	No	Yes
Periodic Broadcaste	No	Yes	Yes
QoS Support	No	No	No
Route Maintained in	Route Cache	Route Table	Route Table
Reactive	yes	No	Yes

3. Solution: Efficient AODV

We propose a solution that is enhancement of basic AODV and others proposed routing protocols. Firstly the proposed protocol will detect the black hole attack and after detection it will make the authorized node.Accroding to the proposed solution the source node will send RREQ to the respective nodes or relaying nodes. Every interme-diate node receives the RREQ from the source node. This intermediate node then sends a RREP packet to the node through which it received the RREQ packet. The RREP packet gets relayed back to the source through the reverse route. Then source node check the route reply message. After this, the source node checks the entire RREP packet. If the RREP packet contain both the Sequence no and Broadcast ID. Then source node transmit the data to this node. If some respective nodes sends only sequence no then it means these are unauthorized nodes or black hole nodes. The source node will request again to send the Broadcast ID.If the intermediate node will send Broadcast ID then it will consider it authorized node and transmit the data through this route. If it will not send the Broadcast ID after again requesting the source node then source node will eradicate it.

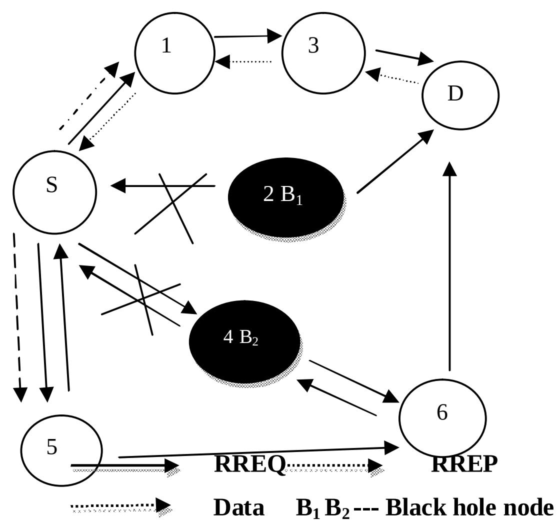


Fig: 8 Propagation of the signal from source to destination node

CONCLUSION AND FUTURE SCOPE

As already mentioned in the previous papers, the solutions has been proposed by using various techniques to attempt to detect the single black hole and co-operative black hole nodes. After detecting these black hole nodes, the data packets had not being send through this route(i.e. to avoid black hole nodes).To reduce the probability it was proposed to wait and check the replies from all the neighboring nodes to find a safe route. We propose a solution that is enhancement of all the proposed solutions. We will implement the efficient AODV routing protocol in order to detect the black hole nodes and after detecting it we will remove those nodes and also find the shortest path from source to destination. By doing this we will be able to improve the throughput, end to end delay and packet delivery ration of the network thus it increases the lifetime of the network structure.

REFERENCES

[1] Hongmei Deng,Wei Li,and Dharma P.Agrawal, “Routing Security in Wireless Adhoc Networks”, University of Cincinnati,IEEE Communications magazins, October 2002.

[2] V.Karpijoki, “Security in Ad Hoc Networks”, Seminar on Net Work Security, HUT TML 2000.

[3] Lidong zhou Zygmunt J. Haas. “Securing Ad Hoc Networks”,
IEEE networks, special issue on network security, Vol.13,no.6,November/December 1999.

[4] Yi-Chun,Adrian Perrig, “A Survey of Secure Wireless Ad Hoc Routing”,IEEE Security and Privacy,1540-7993/04/$20.00
2004 E.May/June 2004.

[5] C.E. Perkins, S.R Das, and E.Royer, “Ad-Hoc on Demand Distance Vector (AODV)”,March 2000, http:/www.ietf.org
internet-drafts/draft-ietf-manet-05.txt

[6]Yi-Chun,Adrian Perrig,David B.johnson, “Ariadna: A secure On-Demand Routing Protocol for Ad Hoc Networks”, sparrow.ece.cmu.edu/~adrian/
Projects/secure-routing/ariadne.pdf,2002.

[7] Charles E.Perkins, Elizabeth M.Belding-Royer,Samir R.Das, Mobile Ad Hoc Networking Working Group, Internet Drafts, February 2003.

[8] Kimaya Sanzgiri,Daniel LaFlamme,Bridget Dahill, Brian Neil Levine, Clay Shields,and Elizabeth M.Belding-Royer,Autenticated Routing for Adhoc Networks,” IEEE jorunal on Selected Areas in Communications,Vol.23 No.3,pp.598-610,2005.

[9]http:/web2.blogtells.com/2008/09/17/
Proactive-routing-protocols-destination-sequenced-distance/

[10]P.jacquet,P.Muhlethaler,T.Clausen, A. Laouiti, A.Qayyum and L. Viennot, “OLSR Protocol for Ad Hoc Networks”, IEEE INMIC, pp.62-68, Dec 2001, last accessed on Oct 2005.

A Survey of Soft Computing Techniques in Engineering Design Optimization

Namarta Kad[1], Dr. Yadwinder Singh Brar[2]
[1]Department of Electrical & Electronics Engineering
Amritsar College of Engineering and Technology, Amritsar
[2]Department of Electrical Engineering
Guru Nanak Dev Engineering College, Ludhiana

Abstract - **Optimization generally means finding an alternative with the most cost effective or highest achievable performance under the given constraints, by maximizing desired factors and minimizing undesired ones by systematically choosing input values from within an allowed set and computing the value of the function. Many real-world optimization problems involve multiple objectives. Improvement of one objective may lead to deterioration of another. Thus, a single solution, which can optimize all objectives simultaneously, does not exist. Instead, the best trade-off solutions, called the Pareto optimal solutions, are important to a decision maker (DM). Optimization problems are often multi-modal that is they possess multiple good solutions.**

***Keywords*: Optimization, Optimization Techniques, GA, PSO, DE, BBO, ACO, MA, HS, TS, FFA, TLBO**

I. INTRODUCTION

In the last few decades, more and more researches suggest that nature is a great source for inspirations to both develop intelligent systems and provide solutions to complicated problems. Taking animals for example, evolutionary pressure forces them to develop highly optimized organs and skills to take advantages of fighting for food, territories and mates. Some of the organs and skills can be well refined as optimization algorithms, and the evolution is a process to fine-tune the parameter settings in the algorithms. In this paper, some nature-inspired algorithms are systematically reviewed; including Ant Colony Optimization (ACO) [19]-[20], Genetic Algorithm (GA) [6], Firefly Algorithm (FA) [26], Particle Swarm Optimization (PSO) [8]-[12], Biogeography Based Optimization (BBO) [16]-[18], etc. In the previous research different soft computing techniques for the optimization problem have been surveyed under the literature [1]-[5].

This paper focuses on recent developments in the various soft computing techniques applicable on the Multi-Objective Optimization Problems. Our major concern is to study the basic concept of these techniques and their application in the problems of various engineering designs. The remainder of this paper is organized as follows. Section II reviews the various nature inspired soft computing techniques. Algorithm frameworks, selection strategies, and offspring reproduction operators are surveyed in this section and finally, the paper is concluded in Section III with the comparison of different techniques and some potential directions for future research.

II. OPTIMIZATION TECHNIQUES

As soft computing is the state-of-the-art approach to artificial intelligence and its role in effect is to model the human mind. In this respect, the soft computing techniques differ from the respective conventional computing techniques in that they are tolerant of imprecision, uncertainty, partial truth, and approximation.

A. Genetic Algorithms

Genetic algorithm is a stochastic optimization technique which is based on the principle of natural selection and genetics [6].

GA starts with a population of candidate solutions chosen randomly within the feasible range, encoded in a binary string that forms chromosomes i.e, to initialize the population that normally is composed of randomly generated individuals covering the entire range of possible solutions, and size of the population is determined by the nature of the problem itself. Each member of the population is then decoded to pass through an evaluation process. The evaluation is based on a fitness function that basically depends on the objective function of the optimization problem. The initial population undergoes three main genetic operations, selection, crossover and mutation. Selection is an operation to choose parent solutions. Crossover operation is applied with a certain probability which combines two parent chromosomes to form two new offspring chromosomes having characteristics from both parents. After a new population has been generated by selection and crossover operations, mutation is applied with small probability. Mutation is used to introduce new information to the population which does not exist in the parents. This process continues until the convergence criterion is met.

GA has been widely adopted to solve complex problems, especially in the areas of scheduling, global optimization and control engineering. The implementation of genetic algorithms is one of the easiest compared to alternative intelligent optimization algorithms, but the building block hypothesis has been criticized that it lacks of theoretical justification. Moreover, in some cases, the termination criterion is not clear. If the fitness function has only 0/1 as results, GA normally cannot solve the problem.

B. Particle Swarm Optimization

PSO was inspired by the movement pattern in a bird flock or fish school, which was first introduced by Dr. Russell C. Eberhart and Dr.James Kennedy in 1995 [8]. PSO has two primary operators: Velocity update and Position update. During each generation each particle is accelerated toward the particles previous best position (Pbest) and the global best position (Gbest). At each iteration a new velocity value for each particle is calculated based on its current velocity, the distance from its Pbest and the distance from the Gbest. The new velocity value is then used to calculate the next position of the particle in the search space. This process is then iterated a set number of times or until a minimum error is achieved [9].

Let x and v denote a particle coordinates (position) and its corresponding flight speed (velocity) in the search space. Therefore, the ***i***-th particle is represented as $x_i = [x_{i1}, x_{i2}, \ldots, x_{id}, \ldots, x_{iM}]$ in the ***M***-dimensional space. Each particle keeps track of its coordinates in the solution space which are associated with the best solution it has achieved so far. This fitness value is called *Pbest*. The best previous position of the ***i-th*** particle is recorded and represented as $Pbest_i = [Pbest_{i1}, Pbest_{i2}, \ldots, Pbest_{id}, \ldots, Pbest_{iM}]$. Another "best" value that is tracked by the particle swarm optimizer is the best value obtained so far by any particle in the neighbors of the particle. When a particle takes all the population as its topological neighbours, the best value is a global best and is called *Gbest*. The index of the best particle among all the particles in the group is represented by the $Gbest_{d.}$ The rate of the velocity for particle i is represented as $v_i = (v_{i1}, v_{i2}, \ldots, v_{id}, \ldots, v_{iM})$. The modified velocity and position of each particle can be calculated using the current velocity and the distance from $Pbest_{id}$ to $Gbest_d$ as shown in the following formulas [10]:

$$\upsilon(t+1)_{id} = w * \upsilon(t)_{id} + c_1 * rand() * (Pbest_{id} - x(t)_{id}) + c_2 * Rand() * (Gbest_d - x(t)_{id}),$$

$$x(t+1)_{id} = x(t)_{id} + \chi * \upsilon(t+1)_{id}, \qquad i = 1, 2, \ldots, N, \quad d = 1, 2, \ldots, M.$$

where N is the number of particles in a group, M is the number of members in a particle, t is the pointer of generations, χ belongs to [0,1] is the constriction factor which controls the velocity magnitude, w is the inertia weight factor, $c1$ and $c2$ are acceleration constants, rand() and Rand() are uniform random values in a range [0,1], $\upsilon(t)_{id}$ is the velocity of particle i at generation t, and $x(t)_{id}$ is the current position of particle i at generation t. The particle swarm optimization concept consists of, at each time step, changing the velocity of (accelerating) each particle toward its *Pbest* and *Gbest* locations. Acceleration is weighted by a random term, with separate random numbers being generated for acceleration toward *Pbest* and *Gbest* locations [11] .

It usually results in faster convergence rates than the genetic algorithm. In recent years, PSO has been successfully implemented to different power system optimization problems including the economic power dispatch problem with impressive success [12].

C. Differential Evolution

DE is a population-based stochastic search technique that works in the general framework of Evolutionary Algorithms and was invented by Price and Storn in 1995 [13]. DE improves population of candidate solutions over several generations using three operations in order to reach an optimal solution. These operations are mutation, crossover and selection operators. DE has the better ability to search for the optimal solution in a faster manner using its different crossover strategies i.e. it has better exploration ability. The design principles of DE are simple, efficient and used real coding. It starts to explore the search space by randomly choosing the initial candidate solutions within the boundary. Then the algorithm tries to locate the global optimum solution for the problem by iterated refining of the population through reproduction and selection.

A brief description of different steps of DE algorithm is given below [14]:

Initialization

The population is initialized by randomly generating individuals within the boundary constraints,

Mutation

As a step of generating offspring, the operations of 'mutation' are applied. 'Mutation' occupies quite an important role in the reproduction cycle. The mutation operation creates mutant vectors by perturbing a randomly selected vector with the difference of two other randomly selected vectors. Scaling factor used to control the amount of perturbation in the mutation process and improve convergence.

Crossover

Crossover represents a typical case of a 'genes' exchange. The parent vector is mixed with the mutated vector to create a trial vector. Crossover constant controls the diversity of the population and aids the algorithm to escape from local optima.

Selection

Selection procedure is used among the set of trial vector and the updated target vector to choose the best. Each solution in the population has the same chance of being selected as parents. Selection is realized by comparing the objective function values of target vector and trial vector. For minimization problem, if the trial vector has better value of the objective function, then it replaces the updated one [15] .

D. Biogeography Based Optimization

Biogeography deals with the geographical distribution of biological organisms [16]. Mathematical models of biogeography describe how species migrate from one habitat to another, how species arise, and how species become extinct. BBO has features in common with other biology-based optimization methods, like GAs and PSO. Geographical areas that are well suited as residences for biological species are said to have a high habitat suitability index (HSI). The variables that characterize habitability are called suitability index variables (SIVs). SIVs can be considered the independent variables of the habitat, and HSI can be considered the dependent variable. Habitats with a high HIS tend to have a large number of species, while those with a low HSI have a small number of species. Habitats with a high HSI have many species that emigrate to nearby habitats, simply by virtue of the large number of species that they host. Habitats with a high HIS have low species immigration rate because they are already nearly saturated with species. Therefore, high HSI habitats are more static in their species distribution than low HSI habitats. Biogeography is nature's way of distributing species, and is analogous to general problem solutions. Suppose that we are presented with a problem and some candidate solutions. The problem can be in any area of life (engineering, economics, medicine, business, urban planning, sports, etc.), as long as we have a quantifiable measure of the suitability of a given solution. A good solution is analogous to an island with a high HSI, and a poor solution represents an island with a low HIS [17].

BBO algorithm

The BBO algorithm can be briefly described with the following algorithm [18]:

(1) Initialize the BBO parameters like habitat modification probability *Pmod*, mutation probability, maximum mutation rate *mmax*, maximum immigration rate *I*, maximum emigration rate *E*, step size for numerical integration *dt*, etc. Also set maximum generation, maximum species count S maximum and an elitism parameter. Generate the SIVs of the given problem within their feasible region. This means deriving a method of mapping problem solutions to SIVs. A complete solution consisting of SIVs is known as habitat H. All these are problem dependent.

(2) Initialize several numbers of habitats depending upon the population size. Each habitat represents a potential solution to the given problem.

(3) Calculate the HSI for each habitat of the population set for given emigration rate l, immigration rate k and species S.

(4) Based on the HSI value elite habitats are identified.

(5) Probabilistically immigration rate and emigration rate are used to modify each non-elite habitat using migration operation. The probability that a habitat H_i is modified is proportional to its immigration rate k_i and the probability that the source of the modification comes from a habitat H_j is proportional to the emigration rate l_j. After modification of each non-elite habitat using migration operation each HSI is recomputed.

(6) For each habitat, the species count probability is updated. Then, a mutation operation is performed on each non-elite habitat and computes each HSI value again.

(7) Go to step (3) for the next generation.

This loop can be terminated after a predefined number of generations or after an acceptable problem solution has been found. After each habitat is modified (steps 2, 5 and 6), its feasibility as a problem solution should be verified. If it does not represent a feasible solution, then some method needs to be implemented in order to map it to the set of feasible solutions.

E. Ant Colony Optimization

Ant colony optimization (ACO) was introduced by Dorigo in 1992. ACO takes inspiration from the behavior of real ant colonies in finding the shortest path and is used to solve optimization problems. Real ants walk randomly until they find food and return to their nest while depositing pheromone on the ground in order to mark their preferred path to attract other ants to follow [19]. If other ants travel along the path and find food, they will deposit more pheromone as to reinforce the path for more ants to follow. In the case of multiobjective optimization, ACO has been applied to traveling salesman problems, vehicle routing problems, flow-shop scheduling problems, portfolio selection and others.

The behavior displayed by real ants to generate a near optimal trail can be explained in four steps: (a) At an initial moment, the environment is ‘‘clean’’, and both ants can choose any of the paths with the same probability; (b) Both ants choose a path, and in this case, one ant chooses a shorter path and the other a longer one. When they move, they deposit chemical substances called pheromones. (c) When the cycle repeats, the shorter path will have a stronger pheromone trail more quickly (since the corresponding ant arrives sooner). To choose the next move, an ant uses a probability function weighted according to the amount of pheromones on the trail. Thus, more ants will choose the shorter trail. (d) After a certain time, the first pheromones that were dropped evaporate, and the pheromone trail on the shorter path becomes dominant. In this case, all the ants will choose the shorter trail. The two main operations are: the pheromone level adjustment (also known as pheromone deposit and pheromone evaporation rules) and a probabilistic rule that chooses a destination based on the pheromone level (the state transition rule) [20].

F. Memetic Algorithms

Memetic algorithms are population-based metaheuristics composed of an evolutionary framework and a set of local search algorithms which are activated within the generation cycle of the external framework [21].

According to the philosophical theory of Richard Dawkins, human culture can be decomposed into simple units namely memes. Thus a meme is a ‘‘brick’’ of the knowledge that can be duplicated in human brains, modified, and combined with other memes in order to generate a new meme. Within a human community, some memes are simply not interesting and then will die away in a short period of time. Some other memes are somewhat strong and then, similar to an infection, will propagate within the entire community. The memes can also undergo slight modifications or combine with each other thus generating new memes which have stronger features and are more durable and prone to propagation. This interpretation of human culture inspired Moscato and Norman in late ’80s, to define Memetic Algorithms(MAs). In their early definition, MAs were a modification of GAs employing also a local search operator for addressing the Traveling Salesman Problem (TSP). Since MAs were not proposed as specific optimization algorithms, but as a broad class of algorithms inspired by the diffusion of the ideas and composed of multiple existing operators, the community started showing an increasing attention toward these algorithmic structures as a general guideline for addressing specific problems. MAs have been successfully applied, in recent years, to solve complex real-world problems and displayed a high performance in a large number of cases.

General structure of memetic algorithms

1. Selection of parents: Selection aims to determine the candidate solutions that will survive in the given generations and be used to create new solutions. Selection for reproduction often operates in relation with the fitness (performance) of the candidate solutions; here, performance typically amounts to the extent to which the solution maximizes/minimizes the objective function(s) (although in some cases fitness may be measured by means of a different guiding function, related to the objective function but not identical). High quality solutions have thus more chances to be chosen. Selection can also be done according to other criteria such as diversity. In such a case, only spread out individuals is allowed to survive and reproduce. If the solutions of the population are sufficiently diversified, selection can also be carried out randomly.

2. Combination of parents for offspring generation: Combination aims to create new promising candidate solutions by blending existing solutions (parents), a solution being promising if it can potentially lead the optimization process to new search areas where better solutions may be found.

3. Local improvement of offspring: The goal of local improvement is to improve the quality of an offspring as far as possible. Candidate solutions undergo refinement which correspond the life-time learning of the individuals in the original metaphor of MAs.

4. Update of the population: This step decides whether a new solution should become a member of the population and which existing solution of the population should be replaced. Often, these decisions are made according to criteria related to both quality and diversity. The policies employed for managing the population are essential to maintain an appropriate diversity of the population, to prevent the search process from premature convergence (i.e., too fast convergence toward a suboptimal region of the search space), and to help the algorithm to continually discover new promising search areas.

Thus, MAs blend together ideas from different search methodologies, and most prominently ideas from local search techniques and population-based search.

G. Harmony Search Algorithm

Harmony search (HS) algorithm has been recently developed in an analogy with an improvisation process where musicians always try to polish their pitches to obtain a better harmony. The HS algorithm is good at identifying the high performance regions of the solution space within a reasonable time. It is a derivative-free meta-heuristic algorithm. In comparison to other meta-heuristics reported in the recent literature, the HS algorithm involves a very few mathematical computations for updating the solutions based on a few logics. So, the HS algorithm executes very fast. The HS algorithm generates a new vector, after considering all of the existing vectors, whereas the GA only considers the two parent vectors. These features increase the flexibility of the HS algorithm and produce better solutions. Therefore, the HS algorithm has captured much attention and has been, successfully, applied to solve a wide range of engineering optimization problems.

The harmony search (HS) algorithm, proposed by Geem et al., is a nature inspired algorithm, mimicking the improvisation of music players [22]. The harmony in music is analogous to the optimization solution vector, and the musician’s improvisations are analogous to the local and global search schemes in optimization techniques. Similar to all population-based optimization algorithms, two main steps are distinguishable for the HS, namely, HM initialization and producing new HM by adopting the principle of HS.

Steps of basic HS algorithm

In the basic HS algorithm, each solution is called a harmony. It is represented by an n-dimension real vector. An initial randomly generated population of harmony vectors is stored in an HM. Then, a new candidate harmony is generated from all of the solutions in the HM by adopting a memory consideration rule, a pitch adjustment rule and a random re-initialization. Finally, the HM is updated by comparing the new candidate harmony vector and the worst harmony vector in the HM. The worst harmony vector is replaced by the new candidate vector if it is better than the worst harmony vector in the HM. The above process is repeated until a certain termination criterion is met. Thus, the basic HS algorithm consists of three basic phases. These are initialization, improvisation of a harmony vector and updating the HM [23].

H. Tabu Search Algorithm

Tabu Search (TS) is one of the recent metaheuristics originally developed for combinatorial optimization problems [24]. However, contributions of TS to solving continuous optimization problems are still very limited compared with other metaheuristics like GAs.

Tabu search algorithm

1. Overview

In general terms, a conventional TS algorithm is an iterative search that starts from an initial feasible solution and attempts to determine a better solution in the manner of a hill-climbing algorithm. The TS has a flexible memory to keep the information about the past steps of the search. The TS uses the past search to create and exploit the better solutions. The main two components of TS algorithm are the tabu list (TL) restrictions and the aspiration criterion (AC).

2. Tabu list restrictions

In order to prevent cycling, repeated search at the same solution, a TL is introduced. The TL stores a set of the tabu (prohibition) moves that cannot be applied to the current solution. The moves stored in TL are called tabu restrictions and used for decreasing the possibility of cycling because it prevents returning in a certain number of iterations to a solution visited recently. In this paper, the size of TL is $n \times 3$ *(row × column),* n is a number of neighborhoods around current solution. In the TL, the first column is used for storing the moves, the second column is the frequency of a move direction, and the last column is the recency (time to keep solutions) of a move.

3. Aspiration criterion

Another key issue of TS algorithm arises when all moves under consideration have been found to be tabued. The tabu status of a move is not absolute, but it can be overruled if certain conditions are met and expressed in the form of AC. If appropriate aspiration criterion is satisfied, the move will be accepted in spite of the tabu classification. Roughly speaking, AC is designed to override tabu status if a move is ‘good enough’.

4. Stopping criterion

Generally, there are several possible conditions to stop searching. Here, the stopping search is used if any of the following two conditions are satisfied: firstly, defined by the limit time. That mean, if the computational time is grater than or equal to the limit time, then the search process will stop. Secondly, defined by the maximum allowable iterations, if the iteration is grater than or equal to the maximum allowable iteration, then the search process will stop.

5. General tabu search algorithm

To solve a combinatorial optimization problem by tabu search, the basic idea is to choose randomly a feasible solution and attempt to find a best neighbor to current solution. During these procedures, the best solution is always updated and stored aside until the stopping criterion is satisfied [25].

I. Firefly Algorithm

Firefly Algorithm, developed by Yang, is used which has been successful to solve mixed variable and constrained engineering problems. It was based on the idealized behavior of the flashing characteristics of fireflies. As all fireflies are unisex so that one firefly is attracted to other fireflies regardless of their sex. Attractiveness is proportional to their brightness, thus for any two flashing fireflies, the less bright one will move towards the brighter one. The attractiveness is proportional to the brightness and they both decrease as their distance increases. If no one is brighter than a particular firefly, it moves randomly. The brightness or light intensity of a firefly is affected or determined by the landscape of the objective function to be optimized [26].

J. Teaching Learning Based Optimization

The TLBO method is based on the effect of the influence of a teacher on the output of learners in a class. Here, output is considered in terms of results or grades. The teacher is generally considered as a highly learned person who shares his or her knowledge with the learners. The quality of a teacher affects the outcome of the learners. It is obvious that a good teacher trains learners such that they can have better results in terms of their marks or grades [27].

Like other nature-inspired algorithms, TLBO is also a population based method which uses a population of solutions to proceed to the global solution. For TLBO, the population is considered as a group of learners or a class of learners. In optimization algorithms, the population consists of different design variables. In TLBO, different design variables will be analogous to different subjects offered to learners and the learners’ result is analogous to the ‘fitness’, as in other population based optimization techniques. The teacher is considered as the best solution obtained so far. The process of working of TLBO is divided into two parts. The first part consists of 'Teacher Phase' and the second part consists of 'Learner Phase'. The 'Teacher Phase' means learning from the teacher and the 'Learner Phase' means learning through the interaction between learners.

1. Teacher Phase

A good teacher brings his or her learners up to his or her level in terms of knowledge. But in practice this is not possible and a teacher can only move the mean of a class up to some extent depending on the capability of the class. This follows a random process depending on many factors. Let M_i be the mean and T_i be the teacher at any iteration *i*. T_i will try to move mean Mi towards its own level, so now the new mean will be T_i designated as M_{new}. The solution is updated according to the difference between the existing and the new mean given by

$$Difference\_Mean_i = r_i[M_{new} - T_F M_i]$$

where T_F is a teaching factor that decides the value of the mean to be changed, and r_i is a random number in the range [0, 1].The value of T_F can be either 1 or 2 which is again a heuristic step and decided randomly with equal probability as $T_F = round[1 + rand(0,1)\{2 \_ 1\}]$. This difference modifies the existing solution according to the following expression

$$X_{new,i} = X_{old,i} + Difference\_Mean_i$$

2. Learner phase

Learners increase their knowledge by two different means: one through input from the teacher and the other through interaction between themselves. A learner interacts randomly with other learners with the help of group discussions, presentations, formal communications, etc. A learner learns something new if the other learner has more knowledge than him or her. Learner modification is expressed as

For $i = 1 : P_n$
Randomly select two learners X_i *and* X_j*, where* $i \neq j$
If $f(X_i) < f(X_j)$
$X_{new,i} = X_{old,i} + r_i(X_i - X_j)$
Else
$X_{new,i} = X_{old,i} + r_i(X_j - X_i)$
End If
End For
Accept X_{new} *if it gives a better function value.*

III. CONCLUSION

As the optimization problem has evolved with time from a totally manual process to computer-based approaches. This paper presents a review of the optimization methods and techniques available for solving the problem. These methods are classified into mathematical programming based methods, Artificial Intelligence (AI) techniques and hybrid methods. The mathematical programming-based or heuristically based, such as the lambda iterative method, gradient projection method, Lagrange relaxation, linear programming, nonlinear programming, etc. fail to obtain a global optimal solution for non smooth and non convex functions. Recently, AI techniques such as GA, DE, EP, BBO, MA, PSO, etc. have successfully been applied for solving nonsmooth or nonconvex functions due to their ability to seek the global optimal solution. However, it may take much time to reach just near a global optimum. Hybrid methods which combine two or more optimization techniques are found to be more effective in finding global optimal solution.

REFERENCES

[1] Xia X and Elaiw AM, Optimal Dynamic Economic dispatch of generation: A review, Electric power systems research, Vol. 80, 2010, 975-986.

[2] Altun H and Yalcinoz T, Implementing soft computing techniques to solve economic dispatch problem in power systems, Expert systems with applications, Vol. 35, 2008, 1668-1678.

[3] Zang H, Zhang S and Hapeshi K, A review of nature-inspired algorithms, Journal of Bionic Engineering, Suppl. 7, 2010, 232-237.

[4] Zhou A, Qu B Y, Li H, Zhao S Z and Suganthan P N, Multiobjective evolutionary algorithms: A survey of the state of the art, Swarm and Evolutionary Computation, Vol. 1, 2011, 32-49.

[5] Roy R, Hinduja S and Teti R, Recent advances in engineering design optimization: Challenges and future trends, CIRP Annals – Manufacturing Technology, Vol. 57, 2008, 697-715.

[6] Nanda J and Narayanan R B, Application of genetic algorithm to economic load dispatch with line flow constraints, Electric power and energy systems, Vol. 24, 2002, 723-729.

[7] Gjorgiev B and Cepin M, A multiobjective optimization based solution for the combined economic-environmental power dispatch problem, Engineering applications of artificial intelligence, 2012.

[8] Eberhart R C and Shi Y, Particle Swarm optimization: Developments, applications and resources, IEEE Evolutionary Computation, Vol. 1, 2001, 81-86.

[9] Safari A and Shayeghi H, Iteration particle swarm optimization procedure for economic load dispatch with generator constraints, Expert systems and applications, Vol. 38, 2011, 6043-6048.

[10] Zhang Y, Gong D W and Ding Z, A bare-bones multi-objective particle swarm optimization algorithm for environmental/economic dispatch, Information sciences, Vol. 192, 2012, 213-227.

[11] Tripathi P K, Bandyopadhy S and Pal S K, Adaptive multi-objective particle swarm optimization algorithm, IEEE congress on evolutionary computation, 2007, 2281-2288.

[12] Wang L and Singh C, Stochastic economic emission load dispatch through a modified particle swarm optimization algorithm, Electric power Systems research, Vol. 78, 2008, 1466-1476.

[13] Wu L H, Wang Y N, Yuan X F and Zhou, Environmental/economic power dispatch problem using multi-objective differential evolution algorithm, Electric Power Systems research, Vol. 80, 2010, 1171-1181.

[14] Noman N and Iba H, Differential evolution for economic load dispatch problems, Electric power Systems research, Vol. 78, 2008, 1322-1331.

[15] Bhattacharya A and Chattopadhyay P K, Solving economic emission load dispatch problems using hybrid differential evolution, Applied Soft Computing, Vol. 11, 2011, 2526-2537.

[16] Roy P K, Ghoshal S P and Thakur S S, Biogeography based optimization for multi-constraint optimal power flow with emission and non-smooth cost function, Expert Systems with Applications, Vol. 37, 2010, 8221-8228.

[17] Simon D, Rarick R, Ergezer M and Du D, Analytical and numerical comparisons of biogeography based optimization and genetic algorithms, Information Sciences, Vol. 181, 2011, 1224-1248.

[18] Bhattacharya A and Chattopadhyay P K, Hybrid differential evolution with biogeography-based optimization algorithm for solution of economic emission load dispatch problems, Expert Systems with Applications, Vol. 38, 2011, 14001-14010.

[19] Neto R F T and Filho M G, Literature review regarding ant colony optimization applied to scheduling problems: guidelines for implementation and directions for future research, Engineering Applications of Artificial Intelligence, 2012, 1-12.

[20] Chandra Mohan B and Baskaran R, A survey: Ant Colony Optimization based recent research and implementation on several engineering domain, Expert Systems with Applications, Vol. 39, 2012, 4618-4627.

[21] Neri F and Cotta C, Memetic algorithms and memetic computing optimization: A literature review, Swarm and Evolutionary Computation, Vol. 2, 2012, 1-14.

[22] Chatterjee A, Ghoshal S P and Mukherjee V, Solution of combined economic and emission dispatch problems of power systems by an opposition-based harmony search algorithm, Electrical Power and Energy Systems, Vol. 39, 2012, 9-20.

[23] Sivasubramani S and Swarup K S, Environmental / economic dispatch using multi-objective harmony search algorithm, Electric Power Systems Research, Vol. 81, 2011, 1778-1785.

[24] Sa-ngiamvibool W, Pothiya S and Ngamroo I, Multiple tabu search algorithm for economic dispatch problem considering valve-point effects, Electrical Power and Energy Systems, Vol. 33, 2011, 846-854.

[25] Hedar A R and Fukushima M, Tabu search directed by direct search methods for non-linear global optimization, European Journal of operational research, Vol. 170, 2006, 329-349.

[26] Yang X, Hosseini S S and Gandomi A H, Firefly Algorithm for solving non-convex economic dispatch problems with valve loading effect, Applied Soft Computing, Vol. 12, 2012, 1180-1186.

[27] Rao R V, Savsani V J and Vakharia D P, Teaching–Learning-Based Optimization: An optimization method for continuous non-linear large scale problems, Information Sciences, Vol. 183, 2012, 1-15.

[28] Rao R V, Savsani V J and Vakharia D P, Teaching–learning-based optimization: A novel method for constrained mechanical design optimization problems, Computer Aided Design, Vol. 43, 2011, 303-315.

Security Attacks and Mechanisms in Wireless Sensor Networks

Dr.Gurjeet Singh
Associate Professor, Deptt of CSE/IT, Desh Bhagat Institute of Engg & Management, Moga
E.mail:- hi_gurjeet@rediffmail.com

***Abstract* — Wireless Sensor networks (WSN) is an emerging technology and have great potential to be employed in critical situations like battlefields and commercial applications such as building, traffic surveillance, habitat monitoring and smart homes and many more scenarios. One of the major challenges wireless sensor networks face today is security. While the deployment of sensor nodes in an unattended environment makes the networks vulnerable to a variety of potential attacks, the inherent power and memory limitations of sensor nodes makes conventional security solutions unfeasible. The sensing technology combined with processing power and wireless communication makes it profitable for being exploited in great quantity in future. The wireless communication technology also acquires various types of security threats. This paper discusses a wide variety of attacks in WSN and their classification mechanisms and different securities available to handle them including the challenges faced.**

Keywords: Wireless Sensor Network; Security Goal; Security Attacks; Defensive mechanisms; Challenges.

I. INTRODUCTION

Sensor networks are primarily designed for real-time collection and analysis of low level data in hostile environments. For this reason they are well suited to a substantial amount of monitoring and surveillance applications. Popular wireless sensor network applications include wildlife monitoring, bushfire response, military command, intelligent communications, industrial quality control, observation of critical infrastructures, smart buildings, distributed robotics, traffic monitoring, examining human heart rates etc. Majority of the sensor network are deployed in hostile environments with active intelligent opposition. Hence security is a crucial issue. One obvious example is battlefield applications where there is a pressing need for secrecy of location

and resistance to subversion and destruction of the network. Less obvious but just as important security dependent applications include:

- **Disasters:** In many disaster scenarios, especially those induced by terrorist activities, it may be necessary to protect the location of casualties from unauthorized disclosure.
- **Public Safety:** In applications where chemical, biological or other environmental threats are monitored, it is vital that the availability of the network is never threatened. Attacks causing false alarms may lead to panic responses or even worse total disregard for the signals.
- **Home Healthcare***:* In such applications, privacy protection is essential. Only authorized users should be able to query and monitor the network.

The major contribution of this paper includes classification of security attacks, security mechanisms and challenges in Wireless Sensor Networks.

II. SECURITY GOALS FOR SENSOR NETWORKS

As the sensor networks can also operate in an adhoc manner the security goals cover both those of the traditional networks and goals suited to the unique constraints of adhoc sensor networks. The security goals are classified as primary and secondary [5]. The primary goals are known as standard security goals such as Confidentiality, Integrity, Authentication and Availability (CIAA). The secondary goals are Data Freshness, Self- Organization, Time Synchronization and Secure Localization.

The primary goals are:

- **Data Confidentiality***:* Confidentiality is the ability to conceal messages from a passive attacker so that any message communicated via the sensor network remains confidential. This is the most important issue in network security. A sensor node should not reveal its data to the neighbors.
- **Data Authentication***:* Authentication ensures the reliability of the message by identifying its origin. Attacks in sensor networks do not just involve the alteration of packets; adversaries can also inject additional

false packets [14]. Data authentication verifies the identity of the senders and receivers. Data authentication is achieved through symmetric or asymmetric mechanisms where sending and receiving nodes share secret keys. Due to the wireless nature of the media and the unattended nature of sensor networks, it is extremely challenging to ensure authentication.

- **Data Integrity:** Data integrity in sensor networks is needed to ensure the reliability of the data and refers to the ability to confirm that a message has not been tampered with, altered or changed. Even if the network has confidentiality measures, there is still a possibility that the data integrity has been compromised by alterations. The integrity of the network will be in trouble when:
- **Data Availability:** Availability determines whether a node has the ability to use the resources and whether the network is available for the messages to communicate. However, failure of the base station or cluster leader's availability will eventually threaten the entire sensor network. Thus availability is of primary importance for maintaining an operational network.

The Secondary goals are:

- **Data Freshness***:* Even if confidentiality and data integrity are assured, there is a need to ensure the freshness of each message. Informally, data freshness [4] suggests that the data is recent, and it ensures that no old messages have been replayed. To solve this problem a nonce, or another timerelated counter, can be added into the packet to ensure data freshness.
- **Self-Organization***:* A wireless sensor network is a typically an ad hoc network, which requires every sensor node be independent and flexible enough to be self-organizing and self-healing according to different situations. There is no fixed infrastructure available for the purpose of network management in a sensor network. This inherent feature brings a great challenge to wireless sensor network security. If self-organization is lacking in a sensor network, the damage resulting from an attack or even the risky environment may be devastating.
- **Time Synchronization:** Most sensor network applications rely on some form of time synchronization. Furthermore, sensors may wish to compute the end-to-end delay of a packet as it travels between two pairwise sensors. A more collaborative sensor network may require group synchronization [4] for tracking applications.

III. ATTACKS ON SENSOR NETWORKS

Wireless Sensor networks are vulnerable to security attacks due to the broadcast nature of the transmission medium. Furthermore, wireless sensor networks have an additional vulnerability because nodes are often placed in a hostile or dangerous environment where they are not physically protected. Basically attacks are classified as active attacks and passive attacks.

A. Passive Attacks

The monitoring and listening of the communication channel by unauthorized attackers are known as passive attack. The Attacks against privacy is passive in nature.

1) Attacks against Privacy

The main privacy problem is not that sensor networks enable the collection of information. In fact, much information from sensor networks could probably be collected through direct site surveillance. Rather, sensor networks intensify the privacy problem because they make large volumes of information easily available through remote access. Hence, adversaries need not be physically present to maintain surveillance. They can gather information at low-risk in anonymous manner. Some of the more common attacks [8] against sensor privacy are:

- **Monitor and Eavesdropping***:* This is the most common attack to privacy. By snooping to the data, the adversary could easily discover the communication contents. When the traffic conveys the control information about the sensor network configuration, which contains potentially more detailed information than accessible through the location server, the eavesdropping can act effectively against the privacy protection.
- **Traffic Analysis***:* Even when the messages transferred are encrypted, it still leaves a high possibility analysis of the communication patterns. Sensor activities can potentially reveal enough information to enable an adversary to cause malicious harm to the sensor network.

B. Active Attacks

The unauthorized attackers monitors, listens to and modifies the data stream in the communication channel are known as active attack.

1) Routing Attacks in Sensor Networks

The attacks which act on the network layer are called routing attacks. The following are the attacks that happen while routing the messages.

- **Selective Forwarding:** A malicious node can selectively drop only certain packets. Especially effective if combined with an attack that gathers much traffic via the node. In sensor networks it is assumed that nodes faithfully forward received messages. But some compromised node might refuse to forward packets, however neighbors might start using another route.[3]
- **Sinkhole Attack***:* Attracting traffic to a specific node in called sinkhole attack. In this attack, the adversary's goal is to attract nearly all the traffic from a particular area through a compromised node. Sinkhole attacks typically work by making a compromised node look especially attractive to surrounding nodes. [3]
- **Sybil Attacks*:*** A single node duplicates itself and presented in the multiple locations. The Sybil attack targets fault tolerant schemes such as distributed storage, multipath routing and topology maintenance. In a Sybil attack, a single node presents multiple identities to other nodes in the network. Authentication and encryption techniques can prevent an outsider to launch a Sybil attack on the sensor network. [3]
- **Wormholes Attacks:** In the wormhole attack, an attacker records packets (or bits) at one location in the network, tunnels them to another location, and retransmits them into the network. [3]
- **HELLO flood attacks:** An attacker sends or replays a routing protocol's HELLO packets from one node to another with more energy. This attack uses HELLO packets as a weapon to convince the sensors in WSN. In this type of attack an attacker with a high radio transmission range and processing power sends HELLO packets to a number of sensor nodes that are isolated in a large area within a WSN. The sensors are thus influenced that the adversary is their neighbor. As a result, while sending the information to the base station, the victim nodes try to go through the attacker as they know that it is their neighbor and are ultimately spoofed by the attacker. [3]

2) Denial of Service

Denial of Service (DoS) is produced by the unintentional failure of nodes or malicious action. DoS attack is meant not only for the adversary's attempt to subvert, disrupt, or destroy a network, but also for any event that diminishes a network's capability to provide a service. In wireless sensor networks, several types of DoS attacks in different layers might be performed. At physical layer the DoS attacks could be jamming and tampering, at link layer, collision, exhaustion and unfairness, at network layer, neglect and greed, homing, misdirection, black holes and at transport layer this attack could be performed by malicious flooding and de-synchronization. The mechanisms to prevent DoS attacks include payment for network resources, pushbac strong authentication and identification of traffic. [2]

3) Node Subversion

Capture of a node may reveal its information including disclosure of cryptographic keys and thus compromise the whole sensor network. A particular sensor might be captured, and information (key) stored on it might be obtained by an adversary. [6]

4) Node Malfunction

A malfunctioning node will generate inaccurate data that could expose the integrity of sensor network especially if it is a data-aggregating node such as a cluster leader [6].

5) Node Outage

Node outage is the situation that occurs when a node stops its function. In the case where a cluster leader stops functioning, the sensor network protocols should be robust enough to mitigate the effects of node outages by providing an alternate route [6].

6) Physical Attacks

Sensor networks typically operate in hostile outdoor environments. In such environments, the small form factor of the sensors, coupled with the unattended and distributed nature of their deployment make them highly susceptible to physical attacks, i.e., threats due to physical node destructions. Unlike many other attacks mentioned above, physical attacks destroy sensors permanently, so the losses are irreversible. For instance, attackers can extract cryptographic secrets, tamper with the associated circuitry, modify programming in the sensors, or replace them with malicious sensors under the control of the attacker.

7) Message Corruption

Any modification of the content of a message by an attacker compromises its integrity.[9]

8) False Node

A false node involves the addition of a node by an adversary and causes the injection of malicious data. An intruder might add a node to the system that feeds false data or prevents the passage of true data. Insertion of malicious node is one of the most dangerous attacks that can occur. Malicious code injected in the network could spread to all nodes, potentially destroying the whole network, or even worse, taking over the network on behalf of an adversary.[9]

9) Node Replication Attacks

Conceptually, a node replication attack is quite simple; an attacker seeks to add a node to an existing sensor network by copying the nodeID of an existing sensor node. A node replicated in this approach can severely disrupt a sensor network's performance. Packets can be corrupted or even misrouted. This can result in a disconnected network, false sensor readings, etc. If an attacker can gain physical access to the entire network he can copy cryptographic keys to the replicated sensor nodes. By inserting the replicated nodes at specific network points, the attacker could easily manipulate a specific segment of the network, perhaps by disconnecting it altogether. [1]

10) Passive Information Gathering

An adversary with powerful resources can collect information from the sensor networks if it is not encrypted. An intruder with an appropriately powerful receiver and well-designed antenna can easily pick off the data stream. Interception of the messages containing the physical locations of sensor nodes allows an attacker to locate the nodes and destroy them. Besides the locations of sensor nodes, an adversary can observe the application specific content of messages including message IDs, timestamps and other fields. To minimize the threats of passive information gathering, strong encryption techniques needs to be used. [8]

IV. SECURITY MECHANISM

The security mechanisms are actually used to detect, prevent and recover from the security attacks. A wide variety of security schemes can be invented to counter malicious attacks and these can be categorized as highlevel and low-level.

A. Low-Level Mechanism

Low-level security primitives for securing sensor networks includes:

1) Key establishment and trust setup

The primary requirement of setting up the sensor network is the establishment of cryptographic keys. Generally the sensor devices have limited computational power and the public key cryptographic primitives are too expensive to follow. Key-establishment techniques need to scale to networks with hundreds or thousands of nodes. In addition, the communication patterns of sensor networks differ from traditional networks; sensor nodes may need to set up keys with their neighbors and with data aggregation nodes. The disadvantage of this approach is that attackers who compromised sufficiently and many nodes could also reconstruct the complete key pool and break the scheme. [1]

2) Secrecy and authentication

Most of the sensor network applications require protection against eavesdropping, injection, and modification of packets. Cryptography is the standard defense. Remarkable system trade-offs arise when incorporating cryptography into sensor networks. For point-to-point communication, end-to-end cryptography achieves a high level of security but requires that keys be set up among all end points and be incompatible with passive participation and local broadcast. Link-layer cryptography with a network wide shared key simplifies key setup and supports passive participation and local broadcast, but intermediate nodes might eavesdrop or alter messages. The earliest sensor networks are likely to use link layer cryptography, because this approach provides the greatest ease of deployment among currently available network cryptographic approaches. [6]

3) Privacy

Like other traditional networks, the sensor networks have also force privacy concerns. Initially the sensor networks are deployed for legitimate purpose might subsequently be used in unanticipated ways. Providing awareness of the presence of sensor nodes and data acquisition is particularly important. [1]

4) Robustness to communication denial of service

An adversary attempts to disrupt the network's operation by broadcasting a high-energy signal. If the transmission is powerful enough, the entire system's communication could be jammed. More sophisticated attacks are also possible; the adversary might inhibit communication by violating the 802.11 medium access control (MAC) protocol by, say, transmitting while a neighbor is also transmitting or by continuously requesting channel access with a request-tosend signal. [1]

5) Secure routing

Routing and data forwarding is a crucial service for enabling communication in sensor networks. Unfortunately, current routing protocols suffer from many security vulnerabilities. For example, an attacker might launch denialof- service attacks on the routing protocol, preventing communication. The simplest attacks involve injecting malicious routing information into the network, resulting in routing inconsistencies. Simple authentication might guard against injection attacks, but some routing protocols are susceptible to replay by the attacker of legitimate routing messages. [6]

B. High-Level Mechanism

High-level security mechanisms for securing sensor networks, includes secure group management, intrusion detection, and secure data aggregation.

1) Secure group management

Each and every node in a wireless sensor network is limited in its computing and communication capabilities. However, interesting in-network data aggregation and analysis can be performed by groups of nodes. For example, a group of nodes might be responsible for jointly tracking a vehicle through the network. The actual nodes comprising the group may change continuously and quickly. Many other key services in wireless sensor networks are also performed by groups. Consequently, secure protocols for group management are required, securely admitting new group members and supporting secure group communication. The outcome of the group key computation is normally transmitted to a base station. The output must be authenticated to ensure it comes from a valid group. [1]

2) Intrusion detection

Wireless sensor networks are susceptible to many forms of intrusion. Wireless sensor networks require a solution that is fully distributed and inexpensive in terms of communication, energy, and memory requirements. The use of secure groups may be a promising approach for decentralized intrusion detection. [1]

3) Secure data aggregation

One advantage of a wireless sensor network is the finegrain sensing that large and dense sets of nodes can provide. The sensed values must be aggregated to avoid overwhelming amounts of traffic back to the base station. For example, the system may average the temperature of a geographic region, combine sensor values to compute the location and velocity of a moving object, or aggregate data to avoid false alarms in real-world event detection. Depending on the architecture of the wireless sensor network, aggregation may take place in many places in the network. All aggregation locations must be secured. [6]

V. CHALLENGES OF SENSOR NETWORKS

The nature of large, ad-hoc, wireless sensor networks presents significant challenges in designing security schemes. A wireless sensor network is a special network which has many constraint compared to a traditional computer network.

A. Wireless Medium

The wireless medium is inherently less secure because its broadcast nature makes eavesdropping simple. Any transmission can easily be intercepted, altered, or replayed by an adversary. The wireless medium allows an attacker to

easily intercept valid packets and easily inject malicious ones. Although this problem is not unique to sensor networks, traditional solutions must be adapted to efficiently execute on sensor networks. [7]

B. Ad-Hoc Deployment

The ad-hoc nature of sensor networks means no structure can be statically defined. The network topology is always subject to changes due to node failure, addition, or mobility. Nodes may be deployed by airdrop, so nothing is known of the topology prior to deployment. Since nodes may fail or be replaced the network must support self-configuration.

Security schemes must be able to operate within this dynamic environment.

C. Hostile Environment

The next challenging factor is the hostile environment in which sensor nodes function. Motes face the possibility of destruction or capture by attackers. Since nodes may be in a hostile environment, attackers can easily gain physical access to the devices. Attackers may capture a node, physically disassemble it, and extract from it valuable information (e.g. cryptographic keys). The highly hostile environment represents a serious challenge for security researchers.

D. Resource Scarcity

The extreme resource limitations of sensor devices pose considerable challenges to resource-hungry security mechanisms. The hardware constraints necessitate extremely efficient security algorithms in terms of bandwidth, computational complexity, and memory. This is no trivial task. Energy is the most precious resource for sensor networks. Communication is especially expensive in terms of power. Clearly, security mechanisms must give special effort to be communication efficient in order to be energy efficient. [5]

VI. CONCLUSION

The deployment of sensor nodes in an unattended environment makes the networks vulnerable. Wireless sensor networks are increasingly being used in military, environmental, health and commercial applications. Sensor networks are inherently different from traditional wired networks as well as wireless ad-hoc networks. Security is an important feature for the deployment of Wireless Sensor Networks. This paper summarizes the attacks and their classifications in wireless sensor networks and also an attempt has been made to explore the security mechanism widely used to handle those attacks. The challenges of Wireless Sensor Networks are also briefly discussed.

REFERENCES

[1] Adrian Perrig, John Stankovic, David Wagner, "Security in Wireless Sensor Networks" Communications of the ACM, Page53-57, year 2004

[2] Al-Sakib Khan Pathan, Hyung-Woo Lee, Choong Seon Hong, "Security in Wireless Sensor Networks: Issues and Challenges**",** International conference on Advanced Computing Technologies**,** Page1043-1045, year 2006

[3] Chris Karlof, David Wagner, "Secure Routing in Wireless Sensor Networks: Attacks and Countermeasures", AdHoc Networks (elsevier), Page: 299-302, year 2003

[4] Ian F. Akykildiz, Weilian Su, Yogesh Sankarasubramaniam, and Erdal Cayirci, "A Survey on Sensor Networks", IEEE Communication Magazine, year 2002

[5] John Paul Walters, Zhengqiang Liang, Weisong Shi, Vipin Chaudhary, "Wireless Sensor Network Security: A Survey", Security in Distributed, Grid and Pervasive Computing Yang Xiao (Eds), Page3-5, 10-15, year 2006

[6] Pathan, A.S.K.; Hyung-Woo Lee; Choong Seon Hong, "Security in wireless sensor networks: issues and challenges" Advanced Communication Technology (ICACT), Page(s):6, year 2006

[7] Tahir Naeem, Kok-Keong Loo, Common Security Issues and Challenges in Wireless Sensor Networks and IEEE 802.11 Wireless Mesh Networks, International Journal of Digital Content Technology and its Applications, Page 89-90 Volume 3, Number 1, year 2009

[8] Undercoffer, J., Avancha, S., Joshi, A. and Pinkston, J. "Security for sensor networks". In Proceedings of the CADIP Research Symposium, University of Maryland, Baltimore County, USA, year 2002
http://www.cs.sfu.ca/~angiez/personal/paper/sensor-ids.pdf

[9] Zia, T.; Zomaya, A., "Security Issues in Wireless Sensor Networks", Systems and Networks Communications (ICSNC) Page(s):40 – 40, year 2006.

Image Encryption using Chaotic Systems

Ratinder Kaur [1] and Simardeep Kaur[2]
[1](Lecturer,SIET,Manawala,Amritsar)
[2](Lecturer,ACET,Manawala,Amritsar)
Email: Er.ratinderkaur@gmail.com, simar.flowers@gmail.com

***Abstract*— Image encryption algorithms are based on chaotic systems (Lorenz or Chen or LU chaotic system based on 16-byte key) to shuffle the position of the image pixels (pixel position permutation) and uses another same chaotic maps confuse the cipher image and plain image which is called as pixel diffusion which further increases the security of the image. This system provides the bigger key space and a very high security to threats.**

Index Terms- Encryption algorithms, chaotic system, pseudo random numbers, pixel diffusion.

I. Introduction

Image encryption schemes have been increasingly studied to meet the demand for real-time secure image transmission over the Internet and through wireless networks. Traditional image encryption [5] algorithm such as data encryption standard (DES), has the weakness of low-level efficiency when the image is large. The chaos-based encryption[5,6] has suggested a new and efficient way to deal with the intractable problem of fast and highly secure image encryption. After Matthews proposed the chaotic encryption algorithm in 1989, increasing researches of image encryption technology are based on chaotic systems Recently there have been many papers on chaotic encryption scheme. Chaotic systems have many important properties, such as the sensitive dependence on initial conditions and system parameters, pseudorandom property, no periodicity and topological transitivity, etc. Most properties meet some requirements such as diffusion and mixing in the sense of cryptography. Therefore, chaotic cryptosystems have more useful and practical applications. One-dimensional chaotic system with the advantages of high-level efficiency and simplicity, such as Logistic map, has been widely used now. But their weakness, such as small key space and weak security, is also disturbing Cryptography studies how to design good (secure and fast) encryption algorithms, and cryptanalysis tries to find security weaknesses of existing algorithms and studies.

An encryption scheme is called a cipher (or a cryptosystem). The encryption and decryption procedure of a cipher is depicted in Figure 1.

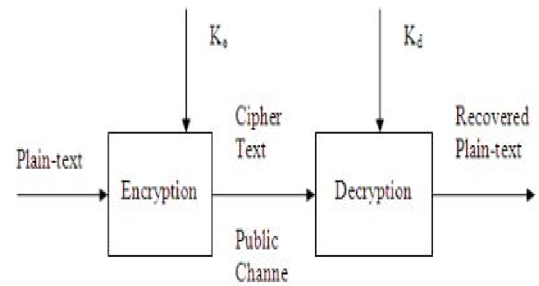


Figure 1. Encryption and Decryption procedure of a Cipher

The message for encryption is called plaintext, and the encrypted message is called cipher-text, which are denoted here by P and C, respectively. The encryption procedure of a cipher can be described as C=Eke(P), where Ke is the encryption key and E(.) is the encryption function. Similarly, the decryption procedure is P=Dkd (C), where Kd is the decryption key and D(.) is the decryption function. When Ke=Kd, the cipher is called a private-key cipher or a symmetric cipher For private key ciphers, the encryption-decryption key must be transmitted from the sender to the receiver via a separate secret channel. When Ke=!Kd, cipher is called a public-key cipher or an asymmetric cipher. . For public-key ciphers, the encryption key Ke is published, and the decryption key Kd is kept secret, for which no additional secret channel is needed for key transfer. The cryptosystems can be classified with respect to the structure of encryption algorithm as stream ciphers and block ciphers.

Stream cipher is the method in which a key generator produces a bit stream (the Key stream) which enciphers the plain-text bit stream by simple modulo 2 additions. A stream cipher system thus hides the plain-text bit by changing the bits of it in a random way. An interceptor, who does not know the key, will not know which bits have been changed (corresponding to the occurrence of “1” in the key stream), or which ones remain unchanged (“0” in the key stream). An ideal stream cipher would use a physical (true) random number generator a Key generator. Since its output cannot be reproduced, however, decipherment would be impossible, unless the whole Key stream, with the same length as the plain-text, is transported to the legitimate receiver via a safe channel. This procedure is often impractical. Therefore mostly so-called pseudo-random number generators with special properties controlled by a relatively short Key have to be used instead as key generators.

Unlike the stream ciphers, where only one bit at a time is ciphered, whole blocks of bits are treated simultaneously. In this case the plain-text information is hidden by the fact that, depending on the key, a cipher-text block can be deciphered to any combination of plain-text bits or to as many combinations as the keys. If the keys are chosen with equal probability, then to the interceptor observing a ciphertext block, all the possible plain-text blocks are equally likely to have occurred.

Cryptography is a permanent field of interest at all time. At present secret communication plays an increasing role in many fields of common life, like banking, industry, commerce, telecommunication etc. Owing to the advance in network technology, information security is an increasingly important problem. Popular application of multimedia technology and increasing transmission ability of network gradually leads to us to acquire information directly and clearly through images. Hence, data security has become a critical and imperative issue. Encryption is such a way that its content can be reconstructed only by a legal recipient. The technology of encryption is called cryptology. Cryptology is the branch of science dealing with the theory of secure communication algorithms. Cryptography is the process of transforming information (plain-text) into unintelligible form (cipher-text) so that it may be sent over insecure channels or it may be stored in insecure files. Cryptographic procedures, can also be used for personal identification, digital signature, access control etc.

II. RELATED WORKS

The chaos-based image cryptosystem mainly consists of two stages [2]. The plain image is given at its input. There are two stages in the chaos- based image cryptosystem. The confusion stage is the pixel permutation where the position of the pixels is scrambled over the entire image without disturbing the value of the pixels and the image becomes unrecognizable. The pixel permutation is carried out by a chaotic system [1,2]. The chaotic behavior is controlled by the initial conditions and control parameters which are derived from the 16-character key.

To improve the security, the second stage of the encryption process aims at changing the value of each pixel in the whole image an important tool to protect image from attackers. The basic idea of encryption[5,6] is to modify the message.

In the diffusion stage, the pixel values are modified sequentially by the sequence generated from one of the three chaotic systems selected by external key. The whole confusion-diffusion round repeats for a number of times to achieve a satisfactory level of security. The randomness property inherent in chaotic maps makes it more suitable for image encryption.

III. ARCHITECTURE OF AN CHAOS BASED IMAGE CRYPTOSYSTEM

The chaos-based image cryptosystem mainly consists of two stages. The plain image is given at its input The typical architecture of the chaos-based image cryptosystems is depicted in Figure 2. There are two stages in the chaosbased image cryptosystem The confusion stage is the pixel permutation where the position of the pixels is scrambled over the entire image without disturbing the value of the pixels and the image becomes unrecognizable.

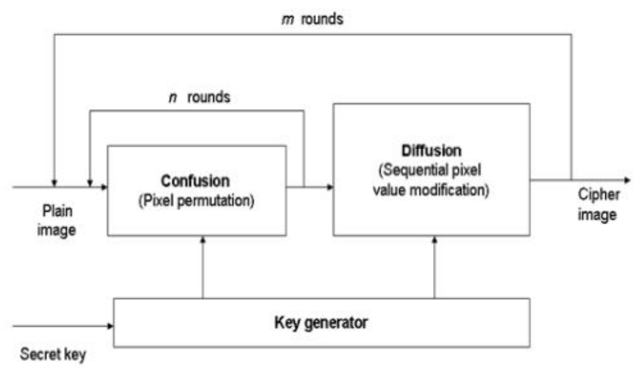


Figure 2. Architecture of proposed Chaos-based image cryptosystem

Therefore these initial conditions and control parameters serve as the secret key. It is not very secure to have only the permutation stage since it may be broken by any attack. To improve the security, the second stage of the encryption process aims at changing the value of each pixel in the whole image. The process of diffusion is also carried out through a chaotic map which is mainly dependent on the initial conditions and control parameters.

In the diffusion stage, the pixel values are modified sequentially by the sequence generated from one of the three chaotic systems selected by external key. The whole confusion-diffusion round repeats for a number of times to achieve a satisfactory level of security. The randomness property inherent in chaotic maps makes it more suitable for image encryption.

IV. PROPOSED CRYPTOSYSTEM

A. Encryption System

The proposed scheme is shown in Figure 3. Different chaotic systems are employed in confusion and diffusion stages. Also complex chaotic maps are chosen rather than the simple ones to further enhance the complexity of the algorithm and thereby improving the security. The input to the cryptosystem is the plain image which is to be encrypted. The cryptosystem consists of two stages.

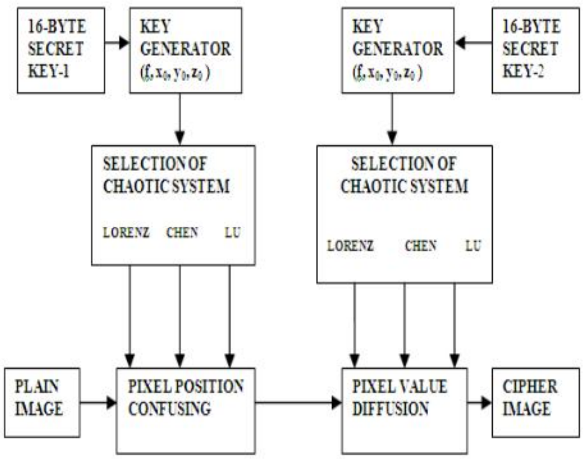


Figure 3. Architecture of proposed Chaos-based image cryptosystem

The first stage is the confusion stage and the second one is the diffusion stage. Among the three chaotic dynamic systems namely Lorenz, Chen and LU one is selected by the system parameter which is obtained from the key and it is applied to the digital color image encryption because of higher secrecy of high-dimension chaotic system. The second step of the encryption process is to encrypt the shuffled image by changing its pixel values based on one of the three high-dimensional chaotic systems (Lorenz, Chen and LU). This is referred to as the diffusion stage.

The initial conditions and the control parameters used to generate the chaos sequence in both the stages serve as the secret key in the two stages. The resulting image is the Cipher image. Separate key is used for permutation and diffusion stages of the encryption process to improve security of the algorithm.

B. Decryption System

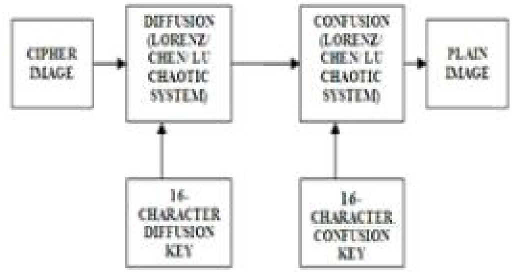


Figure 4. Chaos based Decryption system

The decryption system is illustrated in the Figure 4. The first stage in the decryption process is the diffused image decryption stage. In the encryption process, the pixel value diffusion was carried out with any one of the three chaotic systems. Therefore, in the decryption process to retrieve the original pixel values, again any one of the chaotic system (Lorenz, Chen, Lu) is employed in the first stage of decryption. The first stage of decryption process uses the three dimensional [8] sequence generated by any one of the chaotic system .It is a kind of high-dimensional maps and complex enough The initial conditions that were used in the encryption process should be used here and this serves as the decryption key for the first stage. Second, in the encryption process, the pixel position permutation was carried out with any one of the chaotic system. The initial conditions and control parameters for generating the chaos-sequence were used as the confusion key. Therefore in the decryption process, the same chaotic systems with same confusion key are used to get the original position of the image. The output of the decryption system gives the original image.

V. RESULTS AND DISCUSSION

The proposed image encryption system uses any one of the chaotic system for pixel position permutation and one of the same chaotic system for pixel value modification. Color Lena image of size 256 × 256 was taken as the test image. The original image taken for the work is given in Figure 5. The pixel position permuted image after applying one of chaotic systems such as Lorenz, Chen and Lu was obtained.Then the diffused image was obtained.

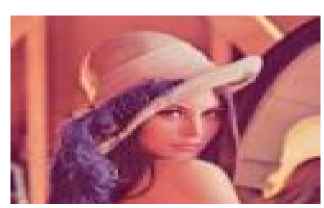


Figure 5. Original image

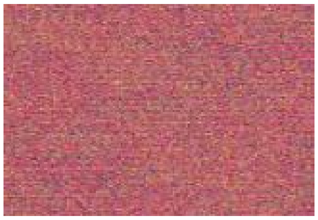


Figure 6. Confused image used for diffusion

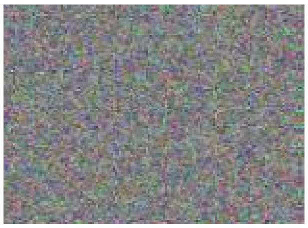


Figure 7. Encrypted image

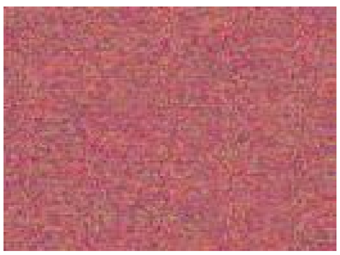


Figure 8. Undiffused image

And this will serve as the input image for the second stage of decryption. The decrypted image after applying Lorenz ,Chen and Lu chaotic systems was obtained.

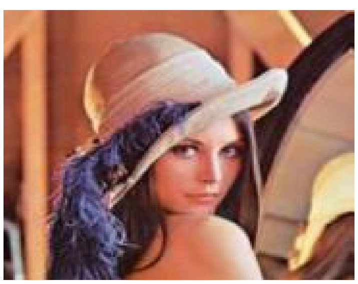


Figure 9. Decrypted image

VII. FUTURE ENHANCEMENT

Currently the chaos-based scheme was designed for still images. The chaos-based image encryption scheme can be applied to moving images as well. The prevalence of multimedia technology in the society has promoted digital images and videos to play a more significant role than the traditional texts, which demands a serious protection of users' privacy. To fulfill such security and privacy needs in various applications, encryption of images and videos is very important to frustrate malicious attacks from unauthorized parties. Hence in the second phase of the paper, it is purposed to design a chaos-based image encryption scheme for moving images (videos).

VIII. CONCLUSION

Based on the design rules discussed earlier, the new image encryption scheme was designed. A suitable chaotic map preserving the properties of chaos after discretization was chosen. By choosing a high dimensional chaotic system, the key space is increased. Complex non-linearity was preserved by choosing suitable chaotic maps. Repeated permutations are avoided but pixel values are changed by the diffusion function. By incorporating all these features, the proposed cryptosystem avoids all the crypto graphical weaknesses of earlier chaos-based encryption systems. Number of security analysis were carried out on the new algorithm and simulation results show that encryption and decryption are good and the algorithm has good security and robustness.

REFERENCES

[1] Alfyorov A.P, Zubov A.Y, Kuzmin A.C., Cheremushkin A.V. of a cryptography basis: the manual, 2-editions : Gelios APB, 2002 480 p.

[2] Moldavjan A. A, Moldovayn N.A. introduction in cryptosystem with an open key. Sank- Petersburg"BhV-Peterburg" 2005г. 288c. Alfyorov A.P, Zubov A.Y, Kuzmin A.C., Cheremushkin A.V. of a cryptography basis: the manual, 2- editions : Gelios APB, 2002 480 p.

[3] Chengqing Li, "On the security of a class of Image Encryption Scheme", IACR's Crytology ePrint Archive:Report 2007/339, August 2007.

[4] Zhang Han, Wang Xiu Feng, Li Zhao Hui, Liu Da Hai, Lin You Chou,"A New Image Encryption Algorithm Based on Chaos System", Proceedings of the 2003 IEEE International Conference on Robotics, Intelligent Systems and Signal Processing, Changsha, China, pp.778-782, October 2003.

[5]Guang ZH, Huang FJ, Guan WJ, "Chaos-based Image Encryption Algorithm", Physics Letters A, Vol.346, pp.153 – 157, 2005.

Security of Digital Signature in E- Commerce Using Elliptic Curve Cryptography

Gagandeep Aul [1], Mandeep Handa[2]
[1]Department of Electronic and Communications Engineering, Punjab Technical University, Jalandhar,
[2]Department of Computer Applications, Punjab Technical University, Jalandhar,
Email: gagandeepaul79@gmail.com, mandeephanda@gmail.com,

***ABSTRACT* - E-commerce security is the task of providing efficient means of dispute resolution to parties engaged in electronic transactions. Security is the critical issue in E-Commerce. The current research and development in this area is still in its infancy. Cryptography [1] and digital signature ensure the confidentially and privacy of message. Digital signatures that allows the verifications of the 'origin' of messages. Today main constraint for cryptography implementation are- Bandwidth, Memory, Execution time constraint. Digital signature from RSA suffer from these constraints as it require exponential calculation which is time consuming method, also generate keys with large size which are difficult to consume in embedding system. ECDSA algorithm removes these constraints to large extent. It removes exponential calculation with "point addition" method on elliptic curve. It generates keys smaller in size. Elliptic curve cryptography requires less space as well as short transmission time as compared to RSA.**

Key Terms:- Cryptography, ECDSA, Elliptic-Curve.

I. INTRODUCTION

Digital transactions [3] have become commonplace, and in some cases inextricably linked to modern life. This technological dependency requires that information be unaltered and confidential. Before the age of web environments, large distributed networks, and the proliferation of electronic data, cryptosystems required only symmetric algorithms. Individuals could maintain the privacy of information with one secret key. With the advent of the Internet era, many obstacles to trust and security appeared. Information security concerns were introduced such as proving the identity of an individual, identifying unwanted modification of data, and securing information without complex key exchanges. Today Technologies are going to emerge a new field. In E-Business Internet play very important roles. After 2002 Microsoft designed a compact Edition of window with Compact Edition of Microsoft SQL server known as MS-SQL(CE) to deals with E-Business over PDA devices such like Mobile Phone, Smart Card etc. Microsoft add these software with ASP.NET 2.0. But we have to ensure for the following properties of these devices.

1. Compact Size- Portability.
2 High Computation speed.
3. Security.

So in this paper my problem is to search a good secure technique to ensure above mentioned characteristic of PDA devices.

The main problem is to determine techniques, which ensure the confidentiality and privacy of message. Cryptography is one efficient way to ensure that if sent message fall into wrong hands, they can not read it. It is the art of secret writing. In E-Commerce, [5] Digital Signatures allow the verification of the 'origin' of messages (Transactions). We use the concept of RSA (by Rivest, Shamir and Adleman.) and Elliptic Curve Algorithms to implement Digital Signature. Our Problem is to find the equation of polynomial such that it is to complex to design its elliptic curve .An elliptic-curve group for cryptography comes from the multiples of a generating point G, a two-dimensional point on an elliptic curve over a finite field. In practice, the finite fields used are either integers modulo large primes, or a similar construction using 0/1 polynomials.

The Elliptic-Curve Digital Signature Algorithm (ECDSA) is a Digital Signature Scheme based on ECC. Like all Digital Signatures, ECDSA is used for authentication and Integrity, but because it's based on ECC, its keys are smaller and its implementation is more efficient The crucial property of an elliptic curve is that we can define a rule for adding" two points which are on the curve, to obtain a 3rd point which is also on the curve. This addition rule satisfies the normal properties of addition. In math jargon, the points and the addition law form a finite Abelian group.

II. Electronic commerce techniques

A brief overview of some of the electronic commerce techniques available today. It is: Electronic Commerce Cryptography is extremely useful to electronic commerce in the areas of payment systems, and transactions over open networks. While several protocols and payment systems exist, the most widely used protocol for Internet transactions is SSL [5].

2.1. Electronic money:

Electronic money (also called electronic cash or digital cash) is a term that is still fairly vague and undefined. It refers to transactions carried out electronically with a net result of funds transferred from one party to another. Electronic money may be either debit or credit. Digital cash per se is basically another currency, and digital cash transactions can be visualized as a foreign exchange market. This is because we need to convert an amount of money to digital cash before we can spend it. The conversion process is analogous to purchasing foreign currency [5].

2.2 The iKP:

The Internet Keyed Payments Protocol (iKP) is architecture for secure payments involving three or more parties. Developed at IBM's T.J. Watson Research Center and Zurich Research Laboratory, the protocol defines transactions of a ``credit card'' nature, where a buyer and seller interact with a third party ``acquirer,'' such as a credit-card system or a bank, to authorize transactions. The protocol is based on public-key cryptography. IKP is no longer widely in use; however it is the current foundation for SET [5].

2.3 SET:

Visa and MasterCard have jointly developed the Secure Electronic Transaction (SET) protocol as a method for secure, cost effective bankcard transactions over open networks. SET includes protocols for purchasing goods and services electronically, requesting authorization of payment, and requesting` `credentials'' (that is, certificates) binding public keys to identities, among other services. Once SET is fully adopted, the necessary confidence in secure electronic transactions will be in place, allowing merchants and customers to partake in electronic commerce. SET supports DES for bulk data encryption and RSA for signatures and public-key encryption of data encryption keys and bankcard numbers.

III. Security in E-Commerce

The successful functioning of e-commerce security depends on a complex interrelationship between several applications development platforms, database management systems, and systems software and network infrastructure [5]. The vast growth of Internet-based e-commerce. has been tempered by legitimate concerns over the security of such a system [5].

3.1. Security needs of E- Commerce[5]

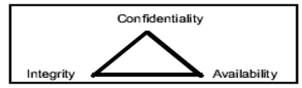


Figure 2. The CIA Triad [1].

Security protection starts with the preservation of the confidentiality, integrity and avail -ability (CIA) of data and computer resources. These three tenets of information security are sometimes represented in the CIA Triad in the Figure

3.2. Including the elements of the CIA

Triad, the six security needs in e-commerce are:

- . Access Control
- . Privacy/Confidentiality
- . Authentication
- . Non Repudiation
- . Integrity
- . Availability

IV. Proposed Approach

The key length for a secure RSA [2] transmission is typically 1024 bits. 512 bits is now no longer considered secure. For more security we are paranoid, use 2048 or even 4096 bits. With the faster computers available today, the time taken to encrypt and decrypt even with a 4096-bit modulus really isn't an issue anymore. In practice, it is still effectively impossible to crack a message encrypted with a 512-bit key. Up to 256 bits. That gives us lots of security. Unlike our simple examples above where we had to deal with a series of integers, to encrypt a 256-bit key with a 1024-bit RSA modulus means we only need a single representative message integer. In fact, you need to pad the 256 bits to ensure that we have a large enough integer before we encrypt it with RSA. 1024 bits is 128 bytes long, so we have quite a handful of data to deal with.

4.1 Security Using Elliptic Curve Cryptography:-

Elliptic curve cryptography's strengths make it most suitable for resource-constrained systems. ECC provides greater security for a given key size and can be efficiently and compactly implemented. These attributes make it well suited for systems with constraints on processor speed, security, heat production, power consumption, bandwidth, and memory. Cell phones, PDAs, wireless devices, laptops, and smartcards are applications that benefit from elliptic curve cryptosystems.

4.2 Elliptic Curve Airthmatic:-

*Abelian Groups:-*An abelian group G, sometimes denoted by {G,*} is a set of elements with a binary operation, denoted by *,that associates to each ordered pair(a,b) of elements (a*b) in G such that the following axioms are obeyed.

(1) Closure: If a and b belong to G, then a * b is also in G.
(2) Associate: a * (b * c) = (a * b) * c for all a, b, c in G
(3) Identity element: There is an element in G such that va*e=e*a=a for all a in G.
(4) Inverse element: For each a in G there is an element in G
(5) Commutative: a * b = b * a for all a, b in

A number of public-key ciphers are based on the use of an abelian group. For example, Diffie-Hellman key exchange involves multiplying pairs of nonzero integers modulo a prime number q. Keys are generated by exponentiation over the gop, with exponentiation defined as repeated multiplication. For example, a mod q = a X a X ….. X a) mod q. To attack Diffie-Hellman, the attacker must determine k given a; this is the discrete log problem.

For elliptic curve cryptography, an operation over elliptic curves, called addition, is used. Multiplication is defined by repeated addition. For example, a * k = (a + a + …. + a), where the addition is performed over an elliptic curve.Cryptanalysis involves determining k given a and (a * k).An elliptic curve is defined by an equation in two variables, with coefficients. For cryptography, the variables and coefficients are restricted toelements in a finite field, which results in the definition of a finite abelian group.

V. Point Implementation using Elliptic Curve

5.1 Adding distinct points P and Q

Suppose [3] that P and Q are two distinct points on an elliptic curve, and the P is not -Q. To add the points P and Q, a line is drawn through the two points.[16]. This line will intersect the elliptic curve in exactly one more point, call -R. The point -R is reflected in the x-axis to the point R. The law for addition in an elliptic curve group is P + Q = R For example

5.2 Elliptic Curve Addition: An Algebraic Approach

Although the previous geometric descriptions of elliptic curves provide an excellent method of illustrating elliptic curve arithmetic, it is not a practical way to implement arithmetic computations. Algebraic formulae are constructed to efficiently compute the geometric arithmetic.

5.3 Adding distinct points P and Q

When P = (xP,yP) and Q = (xQ,yQ) are not negative of each other,

P + Q =R
where s = (yP - yQ) / (xP - xQ)
xR = s2 - xP - xQ and
yR = -yP + s(xP - xR)

Note that s is the slope of the line through P and Q.

5.3 ECDSA Algorithm

ECDSA is a very important application of ECC. As we know, digital signature is an important tool of authentication in E-business. It plays an important rule in data integrity, non-repudiation and anonymity. In some sense, digital signature is more efficient than the traditional way, especially for long messages or documents. Because the traditional way can’t assure that each page is unchanged.

ECDSA was accepted in 1999 as an ANSI standard and in 2000 as IEEE and NIST standards. As digital signatures become more and more important in the’ E-Business, the use of ECDSA will become all pervasive. Elliptic Curve DSA (ECDSA) is a variant of the Digital Signature Algorithm (DSA) which operates on elliptic curve groups. The EC variant provides smaller key sizes for (supposedly) similar security level. On the other hand, the execution time is roughly the same and the signature size is exactly the same: 4*t*, where *t* is the security parameter.

This algorithm is divided into 3 sub algorithm section described below

5.3.1 Key Pair Generation Using ECDSA

Let A be the signatory for a message *M*. Entity A performs the following steps to generate a public and private key:

1. Select an elliptic curve *E* defined over a finite field F_p such that the number of points in $E(F_p)$ is divisible by a large prime *n*.
2. Select a base point, *P*, of order n such that $P \in E(F_p)$
3. Select a unique and unpredictable integer, *d*, in the interval [1, *n*-1]
4. Compute *Q* = *dP*
5. Sender A’s private key is *d*
6. Sender A’s public key is the combination (*E*, *P*, *n*, *Q*)

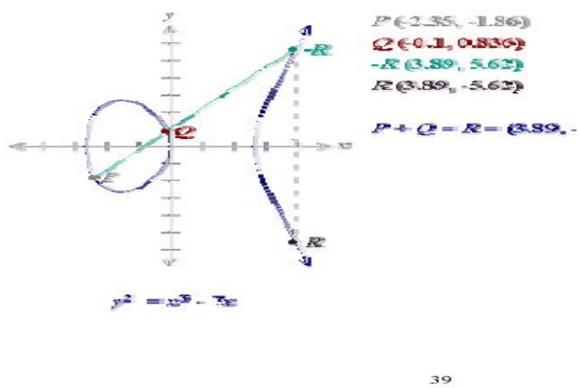
5.3.2 Signature Generation Using ECDSA

Using A’s private key, A generates the signature for message *M* using the following steps:

1. Select a unique and unpredictable integer k in the interval $[1,n-1]$
2. Compute $kP = (x_1,y_1)$, where x_1 is an integer
3. Compute $r = x_1 \bmod n$; If $r = 0$, then go to step 1
4. Compute $h = H(M)$, where H is the Secure Hash Algorithm (SHA-1)
5. Compute $s = k^{-1}\{h + dr\} \bmod n$; If $s = 0$, then go to step1
6. The signature of A for message M is the integer pair (r,s)

5.3.3 *Signature Verification Using ECDSA*

The receiver B can verify the authenticity of A's signature



(r,s) for message M by performing the following:

1. Obtain signatory A's public key (E, P, n, Q)
2. Verify that values r and s are in the interval $[1,n-1]$
3. Compute $w = s^{-1} \bmod p$
4. Compute $h = H(M)$, where H is the same secure hash algorithm used by A
5. Compute $u_1 = hw \bmod n$
6. Compute $u_2 = rw \bmod n$
7. Compute $u_1P + u_2Q = (x_0,y_0)$
8. Compute $v = x_0 \bmod n$
9. The signature for message M is verified only if $v = r$

VI. IMPLEMENTATION ISSUES

6.1 *Generating Elliptic Curve:-*

Whole idea in EC- cryptography is the selection of elliptic curve so that airthmetics operation can be implemented with high precision.

The equation of elliptic curve on pri field Fp where p>3 is given by

$$Y^2 = x^3+ax+b \text{ Mod } P$$
$$\nabla = -16(4a^3+27b^2) \neq 0 .$$
$$J = 1728(4a)^3/\nabla \quad , \text{where } \nabla\neq 0$$

Lemma 1. *Given* $j0 \ \varepsilon$ F*p there is an elliptic curve, E, de.ned over* IF*p such that j*(*E*) = *jo.*
An elliptic curve with a given j-invariant $j0$ is constructed easily. We consider
$j0 \notin \{0, 1728\}$; these special cases are also easily handled. Let $k = j0/(1728 - j0)$, $j0 \ \varepsilon$ Fp then the equation
E: $y2 = x3 + 3kx + 2k$
gives an elliptic curve with j-invariant $j(E)= j0$.

Theorem 1. *Isomorphic elliptic curves have the same j-invariant.*

VII. CONCLUSION

Digital signatures provide cryptographic services that have become a necessity in data and network security. They offer non-repudiation, confidentiality, authentication, and integrity in a format that easily extends to the digital age. As systems become faster and smaller, they will need to maintain their security while using minimal resources. The elliptic curve digital signature algorithm provides the same functionality and security as the standard digital signature algorithm, but delivers it in a smaller key size. This makes it ideal for resource-constrained systems and network technology. Elliptic curves offer major advantages over traditional systems such as increased speed, less memory and smaller key size. Although doing the group operations is slower for an ECC system than for RSA or other discrete log system of the same size, equal security can be provided by much smaller key length using ECC, to this extent that it can be faster than other Verify In addtion, less storage, lesspower and less memory than other systems make it possible to implement cryptography in many special platforms such as wireless devices, laptop Computers and smart cards. So do the situations where efficiency is important.

REFERENCE

[1.] M.Tolga SAKALLI, Ercan BULU and Fatma BUYUKSARACOGLU Cryptography Education for Students Computer Engineering Department, Trakya University, Edirne { tolga. ercanb. fbuyuksaracoglu) @trakya.edu.tr.

[2.] RSA Laboratories. *PKCS #1 v2.1: RSA Encryption Standard.* June 2002, <ftp://ftp.rsasecurity.com/pub/pkcs/pkcs-1/pkcs-1v2-1.pdf

[3.] Johnson, Don B., Alfred J. Menezes, Scott Vanstone. The Elliptic Curve Digital Signature Algorithm (ECDSA). Canada: Certicom.

2001. Available at http://www.comms.scitech.susx.ac.uk/fft/crypto/ecdsa.pdf

[4.] ECC Brainpool. ECC Brainpool Standard Curves and Curve Generation, 2005. Available at http://www.ecc-brainpool.org/ecc-standard.htm, Version 1.0.

[5.] S. KESH, S. RAMANUJAN, AND S. NERUR, "A Framework for Analysing E- Commerce Security" *Information Management and Computer Security*, 2002 Vol.10, no. 4, p 149-158.

[6.] FIPS. 46-3, '"Data Encryption Standard," Federal Information Processing Standard (FIPS), Publication 46-3,National Bureau of Standards, US. Department of Commerce, Washington D.C.. October 25. 1999.

[7.] Wiley & Sons, Inc. New York, 1996.A. Menezes, P.van Oorschot, S. Vanstone, Handbook of [13] ElGamal-Type Signature Schemes in Modular Arithmetic and Galois Fields.

[8.] Hand Book of Cryptography Alfred J. Menezes, Paul C. van Oorschot and Scott A. Vanstone CRC Press

Security Issues in Cloud Computing

Ramandeep Kaur

Department of Electronics and Communication Engineering
Amritsar College of Engineering & Technology, Amritsar, Punjab
raman_johal88@yahoo.com

***Abstract*—Cloud Computing is one of the biggest buzzwords in the computer world these days. It allows resource sharing that includes software, platform and infrastructure by means of virtualization. Virtualization is the core technology behind cloud resource sharing. This environment effort to be dynamic, reliable, and customizable with a guaranteed quality of service. Security is as much of an issue in the cloud as it is anywhere else. Different people share different point of view on cloud computing. Some believe it is unsafe to use cloud. After all, outsourcing means losing significant control over data. But at the same time cloud vendors go out of their way to ensure security. This paper review few major security issues with cloud computing and the existing counter measures to those security challenges in the world of cloud computing.**

I. Introduction

Cloud computing is a pay-per-use model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction.

Typically there are three types of resources that can be provisioned and consumed using cloud: software-as-a-service, platform-as-a-service, and infrastructure-as-a-service.

Cloud computing services themselves fall into three major categories. The first type of cloud computing service is known as Software-as-a-service (SAAS). This service provides capability to the service subscribers to access provider's software applications running on a cloud infrastructure. The service providers manage and control the application. Customer does not have to own the software but instead only pay to use it through a web API. For example, Google Docs relies on JAVA Script, which runs in the Web browser.

The second type of cloud service is called Platform-as-a-service (PAAS). It is another application delivery model.

Lets the consumer to deploy their applications on the providers cloud infrastructure using programming languages and tools supported by the provider. The consumer does not have to manage the underlying cloud infrastructure but has control over the deployed application. A recent example is the Google App Engine, a service that lets developer to write programs to run them on Google's infrastructure.

The third and final type of cloud computing is known as Infrastructure-as-a-service (IaaS). This service basically delivers virtual machine images as a service and the machine can contain whatever the developers want. Instead of purchasing servers, software, data center resources, network equipment, and the expertise to operate them, customers can buy these resources as an outsourced service delivered through the network cloud. The consumer can automatically grow or shrink the number of virtual machines running at any given time to accommodate the changes in their requirement. For example, host firewalls, virtual PC.

There are different kinds of cloud deployment models available. We will discuss three major types of cloud. The first one is Private cloud. This is also known as internal cloud.

Private Cloud refers to on-demand virtualized environments in internally managed data centers. In a private cloud, an organization sets up a virtualization environment on its own servers, either in its own data centers or in those of a managed service provider. This model is driven by concerns about the security and reliability of public clouds.

The second type of cloud is known as Public cloud or External cloud. Public Cloud describes cloud computing in the traditional mainstream sense, whereby computing environment is open for use by the general public and is owned by an organization selling cloud services.

Third type of cloud is known as Hybrid cloud. A hybrid cloud environment consists of multiple internal and /or external providers. It is typical for most enterprises.

This paper is organized as follows. In section 2, we briefly describe the cloud computing architecture. In section 3, we briefly describe the applications of cloud computing. In section 4, we discuss the major security challenges in cloud computing environment and their existing counter measures. In section 5, we discuss the security standards in cloud computing. Finally, in section 6, we conclude.

II Cloud Computing Architecture

Cloud computing system is divided into two sections: the front end and the back end. Theses two ends connect to each other usually through Internet. The front end is the user side and back end is the "cloud" section of the system. The front end includes the client's computer and the application required to access the cloud computing system.

As shown in figure 1, on the back end of the system are the various computers, servers and data storage systems that create the "cloud" of computing services.

A central server administers the system, monitoring traffic and client demands to ensure everything runs smoothly. It follows a set of rules called protocols and uses a special kind of software called middleware.

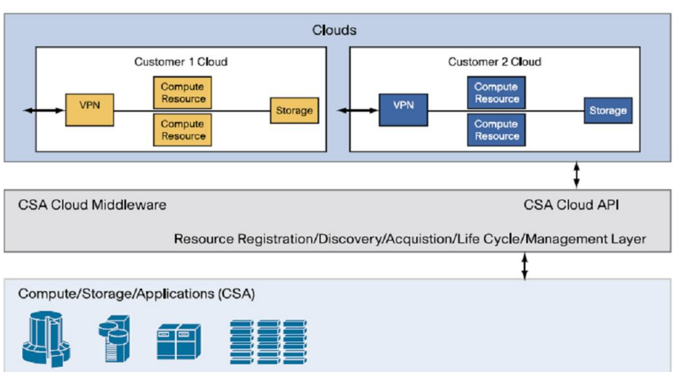


Figure 1. High-Level Cloud Middleware Architecture Example

Cloud middleware also referred to as cloud OS, is the major system that manages and controls services. Middleware allows networked computers to communicate with each other. Google App Engine and Amazon EC2/S3 are examples of cloud middleware.

An Application Programming Interface (APIs) for applications, acquisition of resources such as computing power and storage, and machine image management must be available to make applications suitable for network clouds.

In a simplified vision of the cloud computing architecture, as shown in figure 2. First of all, Client sends service requests. Then system management finds correct resources. After that, system provisioning finds correct resources. After the computing resources are found then the client request is executed. Finally, results of the service requests are sent to the clients.

In the next section we discuss different applications of cloud computing environment and how using cloud can be so beneficial for organizations of all the sizes.

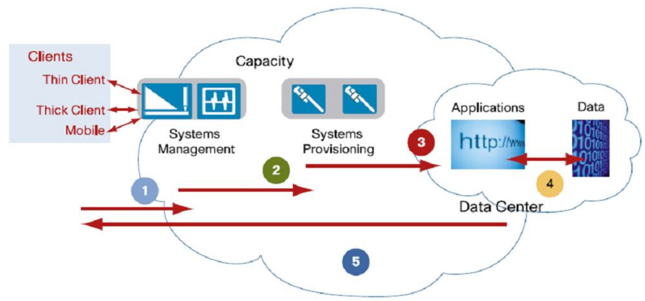


Fig 2. Cloud computing Workflow

III Cloud Computing Applications

The applications of cloud computing are practically limitless. With the right middleware, a cloud computing system can practically run all the applications a personal computer can run.

Clients will be able to access their applications and data at any time from anywhere using any computer linked to Internet. Data and application need not to be confined to a hard drive on client's machines or even on organization's internal network.

Traditionally, Organizations that rely on computers for their operations have to buy all the required software or software licenses for every employee. Cloud computing system gives an option to these organizations to get access to all the required computer applications without even buying those applications. Instead, company can pay a pay-per-use fee to a cloud service provider.

Hardware costs can be minimized. Cloud computing system will reduce the hardware costs on client side. User will not have to buy the computer with most memory, nor he has to buy the large hard drive to store his data. Cloud system will take care of these client's need. Client just have to buy a computer terminal with a monitor, input devices like keyboard and mouse and juts enough processing power to run the middleware necessary to connect to the cloud system.

In most of the companies servers and digital storage devices take up a huge space. Some companies do not have a large physical apace available on-site so they rent space to store their servers and databases. Cloud computing system gives these companies an option to store their data on someone else's (cloud service providers) hardware thus freeing these companies of requirement to have their own physical space on the client side.

Client can make use of cloud system's huge processing power. Like in grid computing, client can send huge complex calculations on cloud for processing. Sometimes complex calculations can take years for individual computer to compute. The cloud system in this case will use the processing power of required number of available computers on the back end to speed up the calculation.

Cloud computing system can also be helpful in reducing costs on IT support. Cloud services are generally effectively organized and simplified and would have fewer problems than the heterogeneous network and operating systems with devices attached form different manufacturers.

Cloud computing offers significant advantage over the traditional computing system but it has its own issues. In the next section we will discuss about few of the major security challenges in the cloud computing environment.

In the next section we discuss about the major security challenges in cloud computing environment and their existing counter measures.

4. Cloud Computing Challenges

Security and privacy are the two major concerns about cloud computing.
In the cloud computing world, the virtual environment lets user access computing power that exceeds that contained within their own physical world. To enter this virtual environment a user is required to transfer data throughout the cloud. Consequently several security concerns arises.

4.1 Information security is concerned with protecting information and information systems from unauthorized access, use, disclosure, disruption, modification or destruction. The goal of information security is to protect the confidentiality, integrity and availability of data regardless of the form the data may take.

4.1.1 Losing control over data

Outsourcing means losing significant control over data. Large banks don't want to run a program delivered in the cloud that risk compromising their data through interaction with some other program.

With any shared storage system, the most common security question is whether unauthorized users can access information either intentionally or by mistake. To ensure that customers have flexibility to determine how, when, and to whom they wish to expose the information they store in Amazon Web Service (AWS) , Amazon Simple Storage Service (S3) APIs provide both bucket- and object level access controls, with defaults that only permit authenticated access by the bucket and/or object creator. Unless a customer grants anonymous access to their data, the first step before a user can access data is to be authenticated using HMAC-Sha1 signature of the request using the user's private key.

An authenticated user can read any object only if the user has been granted Read permissions in an Access Control List (ACL) at the object level. An authenticated user can list the keys and create or overwrite objects in a bucket only if the user has been granted Read and Write permissions in an ACL at the bucket level. Bucket and object level ACLs are independent; an object does not inherit ACLs from its bucket. Permissions to read or modify the bucket or object ACLs are themselves controlled by ACLs that default to creator-only access.

Therefore, the customer maintains full control over who has access to their data. Customers can grant access to their Amazon S3 data to other AWS users by AWS Account ID or email, or Dev-Pay product ID. Customers can also grant access to their Amazon S3 data to all AWS users or to everyone (enabling anonymous access).

4.1.2 Data Integrity

Data integrity means ensuring that data is identically maintained during any operation (such as transfer, storage, or retrieval). Put simply, data integrity is assurance that the data is consistent and correct. Ensuring the integrity of the data really means that it changes only in response to authorized transactions. For example, if the client is responsible for constructing and validating database queries and the server executes them blindly, the intruder will always be able to modify the client-side code to do whatever he has permission to do with the backend database. Usually, that means the intruder can read, change, or delete data at will.

The common standard to ensure data integrity does not yet exists. In this new world of computing users are universally required to accept the underlying premise of trust. In fact, some have conjectured that trust is the biggest concern facing cloud computing.

4.1.3 Risk of Seizure

With the cloud model, you lose control over physical security. In a public cloud, you are sharing computing resources with other companies. In a shared pool outside the enterprise, you do not have any knowledge or control of where the resources run. Exposing your data in an environment shared with other companies could give the government "reasonable cause" to seize your assets because another company has violated the law. Simply because you share the environment in the cloud, may put data at risk of seizure.

The only protection against the risk of seizure for user is to encrypt their data. The subpoena will compel the cloud provider to turn over user's data and any access it might have to that data, but cloud provider won't have user's access or decryption keys. To get at the data, the court will have to come to user and subpoena user. As a result, user will end up with the same level of control user have in his private data center.

4.1.4 Who controls the encryption/decryption keys?

Most customers probably want their data encrypted both ways across the Internet using SSL (Secure Sockets Layer protocol). They also most likely want their data encrypted while it is at rest in the cloud vendor's storage pool. The question arises is if information is encrypted while passing through the cloud who controls the encryption/decryption keys? Is it the customer or the cloud vendor?

Be sure that user control the encryption/decryption keys, just as if the data were still resident on your own servers.

4.1.5 Incompatibility Issue

Storage services provided by one cloud vendor may be incompatible with another vendor's services should you decide to move from one to the other. Vendors are known for creating what the hosting world calls "sticky services" – services that an end user may have difficulty transporting from one cloud vendor to another. For example, Amazon's "Simple Storage Service" [S3] is incompatible with IBM's Blue Cloud, or Google, or Dell.

While cloud computing has the potential to have a positive impact on organizations, there is also potential for lock-in and lost flexibility if appropriate open standards are not identified and adopted. Open Cloud Manifesto establishes a core set of principles to ensure that organizations will have freedom of choice, flexibility, and openness as they take advantage of cloud computing.

Amazon and Microsoft both declined to sign the newly published Open Cloud Manifesto. Amazon and Microsoft pursue interoperability on their own terms.

Amazon CTO Werner Vogels said customers can "mix and match" Amazon Web Services with cloud services from other vendors because Amazon exposes its services as "standard interfaces". Users of Amazon's Elastic Compute Cloud, for example, aren't limited to Amazon's Simple Storage Service as their data-storage option. "You can use EC2 and storage services from another party," Vogels said. " There is no lock-in." He highlighted platform-as-a-service provider Stax Networks-which hosts its service on EC2—as an example of the heterogeneous nature of Amazon's cloud.

Microsoft distinguished engineer Yousef Khalidi gave an update on Microsoft's Azure cloud services, Khalidi said Azure, which is in pre-beta testing, is "designed for interoperability," making use of REST protocols, XML file formats, and support for "all languages".

4.1.6 Constant Feature Additions

Cloud applications undergo constant feature additions, and users must keep up to date with application improvements to be sure they are protected. The speed at which applications will change in the cloud will affect both the SDLC (Software development life cycle) and security.

Emergency, non-routine, and other configuration changes to existing Amazon Web Services (AWS) infrastructure are authorized, logged, tested, approved, and documented in accordance with industry norms for similar systems. Updates to AWS infrastructure are done in such a manner that in the vast majority of cases they do not impact the customer and their Service use.

AWS communicates with customers, either via email, or through the AWS Service Health Dashboard when there is a chance that their Service use may be affected.

4.1.7 Failure in Provider's Security

Failure of cloud provider to properly secure portions of its infrastructure – especially in the maintenance of physical access control – results in the compromise of subscriber systems. To support their cloud's integrity, large providers typically require that users place 100 percent of their data within the provider's cloud. Cloud can comprise multiple entities, and in such a configuration, no cloud can be more secure than its weakest link.

If a cyber criminal can identify the provider whose vulnerabilities are the easiest to exploit, then this entity becomes a highly visible target. The lack of security associated with this single entity threatens the entire cloud in which it resides. If not all cloud providers supply adequate security measures, then these clouds will become high-priority targets for cyber criminals. By their architecture's inherent nature, clouds offer the opportunity for simultaneous attacks to numerous sites, and without proper security, hundreds of sites could be compromised through a single malicious activity.

Nothing guarantees that the cloud provider will, in fact, live up to the standards and processes they profess to support. For example, Amazon's S3 service was recently down for six hours due to an increased volume of unauthenticated calls, which pushed the authentication service over its maximum capacity before Amazon could solve the issue.

It is expected that customer must trust provider's security. For small and medium size businesses provider security may exceed customer security. The provider must protect the underlying infrastructure from break-ins and generally has responsibility for all authentication and encryption. It is generally difficult for the details that help ensure that the right things are being done.

4.1.8 Cloud Provider Goes Down

This scenario has a number of variants: bankruptcy, deciding to take the business in another direction, or a widespread and extended outage. Whatever is going on, subscriber risk losing access to their production system due to the actions of another company. Subscriber also risk that the organization controlling subscriber data might not protect it in accordance with the service levels to which they may have been previously committed.

This was the case with the provider Zimki. The company started in 2006 and by mid-2007 was out of business, causing applications and client data they hosted to be lost.

The only option user have is to chose a second provider and use automated, regular backups, for which many open source and commercial solutions exist, to make sure any current and historical data can be recovered even if user cloud provider were to disappear from the face of the earth.

4.2 Network Security is concerned with protecting the network and the network-accessible resources from unauthorized access, and consistent and continuous monitoring and measurement of its effectiveness combined together. Network security measures are needed to protect data during their transmission, between terminal user and computer and between computer and computer.

4.2.1 Distributed Denial of Service (DDOS) Attack

In DDOS attack servers and networks are brought down by a huge amount of network traffic and users are denied the access to a certain Internet based Service. In a commonly recognized worst-case scenario, attackers use botnets to perform DDOS. In order to stop hackers to stop attacking the network, subscriber or provider face blackmail.

In fact, in Japan, blackmail involving DDOS is on the rise. One major Tokyo firm has to pay 3 million yen (about U.S. $31,000) after the network was brought to a screeching halt by a botnet attack. Because the attack was so dispersed, police have been unable to track down the attackers.

Amazon Web Service (AWS) Application Programming Interface (API) endpoints are hosted on large, Internet-scale, world-class infrastructure that benefits from the same engineering expertise that has built Amazon into the world's largest online retailer. Proprietary DDOS mitigation techniques are used. Additionally, Amazon's networks are multi-homed across a number of providers to achieve Internet access diversity.

4.2.2 Man in the Middle Attack

This attack is a form of active eavesdropping in which the attacker makes independent connections with the victims and relays messages between them, making them believe that they are talking directly to each other over a private connection when in fact the entire conversation is controlled by the attacker.

All of the AWS APIs are available via SSL-protected endpoints which provide server authentication. Amazon EC2 AMIs automatically generate new SSH host certificates on first boot and log them to the instance's console. Customers can then use the secure APIs to call the console and access the host certificates before logging into the instance for the first time. Customers are encouraged to use SSL for all of their interactions with AWS.

4.2.3 IP Spoofing

Spoofing is the creation of TCP/IP packets using somebody else's IP address. Intruder gain unauthorized access to computer, whereby he sends messages to a computer with an IP address indicating that the message is coming from a trusted host. To engage in IP spoofing, a hacker must first use a variety of techniques to find an IP address of a trusted host and then modify the packet headers so that it appears that the packets are coming from the host.

Amazon EC2 instances cannot send spoofed network traffic. The Amazon-controlled, host-based firewall infrastructure will not permit an instance to send traffic with a source IP or MAC address other than its own.

4.2.4 Port Scanning

If the Subscriber configures the security group to allow traffic from any source to a specific port, then that specific port will be vulnerable to a port scan. Since a port is a place where information goes into and out of the computer, port scanning identifies open doors to a computer.

Port scanning has legitimate uses in managing networks, but port scanning also can be malicious in nature if someone is looking for a weakened access point to break into your computer.

There is no way to stop someone from port scanning your computer while you are on the Internet because accessing an Internet server opens a port which opens a door to your computer.

Port scans by Amazon Elastic Compute Cloud (EC2) customers are a violation of the Amazon EC2 Acceptable use Policy (AUP). Violations of the AUP are taken seriously, and every reported violation is investigated. Customers can report suspected abuse. When port scanning is detected it is topped and blocked. Post scans of Amazon EC2 instances are generally ineffective because, by default, all inbound ports on Amazon EC2 instances are closed and are only opened by the customer .

The customer's strict management of security groups can further mitigate the threat of port scans. If the customer configures the security group to allow traffic from any source to a specific port, then that specific port will be vulnerable to a port scan. In these cases, the customer must use appropriate security measures to protect listening services that may be essential to their application from being discovered by an unauthorized port scan. For example, a web server must clearly have port 80 (HTTP) open to the world, and the administrator of the server is responsible for ensuring the security of the HTTP server software, such as Apache.

4.2.5 Packet Sniffing

Packet sniffing by Other Tenants: Packet sniffing is listening (with software) to the raw network device for packets that interest you. When that software sees a packet that fits certain criteria, it logs it to a file. The most common criteria for an interesting packet is one that contains words like “login” or “password”.

It is not possible for a virtual instance running in promiscuous mode to receive or “sniff” traffic that is intended for a different virtual instance. While customers can place their interfaces into promiscuous mode, the hypervisor will not deliver any traffic to them that is not addressed to them.
Even two virtual instances that are owned by the same customer, located on the same physical host, cannot listen to each other's traffic. Attacks such as ARP cache poisoning do not work within Amazon EC2. While Amazon EC2 does provide ample protection against one customer inadvertently or maliciously attempting to view another's data, as a standard practice customers should encrypt sensitive traffic.

4.3 Security issues are more complex in a virtualized environment because you now have to keep track of security on two tiers: the physical host security and the virtual machine security. If the physical host server's security becomes compromised, all of the virtual machines residing on that particular host server are impacted. And a compromised virtual machine might also wreak havoc on the physical host server, which may then have an ill effect on all of the other virtual machines running on that same host.

4.3.1 Instance Isolation

Isolation ensures that different instances running on the same physical machine are isolated from each other.

Virtualization efficiencies in the cloud require virtual machines from multiple organizations to be co-located on the same physical resources. Although traditional data center security still applies in the cloud environment, physical segregation and hardware-based security cannot protect against attacks between virtual machines on the same server. Administrative access is through the Internet rather than the controlled and restricted direct or on-premises connection that is adhered to in the traditional data center model. This increase risk of exposure will require stringent monitoring for changes in system control and access control restriction.

Different instances running on the same physical machine are isolated from each other via Xen hypervisor. Amazon is active in the Xen community, which ensures awareness of the latest developments. In addition, the AWS firewalls reside within the hypervisor layer, between the physical network interface and the instance's virtual interface. All packets must pass through this layer, thus an instance's neighbors have no more access to that instance than any other host in the Internet and can be treated as if they are on separate physical hosts. The physical RAM is separated using similar mechanisms.

4.3.2 Host Operating System

Administrators with a business need to access the management plans are required to us multi-factor authentication to gain access to purpose-built administration hosts. These administrative hosts are systems that are specifically designed, built, configured, and hardened to protect the management plane of the cloud. All such access is logged and audited. When an employee no longer has a business need to access the management plane, the privileges and access to those hosts and relevant systems are revoked.

4.3.3 Guest Operating System

Virtual instances are completely controlled by the customer. Customers have full root access or administrative control over accounts, services, and applications. AWS does not have any access rights to customer instances and cannot log into the guest OS. AWS recommends a base set of security best practices including: customer should disable password-based access to their hosts, and utilize some form of multi-factor authentication to gain access to their instances, or at a minimum certificate-based SSH Version 2 access.

Additionally, customers should employ a privilege escalation mechanism with logging on a per-user basis. For example, if the guest OS is linux, After hardening their instance, they should utilize certificate-based SSHv2 to access the virtual instance, disable remote root login, use command-line logging, and use 'sodu' for privilege escalation. Customers should generate their own key pairs in order to guarantee that hey are unique, and not shared with other customers or with AWS.

AWS Multi-Factor Authentication (AWS MFA) is an additional layer of security that offers enhanced control over AWS account settings. It requires a valid six-digit, single-use code from an authentication device in your physical possession in addition to your standard AWS account credentials before access is granted to an AWS account settings. This is called Multi-Factor Authentication because two factors are checked before access is granted to your account: customer need to provide both their Amazon email-id and password (the first "factor": something you know) AND the precise code from customer authentication device (the second "factor": something you have).

4.4 In addition to the above mentioned issues there are few other general security issues that are delaying cloud computing adoption and needs to be taken care of.

4.4.1 Data Location

When user uses the cloud, user probably won't know exactly where his data is hosted, what country it will be stored in?

Amazon does not even disclose where their data centers are located. They simply clam that ach data center is hosted in a nondescript building with a military-grade perimeter. Even if customer know that their database server is in the us-east-1a availability zone, customer do not know where that data center9s0 behind that availability zone is located, or even which of he three East Coast availability zones us-east-1a represents.

4.4.2 Data Sanitization

Sanitization is the process of removing sensitive information from a storage device. In cloud computing users are always concerned about, what happens to data stored in a cloud computing environment once it has passed its user's "use by date" When a storage device reached the end of its useful life. AWS procedures include a decommissioning process that ensures customer data are not exposed to unauthorized individuals. AWS uses the technique DoD 5220.22-M as per National Industrial Security Program Operating manual to destroy data, as part of the decommissioning process.

When item and attribute data are deleted within a domain, removal of the mapping within the domain starts immediately, and is also generally complete within seconds. Once the mapping is removed, there is no remote access to the deleted data. The storage area is then made available only for write operations and the data are overwritten by newly stored data.

4.4.3 Job Starvation due to some virus or worm

It is where one job takes up a huge amount of resource resulting in a resource starvation for the other jobs.

Customer can reserve the resources in advance. Customer can also reduce the priority of the affected tasks/job.

In the next section of our paper we discuss about the various cloud related working groups and their contribution in the cloud computing environment.

5. Standards for Security in Cloud Computing

Security standards define the processes, procedures, and practices necessary for implementing a security program. These standards also apply to cloud related IT activities and include specific steps that should be taken to ensure a secure environment is maintained that provides privacy and security of confidential information in a cloud environment. Security standards are based on a set of key principles intended to protect this type of trusted environment.

A basic philosophy of security is to have layers of defense, a concept known as defense in depth. This means having overlapping systems designed to provide security even if one system fails. An example is s firewall working in conjunction with intrusion-detection system (IDS). Defense in depth provides security because there is no single point of failure and no single entry vector at which an attack can occur. For this reason, a choice between implementing network security in the middle part of a network (i.e., in the cloud) or at the endpoints is a false dichotomy.

No single security system is a solution by itself, so it is far better to secure all systems. This type of layered security is precisely what we are seeing develop in cloud computing. Traditionally, security was implemented at the endpoints, where the user controlled access. An organization had no choice except to put firewalls, IDSs, and antivirus software inside its own network. Today, with the advent of managed security services offered by cloud providers, additional security can be provided inside the cloud.

5.1 Security Assertion Markup Language (SAML)

SAML is an XML-based standard for communicating authentication, authorization, and attribute information among online partners. It allows businesses to securely send assertions between partner organizations regarding the identity and entitlements of a principal.

SAML standardizes queries for, and responses that contain, user authentication, entitlements, and attribute information in an XML format. This format can then be used to request security information about a principal from a SAML authority. A SMAL authority, sometimes called the asserting party, is a platform or application that can relay security information. The relying party or assertion consumer or requesting party is a partner site that receives the security information. The exchanged information deals with a subject's authentication status, access authorization, and attribute information. A subject is an entity in a particular domain by an email address is a subject, as might be a printer.

SAML is built on a number of existing standards, namely, SOAP, HTTP and XML. SAML relies on HTTP as its communications protocol and specifies the use of SOAP.

5.2 Open Authentication (OAuth)

OAuth is an open protocol, initiated by Blaine Cook and Chris Messina, to allow secure API authorization in a simple, standardized method for various types of web applications. OAuth is a method for publishing and interacting with protected data. For developers, OAuth provides users access to their data while protecting account credentials. It also allows users to grant access to their information, which is shared by the service provider and consumers without sharing all of their identity. OAuth is the baseline, and other extensions and protocols can be built on it.

By design, OAuth Core 1.0 does not provide many desired features, like automated discovery of endpoints, language support, support for XML-RPC and SOAP, standard definition of resource access, OpenID integration, signing algorithms, etc [8]. The core deals with fundamental aspects of the protocol, namely, to establish a mechanism for exchanging a user name and password for a token with defined rights and to provide tools to protect the token. It is important to understand that security and privacy are not guaranteed by the protocol. Infact, OAuth by itself provides no privacy at all an depends on other protocols such as SSL to accomplish that.

5.3 OpenID

OpenID is an open, decentralized standard for user authentication and access control. It allows users to log onto many services using the same digital identity. It is a single-sign-on (SSO) method of access control. OpenID replaces the common log-in process, i.e. a log-in name and a password, by allowing users to log in once and gain access to resources across participating systems. An OpenID is in the form of a unique URL and is authenticated by the entity hosting the OpenID URL.

The OpenID protocol does not rely on a central authority to authenticate a user's identity. Neither the OpenID protocol nor any websites requiring identification can mandate that a specific type of authentication be used; nonstandard forms of authentication such as smart cards, biometrics, or ordinary password are allowed.

6.4 SSL/TLS

Transport Layer Security (TLS) and its predecessor, Secure Sockets Layer (SSL), are cryptographically secure protocols designed to provide security and data integrity for communications over TCP/IP. TLS and SSL encrypt the segments of network connections at the transport layer. The TLS protocol allows client/server applications to communicate across a network in a way specifically designed to prevent eavesdropping, tampering, and message forgery.

TLS provides endpoint authentication and data confidentiality by using cryptography. TLS authentication is one way- the server is authenticated, because the client already knows the server's identity. In this case, the client remains unauthenticated.

TLS also supports a more secure bilateral connection mode whereby both ends of the connection can be assured that they are communicating with whom they believe they are connected. This is known as mutual (assured) authentication. TLS involves three basic steps. The first step deals with peer negotiation for algorithm support. During this phase, the client and server negotiate cipher suites, which determines which ciphers are used. In the next step, key exchange and authentication is decided. During this phase, a decision is made about the key exchange and authentication algorithm to be used, and determine the message authentication codes. The key exchange and authentication algorithms are typically public key algorithms. The finals step is about the symmetric cipher encryption and message encryption. The message authentication codes are made up from cryptographic hash functions. Once these decisions are made, data transfer may begin.

CONCLUSION

The cloud computing phenomenon is generating a lot of interest worldwide because of its lower total cost of ownership, scalability, competitive differentiation, reduced complexity for customers, and faster and easier acquisition of services.

While cloud offers several advantages, people come to the cloud computing topic from different points of view. Some believe the cloud to be an unsafe place. But few people find it safer then their own security provisioning, especially small businesses that do not have resources to ensure the necessary security themselves. Several large financial organizations and some government agencies are still holding back. They indicate that they will not consider moving to cloud anytime soon because they have no good way to quantify their risks.

To gain total acceptance from all potential users, including individuals, small businesses to Fortune 500 firms and government, cloud computing require some standardization in the security environment and third-party certification to ensure that standards are met.

REFERENCES

[1] http://csrc.nist.gov/groups/SNS/cloud-computing/index.html.

[2] Cisco White Paper, http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/white_paper_c11-532553.html, published 2009, pp. 1-6.

[3] John Viega, McAffee, " Cloud Computing and the Common Man," published on the IEEE Journal ON Cloud Computing Security, pp. 106-108, August 2009.

[4] George Reese, "Cloud Application Architectures", First edition, O'Reilly Media, April 2009, ISBN 9780596156367, pp. 2-4, 99-118.

[5] http://en.wikipedia.org/wiki/Cloud_computing.

[6] http://communication.howstuffworks.com/cloud-computing1.htm.

[7] John Harauz, Lori M. Kaufman, Bruce Potter, "Data Security in the World of Cloud Computing," published on the IEEE Journal on Cloud Computing Security, July/August 2009, Vol. 7, No.4, pp. 61-64.

[8] John W. Rittinghouse, James F. Ransome, " Cloud Computing Implementation, Management, and Security", CRC Press, August 17, 2009, ISBN 9781439806807, pp. 147-158, 183-212.

[9] Amazon White Paper, http://aws.amazon.com/about-aws/whats-new/2009/06/08/new-aws-security-center-and-security-whitepaper/ , published June 2009.

[10] Marco Descher, Philip Masser, Thomas Feilhauer, A Min Tjoa, David Huemer, " Retaining Data Control to the Client Infrastructure Clouds", published on the IEEE, 2009 International Conference on Availability, Reliability and Security, pp. 9-15.

[11] David Bernstein, Erik Ludvigson, Krishna Sankar, Steve Diamond, Monique Morrow, "Blueprint for the Intercloud – Protocols and Formats for Cloud Computing Interoperability, submitted to IEEE, 2009 Fourth International Conference on Internet and Web Applications and Services, pp. 328-335.

[12] Liang-Jie Zhang, Qun Zhou, " CCOA: Cloud Computing Open Architecture", published on IEEE, 2009 IEEE International Conference on Web Services, pp. 607-615.

[13] Amazon White Paper, " Introduction to Amazon Virtual Private Cloud", Available: http://aws.amazon.com/about-aws/whats-new/2009/08/26/introducing-amazon-virtual-private-cloud/ , published Aug 26, 2009, pp. 6-8.

[14] Rajkumar Buyya, Chee Shin Yeo, Srikumar Venugopal, "Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities", grid Computing and Distributed Systems and Software Engineering, The University of Melbourne, Australia.

Network Security Issues and Solutions

Harprabdeep singh[1], Navneet Singh Randhawa[2]
Assoc.Prof. Adesh Insititute of Engineering & Technology, Faridkot (Punjab),
Assistant Prof. B.U.C.College Batala(Punjab)
harprabdeeep@yahoo.co.in , navneetrandhawa@yahoo.co.in

***Abstract*- With the evolution in nature and requirement of network security there are many factors contributing to these changes, most important is the shift in focus from so called network level threats such as connection oriented intrusions and Denial of services (DoS) attacks to dynamic, content based threats such as viruses, worms, Trojans, spyware and phishing that can spread quickly and indiscriminately and require sophisticated levels of intelligence to detect. This paper provides an overview of the most common network security threats and its solution which protects you and your organization from threats hackers and ensures that the data travelling across our network is safe.**

Keywords- Audit, Firewall, Security, Vulnerability

I. INTRODUCTION

Computer and network security is new and fast moving technology and as such is still being defined and most probably will always "still defined". Security threats and incidents are arising on daily basis or year by year. As the complexity of network attacks and threats increased so the security measurement is required to protect networks. Securing a modern business network and information technology infrastructure demands protective measures associated to network security vulnerabilities.

Today security problem becomes one of the main problems for computer network and internet developing. However there is no simple way to establish a secure computer network. In fact we cannot find a network in the world which does not have a security holes now-a-days. The infrastructure of cyberspace is vulnerable due to three kind of failure: complexity, accidents and hostile intent. Hundreds of millions of people now appreciate a cyber context for terms like "VIRUSES", "DENIAL OF SERVICES", "PRIVACY", "WORMS" more generally attacks so far have been limited.

Protecting infrastructure systems involves five coupled stages first it's necessary to attempt to deter potential attackers, second if attacked the need is to prevent damage, third since success cannot be guaranteed in prevention next stage is to limit the damage as much as possible, fourth stage having sustained some level of damage from attack the defender must reconstitute the pre-attack state of affairs and last step is for defender to learn from failures. The more specific defense to be discussed may be useful portioned in to two forms:

Passive defense- Essential consists in target hardening. Active defense- In contrast, imposes some risk or penalty on the attacker. Risk or penalty may include identification and exposure investigation and prosecution or pre-emptive counter attacks of various sorts.

II. NETWORK SECURITY AND PREVENTION TECHNIQUES

There are several major drivers that are shaping the new security landscape:

Increasing complexity of networks: Where a network 10 years ago might have consisted of a LAN connected to the Internet through a WAN connection, and maybe a few remote access or site-to-site VPN tunnels, the reality today is much more complex. A common environment today will have multiple access mechanisms into the network, including 802.11 wireless LAN (with myriad Client devices including portable computers, PDAs and Smart Phones), web portals for partners and customers, FTP servers, email servers, end-users using new communication platforms (such as Instant Messaging) and peer-to-peer applications for file-sharing. An example of such a network, and the threats that are present, is illustrated in Fig. 1. In addition, the workforce is becoming more mobile. From telecommuter's who work from a home office to mobile workers who are never in a single location for more than a day, this growing "distributed" model adds a significant amount of risk to the network. To help mitigate these risks, the IT manager must ensure that all remote locations and remote clients are protected with the same level of security as is present in the corporate network. Finally, threats are just as likely to come from inside the local network as they are from the Internet. One trend alone overshadows all others in this regard users are taking their laptops home at night and over the weekend, where they are at increased risk of becoming infected or compromised. When the laptops are brought back into the office, the entire network is at risk since the user entered the network "behind the firewall". This is one of many reasons that an emerging "best practice" in secure network design is to segment the network into separate "security zones" (by physical or logical segmentation) such that attacks can be contained in the event of an outbreak.

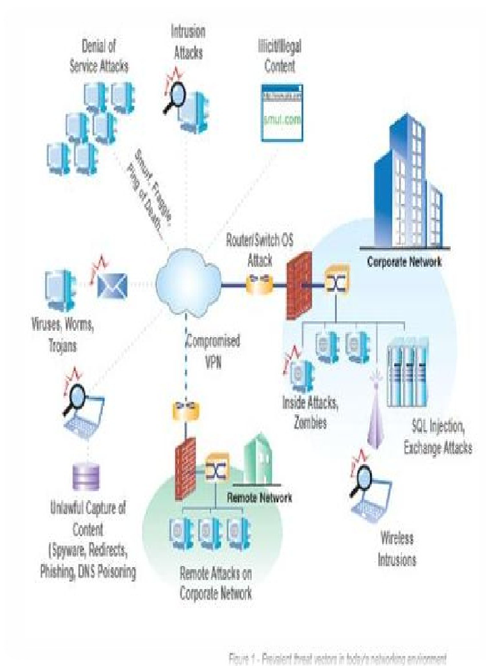


Fig. 1: *Prevalent threat vectors in today's networking environment*

Increasing sophistication of applications and attacks: Applications are growing in complexity. Where Windows NT launched with 5million lines of code in 1994, Windows Vista has over 50 million more than 1000% growth! With this increased complexity comes increased vulnerability, particularly in server systems, which must be patched on a regular basis. While applications are becoming more sophisticated, so are the attacks. A "serious" attack in the early 2000's might have consisted of a simple indiscriminate DoS attack aimed at restricting or temporarily disrupting network access. Today's serious attacks target applications themselves, and in many cases have goals of significant criminal intent, as is demonstrated by the Sasser worm described.

Intrusion Attacks, Worms and Trojans: The "grand-daddy" of them all, the universe of Intrusion attacks is wide and deep. Intrusion attacks are modern threats that target applications and application layer protocols (e.g. using the SMTP protocol to exploit a buffer overflow on an Outlook Exchange server), rather than the networks they are transported on (e.g. DoS attacks that utilize ICMP echo and TCP SYN floods). Examples of common intrusion attacks are Worms, Trojans, web site cross-scripting, SQL injection and tampering, Outlook Exchange server attacks, Apache/IIS buffer overflow attacks, file-path manipulation etc. The Sasser worm, described below, is a classic illustration of an Intrusion attack carried out by a worm. As the Sasser example shows, modern threats are designed to bypass traditional firewalls completely, and instead require an entirely new set of technologies to detect and stop them. An interesting side-note: Sasser also eluded majority of Anti-Virus scanners, which is one example of why AV alone is no longer sufficient protection for Worms and Trojans. As discussed later in this paper, the new technology required to protect against modern threats is Deep Packet Inspection (DPI). DPI gives a security appliance the ability to look not only at the packet headers (like a firewall) but at every bit in the packet payload itself, often across multiple thousands of packets, to detect threats.

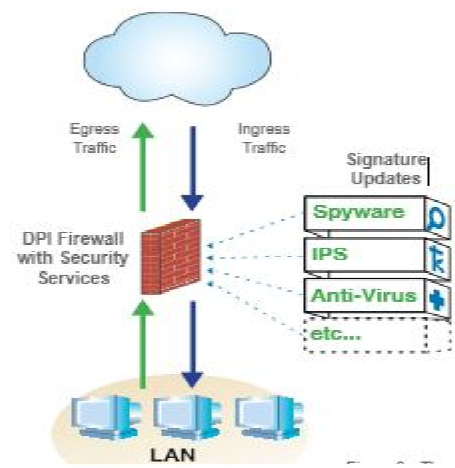


Fig. 2 *The security appliance is now a dynamic system that requires regular signature updates*

Viruses: Viruses (and Worms) are a class of attack whereby an infected attachment or download causes damage to a host system or network. The damage can range from minor (client DoS attack) to catastrophic (full-blown corruption of critical stored information or system registries). A critical trend that is resulting from the increased sophistication of Viruses is the rapidly decreasing "window of infection". In July of 2001, it took the Code Red virus just under 6 hours to infect 359,000clients. Just eighteen months later, the Slammer worm infected 75,000 clients in under 30 minutes. The threats are real… and spread fast. Security vendors have responded by trying to decrease their own "windows of inoculation" which is the time it takes to detect a threat, issue a patch release, and download it to its host systems under management.

There is also a new class of virus-related attack called a 'blended threat'. A blended threat is a 'perfect attack' whereby a virus is accompanied by a number of other attack and intrusion techniques to maximize penetration and damage. A good illustration of this type of attack is the So Big virus detailed below. So Big and Sasser are good examples of how complicated it has become to detect and prevent sophisticated application-layer attacks. To protect against these types of attack, it is mandatory to have IPS and Gateway Antivirus (GAV) installed and activated in the network, whether it is provided by a Deep Packet Inspection.

Financial rewards for hackers with the advent of Spyware and Phishing: The Internet has evolved from being a general information source to a critical enabler of international commerce. Because of the sensitive type of information that now flows freely over the Internet, a new breed of threat aims at obtaining this information sometimes honestly and sometimes with malicious intent. Because the information obtained in these types of attacks has value, hackers are being financially compensated for their work, often by major public corporations; sometimes by organized crime. This is a particularly disturbing trend, since it is attracting the best and the brightest one-time programmers into the black-hat world of hacking and malware generation.

Spyware: Spyware (and Adware) is one of the most misunderstood of the new generation of application-layer threats because there is no consensus on what defines a threat(or more appropriately, what the difference is between 'annoying' Adware and a true threat). There are three general classes of Spyware:

Harmless-but-annoying: Generally consists of actions such as changing the default home page of your browser, or unsolicited/untargeted pop-up ads.

Information-collecting: Cookies are the most common type of information collecting mechanism, but simple keystroke and activity loggers are becoming more common. This class of Spyware is generally interested in collecting basic information about you, the sites you visit, and other preferences so that a 3rd party can send you targeted ads or promotions. There is generally not malicious intent, but many would call this an invasion of privacy.

Malicious: Full keystroke logging and collecting private information with the intent of sending the information to a collection server. The information is collected and sold to 3rd parties who have varying interests. Even today, this type of Spyware can be downloaded instantly on a Client device simply by visiting a URL no further clicking necessary. This type of Spyware is illegal and critical for an organization to detect and stop. To further add to the complexity, there are three major Spyware delivery mechanisms:

Embedded Installs: The most 'honest' of the three mechanisms, embedded installs are typically Spyware/Adware elements that are embedded into programs or services that are downloaded from the web. For example, BigCorp.com might pay a bundling agreement with Claria (Gator eWallet), where they pay Claria $1per client install.

*Drive-by Installs***:** In this method, a banner ad or popup attempts to install software on a PC, usually through the ActiveX controls distributed within Windows and by default enabled in Internet Explorer. Depending on the security settings on the PC browser, the Spyware downloads silently or was downloaded when the user clicked 'Yes' in the installer dialogue box. In many cases, Drive by's also take advantage of browser exploits that can force an unsuspecting PC browser to automatically download and execute code that installs.

Browser Exploit: As described above, targets vulnerabilities in the web browser code to install Spyware. A classic example is the Internet Explorer iFrame vulnerability. Because IE is such a targeted browser, many IT departments are migrating to alternate browsers such as Mozilla's Firefox. This is only putting off the inevitable, however, as every browser that gains in popularity will eventually be the target of Spyware attacks.Spyware is difficult to stop because it requires so many technologies to detect and prevent the exploit. A robust Spyware prevention architecture will consist of both client/server and gateway-based elements. Client and server based Anti-Spyware software will detect and try to prevent users from accessing known bad sites, and to a limited extent provide more advanced functionality to detect suspicious behavior from actual downloads and ActiveX controls. The software will also inspect individual system memory, system registries, start-up files and other stored items to detect and remove Spyware. While necessary, client and server based Anti-Spyware software is not enough. Since Spyware is carried by so many delivery mechanisms and is getting so sophisticated, an additional gateway-based Anti-Spyware element is required. The gateway element not only reinforces URL filtering to prevent access to known bad sites, but provides thorough IPS functionality that detects abnormal behavior from ActiveX Controls and Java Applets and the like, and also provides Anti-virus functionality that inspects attachments for malicious code that installs Spyware. The gateway is also an effective tool for scanning both Instant Messaging (IM) and peer-to-peer protocols/programs, which are a growing target for Spyware and other attacks. Perhaps most importantly, a gateway-based Anti-Spyware solution mitigates the harmful outbound effects of pre-infected client and server devices (that might be attempting to contact a collection server on the Internet to deliver sensitive personal or company data, for instance).

Security as a tool to increase workforce productivity: One of the most profound impacts of security is how it is utilized across all types of organizations to increase operational efficiencies through enhanced work force productivity. There are two main technologies that are helping achieve this:

Web Security and Policy Enforcement: It is no longer a secret that a good amount of an average employee's day can be spent online doing non-work-related activities. Web surfing, online shopping, online gambling, stock trading and even online dating are a few of the more common uses of company Internet resources.

In what many employees might consider a breach of privacy, the company employing URL filtering technology can monitor and report on individual Internet usage, and can also set scheduled restrictions on what types of sites employees are allowed to access throughout the day. If the company is using this type of technology, eSoft highly recommends that the HR department make public notice that this technology is being used, and also clearly state (in the employee handbook, for example) the rules and restrictions of employee Internet usage. The figure above shows a typical screen end user will see when they are trying to access a site that was banned by an IT department employing eSoft Site Filter technology, described later in this document.URL filtering is also a necessary tool for reducing liability that stems from illegal and unethical use of the Internet in public places or organizations. A classic example of this is where an employee (or Internet café patron, for that matter) is accessing a porn site, and another person walks by, witnesses the activity, and uses the company for emotional distress or a hostile work environment. Libraries and schools, by their very nature, MUST have this type of technology deployed. In addition to workforce productivity and liability protection, URL Filtering technology is also the first line of defense at preventing users from accessing Spyware sites. As noted in the previous section, however, Spyware is a much more complicated problem than URL filtering alone can handle.

Spam: Spam has grown into a major problem for all companies and organizations. Spam is especially problematic for public email addresses (listed on a website, for instance), or for common email addresses (support@your_company.com). Spam is also the primary delivery mechanism for Phishing attacks, so its importance has grown over the years. In 2006, over 86% of all e-mail was classified as spam. Over 63% of this spam originates from new or unknown sources. Spam is best dealt with at the security gateway. The reason for this is simple…once Spam emails are inside the network, they are already consuming precious network resources (such as storage, bandwidth and mail server CPU cycles).If prevented before they ever get to a mail server, Spam can become a more manageable nuisance and threat. Another reason Spam is best dealt with at the security gateway is the sophistication of the tools and techniques that are possible to implement at the gateway. Technologies such as word filtering, Bayesian filtering, black and white lists, real-time blackhole lists (RBLs), DNS MX record lookups, reverse DNS lookups, sender policy framework (SFP) compliance and other techniques are all mandatory for effective Spam mitigation. A good gateway Spam filter will reject Spam in such a way that the Spammer will eventually remove the target from their Spam list. For many technologies such as Bayesian filtering, it is necessary to have many, many samples of known spam, and known ham (non-spam) to begin the heuristic process of self-learning. This is another advantage of Anti-Spam technology at the gateway, where there is visibility into every email coming into or exiting the network.

III. FOCUS ON SECURITY

The network security program emphasizes to secure a network. The following background formation in security helps in making correct decisions some areas are concept oriented:

*Attack reorganization: Recognize common attacks such as spoofing denial of service, buffer overflow, etc.

*Encryption techniques: Understand techniques to ensure confidentiality, authenticity, integrity. There must be understood as protocol and at least partial at mathematics or algorithm level in order to select and implement the algorithm matching the organizations need.

*Network security Architecture: Configure a network with security appliances and software such as placement firewalls, intrusion detection system and log management.

To secure a network certain skills must also be practiced:

Protocol analysis: Recognize normal from abnormal protocol sequences using sniffers. Protocols minimal include IP, ARP, ICMP, TCP, UDP, HTTP and encryption protocols SSH, SSL, IP

ACLs (Access Control Lists): Configure and audit routers and firewalls to filter packets accurate and efficient by dropping ,passing or protecting packets based upon their IP and/or port addresses and state. Intrusion detection/Prevention systems

(IDS/IPS): Set and test rules to recognize and report attacks in timely manner

Vulnerability Testing: Test all nodes (routers, server and clients) to determine active application via scanning or other vulnerability test tools and interpret results.

Application Software Protection: Program and test secure software to avoid backdoor entry via SQL injections, buffer overflow etc.

Security Evaluation: Use risk analysis to determine what should be protected and at what cost.

Security Planning: Prepare an audit Plan and report.

Legal Response: Understanding and interpreting the law regarding responding to computer/network attacks corporate responsibility and computer forensics.

DoS Attacks: DoS attacks today are part of every Internet user's life. They are happening all the time and the entire internet user as community has some part in creating them suffering from them or even loosing time and money because of them. DoS attacks don't having anything to do with breaking into computers taking controls over remote hosts over the internet or stealing privileged information like credit card numbers using the internet way of speaking DoS is neither a hack nor a crack . The sole purpose of DoS attacks is to disrupt the services on the internet. Dos attacks are real vandalism against internet services.

Some solution to DoS Attacks: The way DoS and DDoS attacks are perpetrated, by exploiting limitations of protocols and applications is one of the main factor why they are continuous evolving and because of presenting new challenges on how to combat or limit their effects. Even if all of these attacks cannot be completed avoided some basic rules can be followed to protect the network against some and to limit to extent of attack:

i. Make sure the network has firewall up that aggressively keeps everything out except legal traffic.
ii. Implement router filter this lessons the exposure to certain denial of service attacks.
iii. Install patches to guard against TCP/IP attacks. This will substantially reduce the exposure to these attacks but may not eliminate the risk entire.

Observe the system performance and establish baselines for ordinary activity. Use the baseline for ordinary activity. Use the baseline to gauge unusual levels of disk activity, CPU usage or network traffic

IV. CYBERSPACE IS VULNERABLE

The infrastructure of cyberspace is vulnerable due to three kind of failure: complexity, accident, and hostile intent. Very little of it was designed or implemented with assurance or security as primary considerations. Bad things can be done either via network infrastructure or to the infrastructures themselves. These bad things can be characterized by a lot of words destroy, damage, deny, disable, disrupt, distort, degrade, delay and disconnect. We lack a comprehensive understanding of these vulnerabilities.

Absolute defense against cyber attack has rare if ever been achieved in large complex, geographically distributed networks. The complexity of such systems and modes of attack are such that we don't know precisely how to assess and secure them are and this lack of understanding forces defenders to protect themselves in overlapping and in multiple stages. Risk or penalty may include identification and exposure investigation. There will be other tradeoffs, e.g. between detailed and potential cost of individual transaction and waiting to identify and punish attackers over the longer term.

Government will pursue policies that focus on longer term, aspects of protection, seeking to reduce cumulative losses, protecting economies and national security and maintaining law and order.

Protecting Network Boundaries with Firewalls: A Firewall is mechanism by which a controlled barrier is used to control network and out of an organizational internet. Firewall is basically application specific routers. The firewall process can be tight control what is allowed to traverse from one side to another side. Firewalls can range from fairly simple to very complex.

As with most aspects of security, deciding what type of firewall to use will depend upon factors such as traffic levels services needing protection and complexity of rules required. The greater number of services that must be able to traverse the firewall the more complex the requirement become. The difficult for firewalls is distinguishing becomes are distinguishing between legitimate and illegitimate traffic. What do firewall protect against and what protection do the not provide?

If firewalls configured correctly they can be reasonable form of protection from external threats including some denial of services attacks. If not configured correctly then it can be major security holes in organizations.

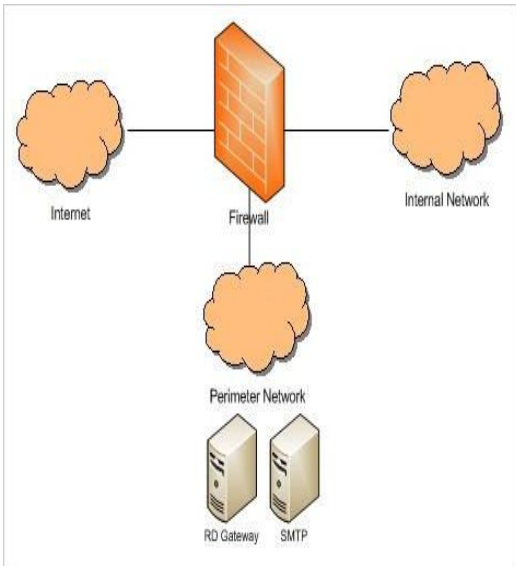


Fig. 3

VI. PREVENTING AN ATACK

There are at least three ways to prevent an attack and all three are ultimately formed of active defense. One is to deter the attacker by having demonstrated capabilities to punish the attackers. This implies that attacker understand the risk of being identified and located that the defender is seen as credible in a resolve to punish and that the cost of punishing is acceptable to defender. A simple situation is when the attacker suffers a large front end loss through discovering during the probe phase and defender can accomplish that discover cheap. When the cost to develop ways of discovering attackers. But the more common situation is when the relatively high costs of legal persecution of single attacker are returned in reduced losses over the longer term.

Second way is to prevent an attack is through establishing cyber attacks as unacceptable behavior among the community of nation.

Third way to prevent an attack is to pre-empt the attacker in ways that the result in abandoning the attack. This implies a great deal by way of national surveillance capabilities to be able to provide strategic warning. So stealth are cyber attacks so widespread is the ability to plan and launch them, so inexpensive are the tools of attacks that pre-emption would not appear to be a practical option at this point. But Should responsible norms of behavior in cyberspace becomes better established the detection and identification of abnormal behavior may become easier.

V. THWARTING AN ATTACK

While preventing attack is largely based on government authority and responsibility the detailed knowledge need to thwart an attack on cyber system to prevent damage rests primarily with its owner. The least complicated case is where the system owner acts individually. Not only must the owner be concerned with defense from outsider, but also need to be recognized that not all authorized users of system may have the owner's interests at heart. There are many ways of defending system against cyber attack and some minimal number must probably employed for the owner to demonstrate due diligence.

This technique such as requiring authorization to center, monitoring the use of system to detect unauthorized activities, periodic checking on the integrity of critical software and establishment and enforcing policies governing System security and responses to unexpected event will be necessary. Owner can limit authorization activities through compartmenting information within the system and maintaining need to know discipline. Owner can provide themselves substantially more rights to monitor inside user by covering access through contractual terms with employees and vendors.

VI. RECONSITUTING AFTER AN ATTACK

Short term recognizaton is set of first steps taken to meet the most urgent threats to life and property. They include assessing damage and implementing an appropriate cover plan. System is restored from backup where possible and residual resources may have to ration. It's possible that additional capacity can be generated as facilities that are idle or in maintenance are brought on line. Online status reporting dispatching of emergency personally and repair equipment notification of users possible lost transaction, an ability to adjust plans in near real time and procedure for secure emergency communication will be required.

VII. HALTING CYBER ATTACKS IN PROGESS

Along with the sharing of information system administrator also need procedure the can use to assist in ending attack already under ways. This need is particularly evident in DoS attacks, which can be of extended duration and which can shut down business operations while they occur. To aid in ending an attack system administrators would profit by working with infrastructure operators to trace the attack to its source and then to block the attacker. Methods for halting attacks in progress as well as those for investigating attacks are constrained by the inability to easily identify and locate attackers. In the case of internet because packet source addresses are easily forged the only way to identify and locate attackers with confidence is to trace the path by packet through the routing infrastructure.

VIII. CONCLUSION

The security issues in our networked systems as described in this paper identify some of the work to be done and the urgency with which concern need to be addressed. Dependence on some of the IT-based infrastructures in several countries is such that serious national consequences could result from the exploration of their vulnerabilities and as the density of network increases the necessity for transnational participation in improving network security increases. The changing technologies and the potential for changing threats is taxing our understanding of threats and how to deal with them. Due to complexity and entanglement among networks and communities internationally an increase in network security must involve the concentrated efforts of as a many nation as possible.

REFERENCES

[1.] "Google Query-serving Architecture" at national conference sponsored by NACC by dixit @ DR.R.K.DIXIT

[2.] Intrusion Control in computer network; Published in National seminar sponsored by higher education of M.P.

[3.] www.abanet.org/scitech/computercrime/cybercrimeobject.html

[4.] Batista E; IDC; Tech Bucks; Hack threats Up Wired news; 23 December 2002

[5.] Council of Europe, convention on cyber crime; ETS no: 185-Explanator report (Article 2 section 2); 23 November 2001: http://conventions.coe.int/treat/en/reports/html/185.html

A Review study of Information Security: Cryptography

Shaweta mahajan[1], Pawanpreet Kaur[2],and Tarun singh[3]
[1]Lecturer in Govt. Polytechnic College, Bhikhiwind.
[2] Lecturer in Lord Krishna Polytechnic College,Subhanpura.
[3] Assistant Professor in Baba Kuma Singh Engg College, Amritsar.
Email: mahajan.shaweta@yahoo.in, pawan.preet11@yahoo.co.in,tarun23.singh@yahoo.co.in

***Abstract-* As the network of the national infrastructure is extended and spreads, and information is widely exchanged and shared, illegal access to information is becoming a serious problem. To address this issue, many nations around the world are researching and developing various techniques and information security policies as a government-wide effort to protect their infrastructures from newly emerging threats. There are many aspects to security. One important aspect for secure communication is that of cryptography. In this paper we discuss the various cryptographic techniques which provides the security to information. Cryptography examines basic cryptography mechanism that are used to disguise information from everyone except who are permitted to see it. It includes shared key and public key cryptography and some of modern algorithms that are used to protect sensitive information.**

***Index terms-* security, cryptography, integrity, confidentiality, non repudation, DES, RSA, DSA.**

I. INTRODUCTION

Information security is the process by which an organization protects and secures its systems, media and facilities that process and maintain information vital to its operation. Following are a few definitions of information security:The process of protecting the availability, privacy, and integrity of information [Wise Geek]. Proper use of data and controls to prohibit accidental or unauthorized use, destruction, or modification of information assets [Peltier 2001]. The process of protecting the confidentiality,integrity and availability of information [Bishop 2003]. “A well-informed sense of assurance that information risks and controls are in balance.” [Anderson 2003. p. 310]. We prefer a more holistic view of information security, incorporating technology, processes, and people [Baskerville 1993; Straub and Welke 1998; Dhillon and Torkzadeh 2006; De Veiga and Eloff 2007]. As such, information security cannot be defined in one sentence.[1] Information security (InfoSec), as defined by the standards published by the Committee on National Security Systems (CNSS), formerly the National Security Telecommunications and Information Systems Security Committee (NSTISSC),[2] is the protection of information and its critical elements, including the systems and hardware that use, store, and transmit that information. The main purpose of information security is to protect the information from unauthorized access.

II. SECURITY OBJECTIVES

Confidentiality, integrity and availability are the three primary security objectives cited in federal law regarding IT security. The Federal Information security management Act of 2002(FISMA) defines “information security” to mean:“Protecting information and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction in order to provide:

- *Confidentiality*: preserving authorized restrictions on access and disclosure, including means for protecting personal privacy and proprietary information. The assigned level of confidentiality is used in determining the types of security measures required for its protection from unauthorized access or disclosure.
- *Integrity:* guarding against improper information modification or destruction, and may include ensuring information and authenticity.The level of impact of unauthorized modification or destruction of information resources describes the importance for maintaining the integrity of a Resource.
- *Availability:* ensuring timely and reliable access to and use of information. The overall importance of availability of a Resource is based on its criticality to the functional operation of a Campus or department or to the priority of that function in continuity plans and disaster recovery strategies. Emergency management planning must take into account the availability requirements of a particular Resource to determine its inclusion in emergency and disaster recovery planning.
- *Non-repudiation:* a service which prevents an entity from denying previous commitments or actions. When disputes arise due to an entity denying that certain actions were taken, a means to resolve the situation is necessary. For example, one entity may authorize the purchase of property by another entity and later deny such autho-rization was granted. A procedure involving a trusted third party is needed to resolve the dispute.

III. CRYPTOGRAPHY

Cryptography is probably the most important aspect of communication security and is becoming increasingly important as a basic building block for computer security [4]. The basic terminology is that cryptography refers to the science and art of designing ciphers; cryptanalysis to the science and art of breaking them; while cryptology, often shortened to just crypto , is the study of both [5]. Cryptography (from Greek kryptos, “hidden”, and graphein, “to write”) is, traditionally, the study of means of converting information from its normal, comprehensible form into an in comprehensible format, rendering it unreadable without secret knowledge, the art of encryption. The art of protecting information (plain text) by transforming it (encrypting it) into an unreadable format is called cipher text. Only those who possess a secret key can dicipher(or decrypt) the message into plain text. Encrypted messages can sometimes be broken by cryptanalysis, also called code breaking, although modern cryptography techniques are virtually unbreakable. Cryptography encrypts the actual message that is being sent. This security mechanism uses mathematical schemes and algorithms to scramble data into unreadable text. It can only be decoded or decrypted by the party that possesses the associated key [6].

Purpose of cryptography:

Cryptography is the science of writing in secret code and is an ancient art; the first documented use of cryptography in writing dates back to circa 1900 B.C. when an Egyptian scribe used non-standard hieroglyphs in an inscription. Some experts argue that cryptography appeared spontaneously sometime after writing was invented, with applications ranging from diplomatic missives to war-time battle plans. It is no surprise, then, that new forms of cryptography came soon after the widespread development of computer communications. In data and telecommunications, cryptography is necessary when communicating over any untrusted medium, which includes just about *any* network, particularly the Internet.[8]

Types of Cryptographic Algorithm:

There are several ways of classifying cryptographic algorithms. For purposes of this paper, they will be categorized based on the number of keys that are employed for encryption and decryption, and further defined by their application and use. The three types of algorithms that will be discussed are:

- Secret Key Cryptography (SKC): Uses a single key for both encryption and decryption
- Public Key Cryptography (PKC): Uses one key for encryption and another for decryption
- Hash Functions: Uses a mathematical transformation to irreversibly "encrypt" information

A. *Symmetric Key Cryptography:*

Symmetric Key Cryptography often known as Conventional Cryptography or secret key cryptography. Symmetric key cryptography is the method where the plaintext is converted to cipher text based on the unique key and the function used. The actual strength of the symmetric key cryptography lies in choosing the non-reversible function that uses the key to produce the cipher text. Like a function Ek(P)=C, Where E is the function, k is the key, P is the plain text and C is the cipher text produced. Then again use the function Dk(C)= P, Where D is the function used to decrypt the message using above components. It is possible to produce the function E=D, and such functions uses the total key length as strength. Though 64 bits may offer a very good security, a 128 bit or 256 bit will definitely use for any mission critical applications. It also aids in improving security because of the increase in the key space, which makes it less prone to brute-force key search and other kind of attacks like plaintext attack, choose plaintext attack, differential plaintext attack etc.

- Stream ciphers are the algorithms that uses the function to encrypt the plaintext in a stream more or less like reading a file character by character and encoding it. Stream ciphers operate on a single bit (byte or computer word) at a time and implement some form of feedback mechanism so that the key is constantly changing. Popular stream ciphers are *vernam ciphers* or also known as one time pad.
- Block ciphers are used to encrypt the plaintext in a pre-defined size of blocks. Say 32bytes, 64 bytes or 128 bytes. It is done to achieve speed. A block cipher is so-called because the scheme encrypts one block of data at a time using the same key on each block. Popular block ciphers are DES, Blowfish etc,

The various attack methods used in the cryptanalysis against symmetric key cryptography are differential cryptanalysis, linear cryptanalysis and algebraic attacks.[7]

Secret key cryptography algorithms that are in use today include:

- *Data Encryption Standard (DES):* The most common SKC scheme used today, DES was designed by IBM in the 1970s and adopted by the National Bureau of Standards (NBS) [now the National Institute for Standards and Technology (NIST)] in 1977 for commercial and unclassified government applications. DES is a block-cipher employing a 56-bit key that operates on 64-bit blocks. DES has a complex set of rules and transformations that were designed specifically to yield fast hardware implementations and slow software implementations, although this latter point is becoming less significant today since the speed of computer processors is several orders of magnitude faster today than twenty years ago. IBM also proposed a 112-bit key for DES, which was rejected at the time by the government; the use of 112-bit keys was considered in the 1990s, however, conversion was never seriously considered.

DES is defined in American National Standard X3.92 and three Federal *Information Processing Standards(FIPS):*

FIPS 46-3: DES

FIPS 74: Guidelines for Implementing and Using the NBS Data Encryption Standard

FIPS 81: DES Modes of Operation

Blowfish: A symmetric 64-bit block cipher invented by Bruce Schneier; optimized for 32-bit Processors with large data caches, it is significantly faster than DES on a Pentium/PowerPC-class machine. Key lengths can vary from 32 to 448 bits in length. Blowfish, available freely and intended as a substitute for DES or IDEA, is in use in over 80 products.[8]

B. Public Key Cryptography:

As the name indicates, uses two kinds of functions and two different keys. The keys are terms as the private key and public key. The public key is the one which is kept 'visible', (i.e.), commonly transmitted over networks, etc. and the other one which is kept *'secret'* the private key, which is never revealed to anybody. Thus the system, uses both keys, it's commonly know as 'public/secret pair algorithms' or 'private/public key algorithms' though the name public key cryptography exists. The functions used in the public key cryptography are like the following.

Consider the function E, EK(P) =C which uses the K, the public key to encrypt plaintext (P) to produce cipher text (C) can only be decrypted using another function D, DS(C)=P which uses S, which is the private key. In some, (but not all) public/private algorithms, you can also use S to encrypt the plaintext that can be later decrypted using K, which is in use by some programs. The major advantage of the public/private pair cryptography is that is solves the issue of key distribution. But since, the functioning is concerned, the pair of functions deployed makes it less secure and there exists a list of attacks against the system. So, even the key strength of 768 bits pose a question threat to information security. Normally a key strength of 1024 and above is secure. But there are programs, which give about 4096 bits of key size which is an uncommon, but providing with a very good security.

Public-key algorithms are slower compared to the speed of symmetric key algorithms or secret-key algorithms. Also, public-key algorithms are prone to attack than the secret-key algorithms. Another main use of public-key algorithms is to produce digital signatures, which plays the lead role in identifying the origin of the message. The *digital signature* thus provides the way to authenticate which is known as message non-repudiation as explained before. Public-key cryptography may be vulnerable to impersonation, but it is the sole responsibility of the user to protect his/her private key securely.[7]

Public-key cryptography algorithms that are in use today for key exchange or digital signatures include:

- *RSA:* The first, and still most common, PKC implementation, named for the three MIT mathematicians who developed it — Ronald Rivest, Adi Shamir, and Leonard Adleman. RSA today is used in hundreds of software products and can be used for key exchange, digital signatures, or encryption of small blocks of data. RSA uses a variable size encryption block and a variable size key. The key-pair is derived from a very large number, *n*, that is the product of two prime numbers chosen according to special rules; these primes may be 100 or more digits in length each, yielding an *n* with roughly twice as many digits as the prime factors. The public key information includes *n* and a derivative of one of the factors of *n*; an attacker cannot determine the prime factors of *n* (and, therefore, the private key) from this information alone and that is what makes the RSA algorithm so secure. Regardless, one presumed protection of RSA is that users can easily increase the key size to always stay ahead of the computer processing curve. As an aside, the patent for RSA expired in September 2000 which does not appear to have affected RSA's popularity one way or the other.

- *Digital Signature Algorithm (DSA):* The algorithm specified in NIST's Digital Signature Standard (DSS), provides digital signature capability for the authentication of messages.

- *Elliptic Curve Cryptography (ECC):* A PKC algorithm based upon elliptic curves. ECC can offer levels of security with small keys comparable to RSA and other PKC methods. It was designed for devices with limited compute power and/or memory, such as smartcards and PDAs.[8]

C. Hash Function:

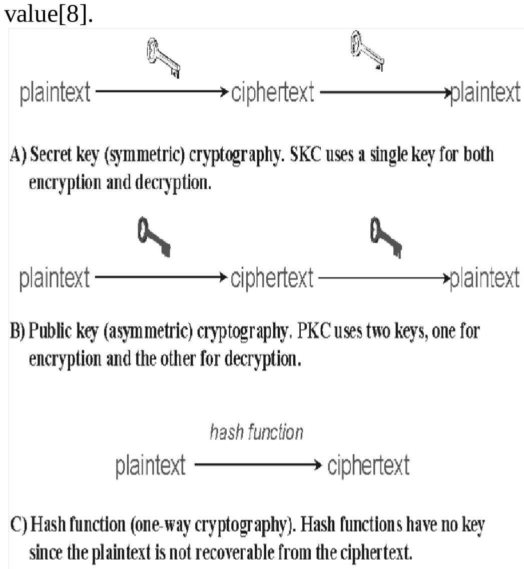
Hash functions, also called message digests and one-way encryption, are algorithms that, in some sense, use no key (Figure 1C). Instead, a fixed-length hash value is computed based upon the plaintext that makes it impossible for either the contents or length of the plaintext to be recovered. Hash algorithms are typically used to provide a *digital fingerprint* of a file's contents, often used to ensure that the file has not been altered by an intruder or virus. Hash functions are also commonly employed by many operating systems to encrypt passwords. Hash functions, then, provide a measure of the integrity of a file. It means to produce a very small value, say H1 from a function M for the plaintext P1. It should be hard to find any other plaintext P2 , which satisfies M(P2)=H2, which gives H1=H2 proves that the plaintext P1=P2.

Example, function M, M(PX)=HX, Where X stands for the different plaintexts, that produces the hashed values. The function M is chosen, such that it is impossible to reverse the function to produce the plaintext from the hashed value (also called as *'hash'* or *'checksum'*).The common use of the hashing is to check the validity of the password. The hash thus produced should be greater that 128 bits to prevent a malicious user to break it. The only know attack against hash is the brute force attack which is almost impossible because of the combination of the plaintext (actually the password) chosen. So by choosing a complicated password with the mixed symbols and alphanumeric characters, adds strength to the hashing algorithm by increasing the available key space to search for the key. The commonly used algorithms are MD4, MD5, SHA and SHA1. As MD stands for Message Digest and SHA stands for Secure Hash Algorithm.

Hash algorithms that are in common use today include:

- *Message Digest (MD) algorithms:* A series of byte-oriented algorithms that produce a 128-bit hash value from an arbitrary-length message.
- *MD2 (RFC 1319):* Designed for systems with limited memory, such as smart cards.
- *MD4 (RFC 1320):* Developed by Rivest, similar to MD2 but designed specifically for fast processing in software.
- *MD5 (RFC 1321):* Also developed by Rivest after potential weaknesses were reported in MD4; this scheme is similar to MD4 but is slower because more manipulation is made to the original data. MD5 has been implemented in a large number of products although several weaknesses in the algorithm were demonstrated by German cryptographer Hans Dobbertin in 1996.
- *Secure Hash Algorithm (SHA):* Algorithm for NIST's Secure Hash Standard (SHS). SHA-1 produces a 160-bit hash value and was originally published as FIPS 180-1 and RFC 3174. FIPS 180-2 describes five algorithms in the SHS: SHA-1 plus SHA-224, SHA-256, SHA-384, and SHA-512 which can produce hash values that are 224, 256, 384, or 512 bits in length, respectively. SHA-224, -256, -384, and -52 are also described in RFC 4634.

Hash functions are sometimes misunderstood and some sources claim that no two files can have the same hash value. This is, in fact, not correct. Consider a hash function that provides a 128-bit hash value. There are, obviously, 2128 possible hash values. But there are a lot more than 2128 *possible* files. Therefore, there have to be multiple files — in fact, there have to be an infinite number of files! — that can have the same 128-bit hash value[8].



A) Secret key (symmetric) cryptography. SKC uses a single key for both encryption and decryption.

B) Public key (asymmetric) cryptography. PKC uses two keys, one for encryption and the other for decryption.

C) Hash function (one-way cryptography). Hash functions have no key since the plaintext is not recoverable from the ciphertext.

Latest cryptographic Advancements

Due to the modern methods of attacks and advancement in the computational speeds, have taken over the well known secure secret-key algorithm, DES, Digital Encryption Standard, which dominated the world of cryptography for decades. A 40 bit version of the DES algorithm can be cracked by any available modern computer and a 56 bit can be broken by any cheaply available Super Computer or by a simple beo-wulf cluster. That pose a security threat and people have developed, developing various other algorithms like 3DES also known as Triple DES etc,. The NIST and NSA plays an important role in deciding the algorithms. They analyse and recommend the algorithm for usage. The latest trends have opened for the AES Advanced Encryptions Standards and many algorithms have been submitted for the review. Some of them approved and under analysis. They are IDEA, RC5, RC6 Rjindael, MARS, Two-fish, Blowfish, Diamond, Sapphire etc[7].

IV. CONCLUSION

The real power can be brought out by using the combination of secret-key and public-key cryptosystems. Like using a secret-key algorithm to encrypt the data/file/message and then signing it with the public-key algorithm can be the highest possible security that can be given to a data/file/message. Also compressing the data/file/message with/without encryption (which uses some patented symmetric key algorithms!) will be more secure because the integrity is maintained to the lowest possible level.

REFERENCES

[1.] Zafar, Humayun and Clark, Jan Guynes (2009) " Current State of Information Security Research In IS," Communication of the Association for Information Systems : vol.24, Art.34. Available at; http://aisel.aisnet.org/cais/vol24/iss1/34.

[2.] National Security Telecommunication and Information Systems Security, National Training Standard for Information Systems Security (Infosec) Professional, 20 june 1994, file, 4011. Accessed 8 feb 2007 from www.cnss.gov/assets/pdf/nstissi_4011.pdf.

[3.] Computer at Risk : Safe Computing in the Information Age, National Research council, 1991.

[4.] Security Engineering: A guide to building Dependable Distributed System.

[5.] William Stalling, Cryptography and Network Security, Principles and Practice, Third eddition, Pearson education, Singapore,2003.

[6.] Cryptography: the myth, avaliable at http://www.comms.engg.susx.ac.uk/fft/crypto/cryptography.pdf

[7.] Gary C.kessler" An Overview Of Cryptography"availableat: http://www.garykessler.net/library/crypto.htm.

Security Advances in Computer Networks: A Review

Er.Simranjit Singh,
Lecturer, CE/IT Department, Baba Banda Singh Bahadhur Polytechnic College Fatehgarh Sahib
Simran88204@gmail.com

Abstract: **-Security is the primary issue in today's computer era. Different types of security measures are applied for the same. This review paper highlights different threats to computer networks and the ways to improve the security of different networks. At the end, it concludes the paper by reviewing the importance of network security**.

I. INTRODUCTION

The need for network security is a relatively new requirement. Prior to the 1980s most computers were not networked. It was not due to lack of desire to network them; it was more a result of the lack of technology. Most systems were mainframes or midrange systems that were centrally controlled and administered. Users interfaced with the mainframe through "dumb" terminals. The terminals had limited capabilities. Terminals actually required a physical connection on a dedicated port. The ports were often serial connections that utilized the RS-232 protocol. It usually required one port for one terminal. IBM, Digital Equipment, and other computer manufacturers developed variations on this architecture by utilizing terminal servers, but the basic concept was the same. There was nothing equivalent to what we experience today where hundreds if not thousands of connections can reach a system on a single network circuit.

In the 1980s, the combination of the development of the personal computer (PC), the development of network protocol standards, the decrease in the cost of hardware, and the development of new applications made networking a much more accepted practice. As a result, LANs, WANs, and distributed computing experienced tremendous growth during that period.

When first deployed, LANs were relatively secure-mainly because they were physically isolated. They were not usually connected to WANs, so their standalone nature protected the network resources.

WANs actually preceded LANs and had been around for some time, but they were usually centrally controlled and accessible by only a few individuals in most organizations. WAN sutilizing direct or dedicated privately owned or leased circuits were relatively secure because access to circuits was limited.

The Internet is the largest and best known of this type of network. The Internet utilizes TCP/IP and was primarily designed to connect computers regardless of their operating systems in an easy and efficient manner. Security was not part of the early design of TCP/IP, and there have been a number of widely publicized attacks that have exploited inherent weaknesses in its design. One well-known event was the Internet Worm that brought the internet to its knees back in 1986. Today, security has to be more important than ease of access. Network security is concerned, above all else, with the security of company information assets. We often lose sight of the fact that it is the information and our ability to access that information that we are really trying to protect-and not the computers and networks.

A simple definition for information security:

Information security = confidentiality + integrity + availability + authentication

There can be no information security without confidentiality; this ensures that unauthorized users do not intercept, copy, or replicate information. At the same time, integrity is necessary so that organizations have enough confidence in the accuracy of the information to act uponit. Moreover, information security requires organizations to be able to retrieve data; security measures are worthless if organizations cannot gain access to the vital information they need to operate when they need it. Finally, information is not secure without authentication determining whether the end user is authorized to have access.

II. THREATS TO NETWORK SECURITY

Networks and systems face many types of threats. There are viruses, worms, Trojan horses, trap doors, port scanning, sniffing, war dialing, denial-of-service attacks, and other protocol-based attacks.

A virus, a parasitic program that cannot function independently, is a program or code fragment that is self-propagating. It is called a virus, because like its biological counterpart, it requires a "host" to function. In the case of a computer virus the host is some other program to which the virus attaches itself. A virus is usually spread by executing an infected program or by sending an infected file to someone else, usually in the form of an e-mail attachment.

A worm is a self-contained and independent program that is usually designed to propagate or spawn itself on infected systems and to seek other systems via available networks. The main difference between a virus and a worm is that a virus is not an independent program.

A Trojan horse is a program or code fragment that hides inside a program and performs a disguised function. This type of threat gets its name from Greek mythology and the story of the siege of Troy. The story tells of how Odysseus and his men conquered Troy by hiding within a giant wooden horse. A Trojan horse program hides within another program or disguises itself as a legitimate program. This can be accomplished by modifying the existing program or by simply replacing the existing program with a new one. The Trojan horse program functions much the same way as the legitimate program, but usually it also performs some other function, such as recording sensitive information or providing a trap door.

A trap door or back door is an undocumented way of gaining access to a system that is builtinto the system by its designer(s). It can also be a program that has been altered to allowsomeone to gain privileged access to a system or process. A logic bomb is a program or subsection of a program designed with malevolent intent. It isreferred to as a logic bomb, because the program is triggered when certain logical conditionsare met. This type of attack is almost always perpetrated by an insider with privileged accessto the network. The perpetrator could be a programmer or a vendor that supplies software.

Like a burglar casing a target to plan a break-in, a hacker will often case a system to gatherinformation that can later be used to attack the system. One of the tools that hackers often usefor this type of reconnaissance is a port scanner. A port scanner is a program that listens towell-known port numbers to detect services running on a system that can be exploited tobreak into the system.

III. DIFFERENT WAYS FOR PROVIDING SECURITY

There are enormous ways to provide security to a standalone computer or a network(LAN or MAN or WAN).

A. PASSWORDS

Making sure that certain areas of the network are“password protected”—only accessible by those withparticular passwords—is the simplest and most commonway to ensure that only those who have permission canenter a particular part of the network. In the physicalsecurity analogy above, passwords are analogous tobadge access cards. However, the most powerful network security infrastructures are virtually ineffective if people do not protect their passwords. Many users choose easily -remembered numbers or words as passwords, such as birthdays, phone numbers, or pets’ names, and othersnever change their passwords and are not very carefulabout keeping them secret.

B. CRYPTOGRAPHY

Traditionally, cryptography conjures up thoughts of spies and secret codes. In reality, cryptography and encryption have found broad application in society.Encryption technology ensures that messages cannot beintercepted or read by anyone other than the authorizedrecipient. Encryption is usually deployed to protect datathat is transported over a public network and usesadvanced mathematical algorithms to “scramble”messages and their attachments. Several types ofencryption algorithms exist, but some are more securethan others. Encryption provides the security necessary tosustain the increasingly popular VPN technology. VPNsare private connections, or tunnels, over public networkssuch as the Internet. They are deployed to connecttelecommuters, mobile workers, branch offices, andbusiness partners to corporate networks or each other.

C. FIREWALLS

A firewall is a hardware or software solution implemented within the network Infrastructure to enforce anorganization’s security policies by restricting access tospecific network resources. In the physical securityanalogy, a firewall is the equivalent to a door lock on aperimeter door or on a door to a room inside of thebuilding—it permits only authorized users, such as thosewith a key or access card, to enter. Firewall technology iseven available in versions suitable for home use. Thefirewall creates a protective layer between the networkand the outside world. In effect, the firewall replicates thenetwork at the point of entry so that it can receive andtransmit authorized data without significant delay. However, it has built-in filters that can disallow Unauthorized or potentially dangerous material fromentering the real system. It also logs an attemptedintrusion and reports it to the network administrators.

D. INTRUSION DETECTION

Organizations continue to deploy firewalls as their centralgatekeepers to prevent unauthorized users from enteringtheir networks. However, network security is in many ways similar to physical security in that no one technologyserves all needs—rather, a layered defense provides thebest results. Organizations are increasingly looking toadditional security technologies to counter risk andvulnerability that firewalls alone cannot address. Anetwork-based intrusion detection system (IDS) providesaround-the-clock network surveillance. An IDS analyzespacket data streams within a network, searching forunauthorized activity, such as attacks by hackers, andenabling users to respond to security breaches beforesystems are compromised. When unauthorized activity isdetected, the IDS can send alarms to a managementconsole with details of the activity and can often orderother systems, such as routers, to cut off the unauthorizedsessions. In the physical analogy, an IDS is equivalent to avideo camera and motion sensor; detecting unauthorized orsuspicious activity and working with automated responsesystems, such as watch guards, to stop the activity.

E. HONEYPOTS

One technique that many administrators employ is the use of "honeypot" systems. Honeypotsare decoy or lure systems. They are basically deception systems that contain phony services,files, and applications designed to emulate well-known holes with the goal of entrappinghackers. They are designed to attract hackers, hence the name "honeypot." The honeypot isintended to make hackers believe that they have discovered a real system. The system isdesigned to lure a hacker into a "safe" network or server that impersonates importantapplications or information. When the hacker enters the honeypot the trap is sprung and thealarm is sounded. For it to work properly, the system has to be interesting enough to occupythe hacker long enough so that a security administrator can trace the hacker.

IV. CONCLUSION

The Internet has undoubtedly become the largest publicdata network, enabling and facilitating both personal andbusiness communications worldwide. The volume of traffic moving over the Internet, as well as corporatenetworks, is expanding exponentially every day. Moreand more communication is taking place via e-mail;mobile workers, telecommuters, and branch offices areusing the Internet to remotely connect to their corporatenetworks; and commercial transactions completed overthe Internet, via the World Wide Web, now account forlarge portions of corporate revenue.While the Internet has transformed and greatly improvedthe way we do business, this vast network and its associatedtechnologies have opened the door to an increasing numberof security threats from which corporations must protectthemselves. Although network attacks are presumably moreserious when they are inflicted upon businesses that storesensitive data, such as personal medical or financial records,the consequences of attacks on any entity range from mildlyinconvenient to completely debilitating—important datacan be lost, privacy can be violated, and several hours,or even days, of network downtime can ensue.

REFERENCES

[1] R. Merkle, “Secure communication over an insecure channel,”submitted to Communications of the ACM.

[2] D. Kahn, TheCodebreakers, The Story of Secret Writing. NewYork: Macmillan, 1967.

[3] C. E. Shannon, “Communication theory of secrecy systems,” BellSyst. Tech. J., vol. 28, pp. 656-715, Oct. 1949.

[4] M. E. Hellman, “An extension of the Shannon theory approach tocryptography,” submitted to IEEE Trans. Inform. Theory, Sept.1975.

[5] W. Diffie and M. E. Hellman, “Multiuser cryptographic techniques,”presented at National Computer Conference, New York, June 7-10,1976.

[6] D. Knuth, The Art of Computer Programming, Vol. 2, Semi-Numerical Algorithms. Reading, MA.: Addison-Wesley, 1969.

[7] The Art of Computer Programming, Vol. 3, Sorting andSearching. Reading, MA.: Addison-Wesley, 1973.

[8] S. Pohlig and M. E. Hellman, “An improved algorithm for computingalgorithms in GF(p) and its cryptographic significance,”submitted to IEEE Trans. Inform. Theorv.

[9] M. V. Wilkes, Time-Sharing Computer Systems. New York: Elsevier,1972.D

Security in Grid Computing

Gurbhej Singh[1] Sumit Sharma[2] Simarpreet Kaur[3]
Department of Electronics & Communication Engineering, ACET, Amritsar, Punjab
Email: gurbhej86@gmail.com, ersumitsharma@gmail.com, sim.preet330@gmail.com

***Abstract*- Grid computing provides high computing power, enormous data storage, and collaboration possibilities to its users. In the networked access to computation with a single-sign-on system as the portal to the possibilities of world wide computing grids security plays an important role. This paper provides an overview about the state-of-the-art of security in current grid technologies and a discussion about possible weaknesses which have to be considered at the current approach.**

I. INTRODUCTION

When in 1969 the project ARPANET, which subsequently became the internet, went online Leonard Kleinrock was cited in a press release from UCLA that computer networks are still in their infancy. But as they grow up and become more sophisticated, we will probably see the spread of computer utilities which, like present electric and telephone utilities, will service individual homes and offices across the country [1]. 20 years later with the introduction and standardization of HTTP and HTML [2] the internet and the world wide web started to spread out to every home.

The growing number of devices and thus mostly unused resources connected to the internet triggered many different ideas to share available computing and storage resources. In 1998 Ian Foster and Carl Kesselman defined in the book The Grid: Blueprint for a New Computing Infrastructure a computational grid as a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities. [3] This definition was subsequently refined.

In general, grid computing provides users with the ability to divide and spread large computations across multiple machines as well as access to distributed storage and collaboration possibilities within virtual organizations [4]. As in [5] stated, grids allow the simultaneous use of large numbers of resources, dynamic requirements, use of resources from multiple administrative domains, complex communication structures, and stringent performance requirements. The coordination, administration, and subsequently also the billing of resource usage in a grid is not only challenged by the decentralized control and ownership of the computing and storage structure but also by the aim to provide an efficient and usable environment to its users.

Current grid middleware like the Globus Toolkit, legion, gLite, or BOINC offer a basic software infrastructure and tools for hiding the complexity of the heterogeneous infrastructure of a grid from the users, for easing the administration and configuration of the participating systems, for coordinating and monitoring available resources, and for providing a basic layer for developers to create grid applications.

In the following a closer look will be taken at security threats, countermeasures, and detection in grids and how different grid middleware solutions approach these issues.

II. Components of Grid Middlewares

The decentralized approach of grids, the flexibility and the heterogeneous nature of their infrastructure, and the aim of being a general purpose system are a challenge to grid middle wares to provide a manageable distributed, secure, stable, and high quality-of-service system to its users. For achieving this most approaches of grid middleware are split up into the following components:

basic middleware: This module provides the basic abstraction layer from the system integrated in the grid and also the API for developing and running applications on the grid.

authentication and authorization system: This module is responsible for the authentication and authorization of users, virtual organizations, and processes accessing the grid.

workload management/scheduling: This module manages the scheduling, distribution, and prioritization of jobs and processes running on the grid.

data management: This module manages the data storage and also the access to data on the grid.

fabric management: This module provides tools for installation management of grid applications and basic resource management, monitoring, and configuration.

Information system: This module collects available information about the grid like availability and status of resources, the job queue and the status of active jobs, information about users and virtual organizations, etc. These systems allow to monitor the grid and also provide tools allowing users to interact with their submitted jobs.

Due to the characteristic of grids each of these components provides a challenge for security in informations systems to prevent manipulation, misuse, unauthorized access, denial of service, hijacking, stalling of processes, and stealing of information stored on the grid, computing power provided by the grid, and in general devices connected to the grid. The large and dynamic user population and resource pool, the dynamic acquisition and release of distributed resources during computation, and the different authentication and authorization mechanisms in the heterogeneous environment of grids require a broad view on security and contain many security challenges

III. Security areas

Each category of grid middleware modules is confronted with security issues; either with issues concerning only its category or issues concerning several categories. But as soon as one module is compromised or provides a security hole the whole grid infrastructure may be compromised by attackers, which have the chance to hijack the grid and/or get into a distributed computing infrastructure which is designed for high performance computing. Starting out with taking care of user authentication in grids [7][5] the security working group of the grid forum started to list issues, which also have to be addressed when looking at the security of grids

3.1. Authentication and authorization

Most security considerations in grids are focused on the authentication and authorization to access the available resources on the grid. From a usability point of view to access the grid and all its resources users should enter their login credentials only once. This single-sign-on approach to access the heterogeneous and distributed environment of grids is a corner stone of the success and further growth and distribution of grids [7][5]. Using public key infrastructure (PKI) based on X.509 certificates has become the standard for grid middleware – like Globus or gLite – to implement the single-sign-on approach. The usage of this PKI establishes a mutual trust relationship between the user and the entry point to the grid allowing not only the grid to check the user's certificate but also vice versa allow the user to verify the entry to the grid via it's certificate.

In addition to this basic authentication the middle wares Globus and gLite use a user proxy approach to delegate the credentials to the systems either used for computations by the user or the user's processes or containing data required by the user or the user's processes. Using a proxy the user delegates the rights to it which again can delegate the rights to processes started by the user and needing to access other systems of the grid infrastructure. To avoid the exposure and publication of the user's credentials the proxy uses its own credentials which are only valid for short period of time, usually for about 12 hours. On the systems themselves the grid users are mapped to local user accounts, which allow the execution of the requested jobs and access to data necessary for the execution of the request. In Globus and gLite this user-mapping is based on gridmap-files. In Globus and gLite the access to resources can also be limited by requiring users to be members of virtual organizations (VO). By restricting the access to systems of the grid infrastructure to special VOs, only members of those VOs are authorized to access them.

3.2. Scheduling

The scheduling of jobs and managing a job's access to data, especially if the executed process is very data intensive [13], is an important topic in grid computing. Processes running in grid environments do not only require CPU time but also bandwidth and data storage, which should be reserved for the processes. Due to the structure of grids and resources which are managed, distributed scheduling of tasks improves the scheduling performance and according to [14] makes a system portable, secure, and capable of distributing scheduling workload among an array of computational sites in the system.

When looking at different grid middleware implementations, not all of them support security- and policy-based scheduling. Although Globus and gLite use GSI as a basic layer for all processes and users have the ability to specify several conditions which have to be met by resources to decide which ones should be used [6] the scheduling is by default based on the information specified by the users in the job description file and the user's virtual organization.

3.3. Execution

After a job is submitted and scheduled for execution it is submitted to the designated resource/computing element for execution. Several security aspects have to be considered at this stage. From an administrator's perspective the job should have no possibilities to do any harm to the resource it is running on. It should not be able access data and other jobs it is not allowed to access as well as it should not be able to consume so many resources on the computing element that other locally originated jobs starve due to resource shortage. Several methods like application-level sandboxing, virtualization, user-space sandboxing, or flexible kernels can be used to protect data on the computing element. Grid middleware like Globus or gLite only allow accountable users to submit jobs, develop new applications, and upload and access those jobs and applications.

3.4. Data Access and Management

Besides distributing and parallelizing computation on the grid the access to and storage of data in the heterogeneous environment of a grid also provides many challenges regarding distribution, replication, and performance as well as concerning security issues. Globus and gLite use GridFTP as transportation protocol, which is a FTP solution, is based on GSI, and uses transport layer security (TLS) for securing the file transfer between clients, storage elements,
and computing elements.

3.5. Information and monitoring

Collecting information about available and used resources-computing and storage resources, the status of jobs, active services, etc. is a vital part of managing and also using a grid infrastructure. Each grid middleware has tools to collect information provided optionally by computing and storage elements and information which must be available to schedulers and make this information via a public interface available to administrators and users of the grid. gLite for example includes the web information and monitoring tool GridICE[6], which can provide information about available memory, number of CPUs, storage size, etc. for computing and storage elements, which is useful information for administrators and users, but also possibly useful information for attackers. Apart from this detailed information about resources of the grid infrastructure, which is publicly available on websites, grid portals also provide the possibility to submit jobs and interact with them.

In addition to security issues of the grid middleware itself these grid portals introduce security issues of web applications to the grid infrastructure. Although breaking into a Grid (...) may not necessarily allow the attacker access to backend Grid resources, but as most grid portals allow users to access grid resources and manage their credentials, monitor and maybe even interact with their running jobs, breaking into grid portals provides the same rights to an attacker as the grid user has on the grid portal. Generally speaking, providing access to and control over grid resources to users via a web portal increases the security risk of a grid infrastructure.

IV. Conclusion

Due to the heterogeneous environment, the distributed infrastructure, the decentralized administration and monitoring, and the available computing power in grids security issues in grid computing are manifold. Apart from security issues on single grid elements, security issues of the grid middleware may affect all connected grid elements. Therefore, constant monitoring, vigilant distributed intrusion detection, thoughtful rights management, regular updates, etc. are even more necessary than in other computing environments. Basic tools for monitoring and rights management are available. Still progress has to be made regarding data protection and data privacy in grids and sandboxing of processes on computing elements. Many steps are necessary to provide a secure grid environment to its users and maybe even more to move the grid from mainly scientific applications to private and industrial application of publicly available grid computing and manage the step with grid computing, which the World Wide Web has achieved in the 1990ies.

REFERENCES

[1.] Kleinrock, L.: An internet vision: the invisible global infrastructure (2003)

[2.] Berners-Lee, T., Cailliau, R.: Worldwideweb: Proposal for a hypertext project. Technical report, CERN (1990)

[3.] Foster, I.: What is the grid? - a three point checklist. GRIDtoday 1(6) (July 2002)

[4.] Humphrey, M., Thompson, M.R.: Security implications of typical grid computingusage scenarios. In: HPDC ’01: Proceedings of the 10th IEEE International Symposium on High Performance Distributed Computing, Washington, DC, USA, IEEE Computer Society (2001) 95

[5.] Foster, I., Kesselman, C., Tsudik, G., Tuecke, S.: A security architecture for computational grids. In: CCS ’98: Proceedings of the 5th ACM conference on Computer and communications security, New York, NY, USA, ACM (1998) 83–92

[6.] Burke, S., Campana, S., Lorenzo, P.M., Nater, C., Santinelli, R., Sciaba, A.: gLite 3.1 user guide. Members of the EGEE-II Collaboration. 1.2 edn. (April 2008)

[7.] Butler, R., Welch, V., Engert, D., Foster, I., Tuecke, S., Volmer, J., Kesselman, C.: A national-scale authentication infrastructure. Computer 33(12) (2000) 60–66

[8.] Humphrey, M., Thompson, M.: Security implications of typical grid computing usage scenarios (2001)

[9.] Park, N., Moon, K., Sohn, S.: Certificate validation service using xkms for computational grid. In: XMLSEC ’03: Proceedings of the 2003 ACM workshop on XML security, New York, NY, USA, ACM (2003) 112–120

[10.] Novotny, J., Tuecke, S., Welch, V.: An online credential repository for the grid: Myproxy (2001)

[11.] Welch, V.: Globus toolkit version 4 grid security infrastructure: A standards perspective. Technical report, Globus Alliance (2005)

[12.] Kim, B.J., Hong, S.J., Kim, J.: Ticket-based fine-grained authorization service in the dynamic vo environment. In: SWS ’04: Proceedings of the 2004 workshop on Secure web service, New York, NY, USA, ACM (2004) 29–36

[13.] Xue, Y., Wan, W., Li, Y., Guang, J., Bai, L., Wang, Y., Ai, J.: Quantitative retrieval of geophysical parameters using satellite data. IEEE Computer 41(4) (2008) 33–40

[14.] Xie, T., Qin, X.: Enhancing security of real-time applications on grids through dynamic scheduling. In Feitelson, D.G., Frachtenberg, E., Rudolph, L., Schwiegelshohn, U., eds.: JSSPP. Volume 3834 of Lecture Notes in Computer Science, Springer (2005) 219–237

[15.] Xie, T., Qin, X.: Security-driven scheduling for data-intensive applications on grids. Cluster Computing 10(2) (2007) 145–153

Cloud Computing: Secure Architectures

Jasmeen Kaur[1], Aditi Aggarwal[2] , Harpreet Kaur[3], Ramanjyot Kaur[4]
Sant Baba Bhag Singh Institute Of Engineering And Technology, Jalandhar
E-mail: [1]jassnarwal30@yahoo.com, [2]aditi.aggarwal17@yahoo.in,
[3]er.harpreetarora@gmail.com , [4]ramanjyotk@yahoo.com

Abstract- In cloud computing, data storage and processing are offered as services, and data are managed by external providers that reside outside the control of the data owner. The use of such services reduces the burden of the owners in managing their data, and may provide significant cost savings. However, cloud computing introduces new security and privacy concerns. In fact, there is little consensus on how to guarantee the confidentiality, integrity, and availability of data in cloud computing scenarios. Also, it is unclear to what extent parties can be held accountable in case something goes wrong. In this paper, we searched for architectures, modeling approaches, and mechanisms that can help in providing guarantees for cloud security. We proposed the concept of verification-as-a-service that can guide architectures for verification of cloud architectures and configurations, as well as results of computations. We also proposed architectures for organizing customizability of security and privacy for cloud customers.

1. Introduction

In cloud computing, data storage and processing are offered as a service, and the data resides outside the control of the owner. It is often argued that clouds improve security, as the providers have more security expertise than their (smaller) customers. However, despite theoretical breakthroughs in cryptography, there is little consensus on how we can provide architectural solutions guaranteeing that cloud data remains confidential, uncorrupted, and available. The main question was which cloud-specific security architectures should and could be devised, and how they can be matched to security policies.

1.1 Data protection: Data outside the data owner's control implies that privacy and even integrity can be put at risk. Guaranteeing the privacy and integrity of the data, whether stored in the system or communicated to external parties, becomes a primary requirement, and has raised the attention of both individuals and legislators. Cloud providers have to properly protect the privacy of (possible sensitive) information when storing, processing or sharing it with others and have to adopt adequate access control solutions for enforcing selective access to the data. New approaches have emerged for identifying persons and roles and linking them to access privileges, such as identity, attribute, claims, and data-based access control.

1.2 Simulating physical constraints in the cloud: In the cloud, we cannot easily enforce where data is stored and how long, and from where it is accessed. Location-based access control aims at limiting access to specific locations, thereby seemingly putting physical limitations back in place. Measures proposed include use of GPS, trusted platform modules (TPMs), but also physically unclonable functions (PUFs). Also, data could be moved away from attacks. With respect to time, mechanisms have been proposed to assure deletion of data in the cloud. We assessed to which extent these approaches are sufficient to simulate physical constraints, and which architectural solutions are needed to make such forms of assurance possible in practice.

1.3 Misuse detection: Many methods have been proposed for intrusion detection, penetration testing and digital forensics. The necessary adaptations to system and threat models as well as security metrics, to adequately indicate which attacks are possible and which are actually happening, and thereby reduce cyber crime.

1.4 Splitting the clouds: Public clouds, containing data from different parties, are not deemed suitable for particularly sensitive information. This means that decisions will have to be made about which data to put in the cloud and which data not, which security properties to outsource and which not, and how to make sure that the entire system conforms to the security requirements.

2. Main Findings

As a general observation, we concluded that clouds require a different kind of architectural decisions than traditional information systems. In complex systems such as clouds, we cannot do lots of things manually anymore. For example, there is usually no way to inspect a cloud for evidence manually after an incident.

This means that the architecture needs to allow for automation of such tasks, by providing not only functional services, but also meta-services to perform automated maintenance, recovery, etc. Moreover, the processes that make use of such meta-services need suitable architectures themselves. In particular, the following meta-services are needed:

- Automated policy checking,
- Automated configuration verification,
- Automated incident management,
- Automated auditing, and
- Automated forensics.

These processes could be deployed again in (different) clouds, but then the same security concerns apply to them as well.

In particular, we proposed the concept of verification-as-a-service, which can refer to both the verification of the results of computations, as well as the verification of the (security) architecture and configuration in place at the cloud provider. The former is well-known in the field of electronic voting systems; the latter resonates with the practice of security auditing. We can do is to integrate GPS with trusted hardware (such as TPM) to prove locations. Verification-as-a-service provides a paradigm to organize accountability in the cloud. This could be realized by different techniques, for example by: Transparency of architecture/configuration (inspection/attestation), Forensics (e.g., watermarking), Regulation (precaution) and enforcement, Incident response (logging), or Creating incentives.

Verifying the integrity of data seems to be more intuitive than verifying its confidentiality. With integrity, it is possible, for example, to compare two different copies. With confidentiality, one would have to prove that only certain parties possess a copy. It only seems to be possible to falsify this after the fact, when it is indeed discovered that data has been leaked. Even in that case, one would need some kind of watermark to prove who leaked the data, for it might have been the user as well as the provider. How to develop a service that provides such watermarking in relation to confidentiality-as-a-service has been identified as an open problem. Especially on the user side, accountability can be further enhanced by modifiability, or customizability, which allows the user to adapt services to his or her own policies. This requires negotiation on policies, not only between the user and the initial provider, but also between providers within the supply chain. Again, special services can be set up that allow the user to achieve this for multiple cloud services at the same time, which would amount to modifiability-as-a-service. Such services could be standardized to make sure that they really empower the user, by employing certain privacy policies themselves, and providing an understandable interface. We would then have achieved “standardized customizability”.

We formulated several attacker models that lie behind these proposals. Many standard attacker models are problematic in the cloud. An evil/malicious cloud service provider implies that we cannot solve anything without advanced encryption methods, which are costly or even infeasible in many scenarios. Assuming that computations are performed in the clear, we have to assume that the cloud service provider is indifferent, not curious. Thus, we trust the cloud provider on the issue of confidentiality, in the sense that we do not expect the provider to leak or misuse data intentionally. However, the provider may still be a:

- Sloppy provider (makes mistakes),
- Lazy provider (simplifies computations), or
- Greedy provider (reduces security to save money).

The sloppy and lazy provider might compromise the integrity of the result of computations. Verification of results would be a countermeasure here, for example by executing the computations on multiple, independent clouds.

Greedy providers are willing to violate policies for economic reasons, thereby exposing the data to insider or outsider threats. Although we do not assume malice on the side of the provider, we do assume malice on the side of other cloud users, who may or may not have specialized access (e.g., administrators). In relation to the greedy provider, one would want to have some means to verify the architecture in place.

Especially if services have been customized, one would want to have some kind of assurance that there is actually a change in configuration taking place based on the customization. We proposed the development of a tool suite to support remote measurements of architectural variables, which would include existing proposals. Care needs to be taken that acquiring such information does not violate customer privacy or company property rights. Also, even if the architecture would be (partly) known, the user would then need meaningful support to choose among different providers (and thereby different architectures).

This provides another incentive to develop quantitative models that can indeed calculate overall security risks from system architectures, based on existing qualitative approaches. The user can then compare risks and costs to make decisions. Such decisions could even be made in real-time based on information on the current security situation, leading to what has been called fluid information systems.

A remaining question is how to create incentives to invest in cloud security. If there is no immediate impact, investments may lag behind with respect to threat levels. Ironically, you can gain a competitive advantage by making your competitors invest in security. Do we really need big scandals to improve security? In any case, achieving more security by (self-) regulation, whether by law, seals, or otherwise, requires architectures.

3. Adaptive Information Security for Cloud Services: Relating Security Requirements to Design

Information security involves protecting valuable information assets from possible harm. With the increasing use of cloud computing services, the technical and social contexts in which software applications are expected to operate become increasingly dynamic. As a result, the assets, their values, and attack scenarios can easily change. This increases the challenge of finding out what the information assets are, who their owners are, where in the system vulnerabilities lie, and the extent to which the security requirements need to be enforced. In such an environment, information security has to be highly context-sensitive: software applications must adapt to the changing contexts and respond quickly and appropriately to ensure that the requirements for information security are not violated. We call this notion Adaptive Information Security, and focus on three of its prerequisites in the context of cloud computing: (1) understanding user requirements for cloud applications; (2) traceability between requirements, design and implementation of cloud services; and (3) adaptive design for dynamic contexts.

4. Security Assurance in Virtualized Infrastructures

Cloud computing and virtualized infrastructures are often accompanied by complex con-figurations and topologies. Dynamic scaling, rapid virtual machine deployment, and open multi-tenant architectures create an environment, in which local misconfiguration can create subtle security risks for the entire infrastructure. This situation calls for automated deployment as well as analysis mechanisms. We present a platform that combines a static information flow analysis and a virtualization assurance language with state-of-the art verification methods. The system discovers the actual configurations of diverse virtualization environments and unifies them in a graph representation. Using graph traversal, it computes the transitive closure of information flow. The language integrates descriptions of virtualized infrastructures, their transformations, their desired security goals, and evaluation strategies. The different verification tools range from model checking to theorem proving; this allows us to exploit the complementary strengths of methods.

5. Versatile Key Management for Secure Cloud Storage

Storing data on cloud-based infrastructures facilitates infinite scalability and all-time availability. Putting data in the cloud additionally offers a convenient way to share any information with user-defined third-parties. However, storing data on the infrastructure of commercial third party providers, demands trust and confidence. Often simple approaches, like merely encrypting the data by providing encryption keys, which at most consists of a shared secret supporting rudimentary data sharing, do not support evolving sets of accessing clients to common data.

Based on well-established approaches regarding stream-encryption, we propose an adaptation for enabling scalable and flexible key management within heterogeneous environments like cloud scenarios. Representing access-rights as a graph, we distinguish between the keys used for encrypting hierarchical data and the encrypted updates on the keys enabling flexible join-/leave-operations of clients. This distinction allows us to utilize the high availability of the cloud as updating mechanism without harming any confidentiality. Our graph-based key management results in a constant adaptations of nodes related to the changed key. The updates on the keys generate a constant overhead related to the number of these updated nodes.

6. Energy Efficiency in Cloud and Related Security Issues

Cloud computing is a promising approach for implementing scalable on- demand computing infrastructure. It includes business aspects like SLAs and customer-provider relationship, as well as organizational issues like scheduling, resource allocation, all the way to technical details like VM monitoring and application deployment. While energy efficiency is mostly managed on an organizational level, it is realized by actions on the level of clusters, physical machines, VMs or even a single application.

By monitoring customer's applications for a purpose of more efficient scheduling, provider reaches the privacy border. Also, by applying energy efficient measures like time-sharing VMs and running multiple VMs on a single physical machine, provider creates vulnerable environments for customer's applications. Can customer trust provider's measurements; how secure is his application; is customer's privacy being threatened; these are all the questions which cannot be neglected for benefit of energy efficiency, but should certainly be considered.

Conclusion

This paper proposed architectures for verifying the results of cloud computations, verifying the configuration of cloud architectures, and supporting customizability of cloud services in terms of security. These were defined in relation to cloud-specific attacker models.

References

[1.] S. Bleikertz, T. Groß, and S. Mödersheim. "Automated verification of virtualized infrastruc-tures. In CCSW'10: Proceedings of the 2010 ACM Cloud Computing Security Workshop,pp. 47–58. ACM, 2010.

[2.] S. Bleikertz, M. Schunter, C.W. Probst, D. Pendarakis, and K. Eriksson. Security audits of multi-tier virtual infrastructures in public infrastructure clouds. In CCSW'10: Proceedings of the 2010 ACM Cloud Computing Security Workshop, pp. 93–102. ACM, 2010.

[3.] T.D. Breaux and C.B. Lotrionte. Towards a privacy management framework for distributed cybersecurity in the new data ecology. In HST'11: Proceedings of the IEEE International Conference on Technologies for Homeland Security