IJACSA
WHERE WISDOM SHARES
International Journal of Advanced Computer Science and Applications

Volume 2 Issue 3
March 2011

ISSN 2156-5570(Online)
ISSN 2158-107X(Print)

SAI
www.ijacsa.thesai.org

IJACSA Editorial

From the Desk of Managing Editor…

It is a pleasure to present to our readers the March 2011 issue of International Journal of Advanced Computer Science and Applications (IJACSA).

With monthly feature peer-reviewed articles and technical contributions, the Journal's content is dynamic, innovative, thought-provoking and directly beneficial to readers in their work.

The number of submissions we receive has increased dramatically over the last issues. Our ability to accommodate this growth is due in large part to the terrific work of our Editorial Board.

In order to publish high quality papers, Manuscripts are evaluated for scientific accuracy, logic, clarity, and general interest. Each Paper in the Journal not merely summarizes the target articles, but also evaluates them critically, place them in scientific context, and suggest further lines of inquiry. As a consequence only 28% of the received articles have been finally accepted for publication.

We have high standards for the quality of papers we publish and for the reviewers that help us. The Reviewer Board reflects both the journal's truly international readership and wide coverage.

At The Science and Information Organization, we are constantly upgrading to the latest technology to speed up the publication process and add greater functionality to the published papers.

The success of authors and the journal is interdependent. While the Journal is advancing to a new phase, it is not only the Editor whose work is crucial to producing the journal. The editorial board members , the peer reviewers, scholars around the world who assess submissions, students, and institutions who generously give their expertise in factors small and large— their constant encouragement has helped a lot in the progress of the journal and earning credibility amongst all the reader members.

We hope to continue exploring the always diverse and often astonishing fields in Advanced Computer Science and Applications

Thank You for Sharing Wisdom!

Managing Editor
IJACSA
Volume 2 Issue 3, March 2011
editorijacsa@thesai.org
ISSN 2156-5570(Online)
ISSN 2158-107X(Print)

IJACSA Associate Editors

IJACSA Reviewer Board

CONTENTS

Quality of Service Management on Multimedia Data Transformation into Serial Stories Using Movement Oriented Method

A. Muslim
Etudient Doctoral Informatique
Le2i-UMR CNRS 5158 Université de Bourgogne
9 Av. Alain Savary 21078 Dijon - France
Aries.muslim@etu.u-bourgogne.fr

A.B. Mutiara
Faculty Computer Science and Information Technology
Gunadarma University
Jl. Margonda Raya No. 100 Depok 16424 - Indonesia
amutiara@staff.gunadarma.ac.id

***Abstract*— Multimedia data transformation into serial stories or story board will help to reduce the consumption of storage media, indexing, sorting and searching system. Movement Oriented Method that is being developed changes the form of multimedia data into serial stories. Movement Oriented Method depends on the knowledge each actor who uses it. Different knowledge of each actor in the transformation process raises complex issues, such as the sequence, and the resulted story object that could become the standard. And the most fatal could be, the resulted stories does not same with the original multimedia data. To solve it, the Standard Level Knowledge (SLK) in maintaining the quality of the story could be taken. SLK is the minimum knowledge that must be owned by each actor who will perform this transformation process. Quality of Service management could be applied to assess and maintain the stability and validity of the level of the each system to SLK.**

Keywords-multimedia data; serial stories; movement oriented; standard level knowledge;quality of service

I. INTRODUCTION

In the decades when the use of digital media is growing rapidly, both in size and type of data are not only on the text but also on the image, audio and video. Along with the increasing use of digital media, especially video, is required technical data management and retrieval of image effectively. Volume of digital video produced by the field of scientific, education, medical, industrial, and other applications available for its users increased dramatically as a result of the progress in sensor technology, the Internet and new digital video. Engineering tools annotate video with text and image search using the text-based approach. Through the description text, image can be organized by semantic hierarchy for easy navigation and search based on standard Boolean queries. Because the description text for a broad spectrum of video cannot be obtained automatically, the system mostly text-based video retrieval requires annotation manually. Indeed, manual video annotation is a difficult and expensive work for a large video database, and is often subjective, context-sensitive and not perfect [1].

There are two approaches that can be used to represent the video: (1) metadata-based and (2) content-based. Techniques required for the retrieval (query) from the two approaches which can be divided into three namely: (1) Context-based, (2) Semantic-based and (3) Content-based.

Using a content-based approach to represent a video into serial stories, the standard requires knowledge that can eliminate the error. And to keep the system knowledge, it is required of a quality level that can maintain the quality of the resulting story. The method that can keep the standard of knowledge to reaches a certain level is the quality of service management.

II. MOVEMENT ORIENTED METHOD

Multimedia (video) has three data elements, namely, (1) Image, (2) Vice (3) Text or character of a display to other combinations of these three elements [10]. Multimedia system design is a very complex problem, where the system complexity is growing a lot of elements on each linier changes or nonlinear of the results presentation and interaction system. Movement Oriented Design is a new paradigm that will help to manage the level of complexity in the design problem, and so that searching the multimedia data (video) is so diverse [2].

Data/video retrieval on the number of video database with a very large system using content-based experience will be a problem on identifying the time to give each frame of video files, which is a series of identifying each textual data object that will represent the video frame, where a combination of data is a metadata that will add more storage media needs that are not less [3]. Textual process of adding data to the multimedia data can be described as figure 1.

Movement Oriented Design is a basic concept of the form of multimedia system, whether they are formal (educational) or a game (game), all of which have a certain flow of the story itself. Art establishment flow must be a story of integration of 3 scientific disciplines, namely 1) art and storytelling ability / art storytelling, 2) Cognitive psychology and 3) technology / programming [2].

At the time of this object-oriented program is a key tool in making the system of a multimedia/video that you can manipulate the object of a story, either in part or whole flow of multimedia stories. Changes can be made based on the interpretation of the respective manufacturer/user of a series of multimedia stories. These changes are very dependent on the

level of knowledge, experience, and psychological situation of the personnel to do so. This design will produce a series of stories that come from each multimedia frame that will form the collection of data/database textual which is the identification of each frame of the overall display multimedia data. Collection of data is the alphanumeric data textual that make us easy to do with the process and indexing query during search process of data.

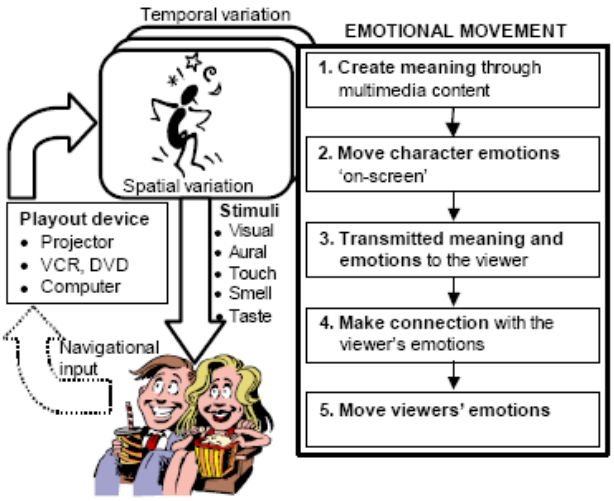


Figure 1. Multimedia Data transformation in the form of Textual Data

Quality of Service management is method to maintain the quality of service to consumers. QoS widely applied in telecommunications service and web processing. Referring to this, QoS can be used to maintain the standardization of database knowledge to the transformation of video data into serial stories, without reducing the significance with exist in the real video data. As studies on the E-Service Composer applying BPEL engine, which they express the needs of clients with their respective state standard [4].

A. *Problem*

- Story collection story are multimedia representation of the set of statements/character in the form of a database. It is relatively large because of the various capabilities of each of the personal view [3]. This is like the figure 2.
- Type of the multimedia data story of multimedia data will reflect the distinctive character, which of course will vary according to table I.
- Indexing the form of navigation and data processing work will require a relative faster speed, because each frame/stage will result in some textual statement. It requires the optimal index algorithm of data collection of stories, when it is applied on the search process or queries. An example is shown in table II. and figure 3.
- It is required the standard of knowledge base to transformation process to get story in accordance with the original video.
- It is required the management to the standardization of knowledge.

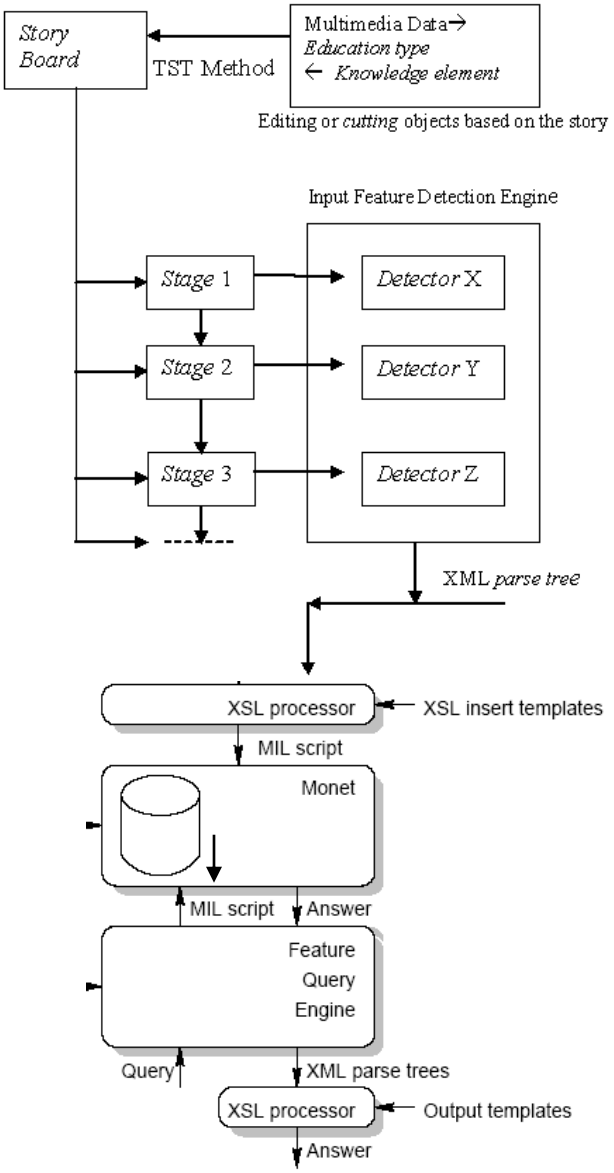


Figure 2. Diagram process data fragmentation video

TABLE I. THE CHARACTERS IN THE TABLE NEED A MULTIMEDIA DATA

Story Type	Charakter Type
Humanistic	Human Beings
Animated	Animation Beings
Game	Game Beings
Education	Knowladge elements
Song	Word, metaphors
Musik	Notes, Movements
Multisensory story	” + Touch, Smell & Taste
Formal Story	Any of the above

TABLE II. MAIN PROBLEM: TO EXPLAIN AND DEMONSTRATE IMPORTANT ASPECTS OF ELECTRIC CURRENT TO HIGH SCHOOL STUDENTS.

STAGE-1
Problem: To explain the meaning of electric current
B1 Importance of Electric Current (EC)
M1 Define and exemplify EC
E1 Link to Ohm's Law. Explain AC & DC Each of the Stage-1 Story Units, i.e.
B1, M1 and **E1**can now be expanded further. For example, '**B1,B2**' in Stage-2 is the Begin of the Story Unit '**B1**' in stage-1.

STAGE –2
B1 Problem: Why is EC important
Explain the importance of Electric Current (EC)
B1,B2 Many people die of electric shock
B1,M2 Understand and respect EC, not be afraid of it.
B1,E2 EC is useful for running appliances
M1 Problem: How is EC defined
Define and exemplify EC
M1,B2 Amperes = Coulombs / second
M1,M2 It's like watching Coulombs go past and counting how many go past in one second.
M1,E2 Demonstrate the effect of EC through multimedia and ***multisensory experience***.
E1 Problem: What determines EC strength
Link to Ohm's Law. AC/DC
E1,B2 Current depends upon voltage and resistance
Coulombs go past and
counting how many go past in one second.
M1,E2 Demonstrate the effect of EC through multimedia and ***multisensory experience***.
E1 Problem: What determines EC strength
Link to Ohm's Law. AC/DC
E1,B2 Current depends upon voltage and resistance
E1,M2 Ohm's Law: I = V/R
E1,E2 Current can be direct or alternating.
Some of the Story Units at Stage-2 have been expanded further in Stage-3; whereas some of these have been left out, signifying that Story Units can be instantiated in more ways than one.

STAGE -3
B1 Problem: Why is EC important
The importance of Electric Current (EC)
B1, B2 Problem: Effect and cause of electric shock
Many people die of electric shock
B1,B2,B3 Video clip of a person getting a shock
B1,B2,M3 Explain the reason for the shock
B1,B2,E3 Ask, "So what is electric current?"
B1, M2 Problem: How should we treat EC
Understand and respect EC, not be afraid of it.
B1,M2,B3 (Left un-instantiated
B1,M2,M3 for the user to
B1,M2,E3 try out some options)
B1, E2 Problem: How do we use EC
EC is useful for running appliances
B1,E2,B3 (Left un-instantiated
B1,E2,M3 for the user to
B1,E2,E3 try out some options)
M1 Problem: How is EC defined
Define, exemplify and inject EC

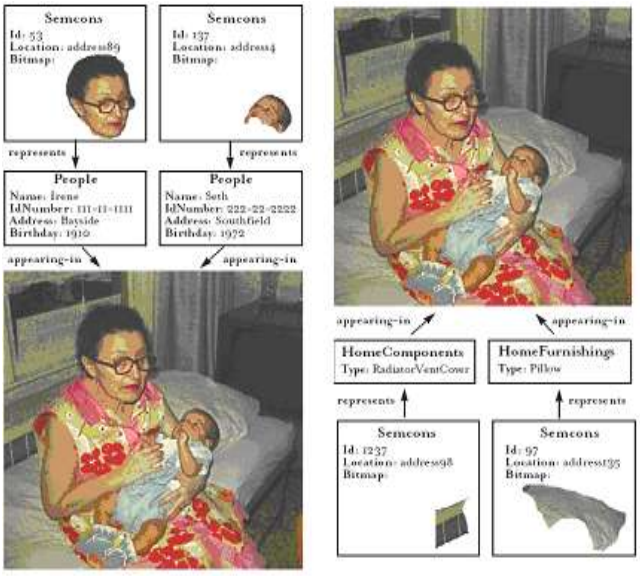


Figure 3. Content Base Method (Semcon) Identify Multimedia Data

III. RESEARCH METHOD

Refer to the problems in the process of change to the multimedia data in the textual data with the movement oriented method, which results in a textual database as multimedia identifying data that will be in the index, so the conducted steps in research are as following:

- We use a data-based education about robot video (MPG, AVI , MPEG3 / 4)
- Cutting the video data objects is based on the main story (Video Editing, Ulead Video Studio)
- Building a database metadata (story board, image and sound) which comes from the editing/cutting data to determine tuple video that will be the object index
- Building a video database with object oriented programming architecture based on ACOI (A System for Indexing Multimedia Object) to get tuple to process the index.
- Determining the stage for a major input in the feature detector Detection Engine.
- Making a simulation with the available databases, to design algorithms index of the overall data.
- Building a knowledge database architecture and design process of matching serial stories, image, sound and movement.
- To measure the success of the index algorithm design, we conducted the simulation / testing query data on the original database with a database that has been in the index, namely when DBA T < T dBi, then the index is not successful, the algorithm index enhanced DBA Q >> Q dBi, then the index is successful. The process method with the index movement oriented to the multimedia data can be done.

Movement Oriented as new paradigm for multimedia design, called the Movement Oriented Design, will change frame-per-frame data into a series of multimedia story (text only/stage) based on the movement, making the process identifying, indexing and searching data to be simpler. This is because the data will be textual data exploration. It is the ideal form of notice that is required by a database system, whether small or large. Multimedia data of education has a specification of a particular approach, so that changes every frame of the story plot does not become too diverse. This will facilitate the process to identify image and voice into a series of story that describes the stage of multimedia data.

Identifying a frame that has been based textual data will be easier to classify based on the limitations of the multimedia data. For example learning about the movement of robots, the classification can be drawn is that, 1) type of robot, 2) the form of a robot, 3) type of movement, 4) a lot of movement, without having to see the background, about the condition or feelings creator element. Object Oriented Programming is a technique that supports the process of classification into objects that can be used as a restriction of the index data.

Quality of Service (QOS) is the important part that must be a reference Identify when the multimedia data into textual data does not have the same level of knowledge of each personal, so that interpretation is not a frame out of the path of operation [5].

Algorithm for parallel process video/data real-time video with search range 40 ms/period requires a technique that is very complex, where the frame of an object is based on the main level partition needs [6]. Partition methods have been developed, namely, (1) AST - Almost Square Tiles data partitioning algorithm, (2) AST war - Almost Square Tiles data partitioning algorithm with aspect ratio, and (3) Lee-Hamdi method.

Following algorithm is AST

1) k ← least greater or equal square(partitions)
2) first square ← squareroot(k) (A)
3) cols ← first square
4) if (partitions is a square of an integer) rows ← first square (B)

   Else if ((rows-1)*cols partitions) rows ← first square -1
5) irr col ← cols * rows - partitions (C)
6) a ← image height/rows
7) ap ← image height - a * (rows -1)
8) ar ← image height/num rows - 1 (D)
9) arp ← image height - ar * (rows - 2)
10) b ← (image width/((ar/a) * (cols-irr cols) + irr cols)) * (ar/a)
11) bp ← (image width - b * (cols-irr cols)) / irr cols (E)
12) bpp ← image width - b * (cols-irr cols) - bp * (irr cols - 1)

RST - regular standard ones (a ∗ b),

RET - regular excess ones (ap ∗ b),

IST - irregular standard one (bp ∗ ar),

ICET - irregular column excess ones (bpp ∗ ar),

IRET - irregular row excess ones (bp ∗ arp),

IRCET - irregular double excess one (bpp ∗ arp);

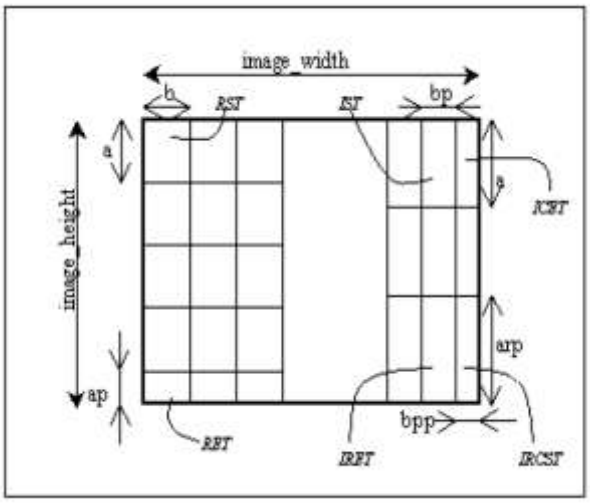


Figure 4. Picture of the position of video data on Algorithm AST

From the discussion above that the algorithm developed will be very difficult if the video database in large numbers, because will be very many objects to which the process of identification constrain index. Figure 4 shows the process identification of the image position using the AST algorithm. In the image shown in the process of collecting similar images, the regular and collected standard is called RST. For those that are not regularly, are collected into IST. So it is on both the form of row or columns.

A. Indexing Multimedia Objects

Creating Index-based multimedia objects, which provides freedom in the grammar will form a new structure on the semi-main memory database system, using the derived parse tree to make the index on the source of multimedia data [7].

Detector X represents the voice data, the detector is variable among others; 1) Voters people based on gender, region of origin, age, and so forth. 2) Hearing music: music instruments, stringed instrument, musical instrument quotation, 3) natural Voters, and others.

Detector Y represents the image data, with variable detector, among others: 1) Form of the picture, 2) type of image, which can be detected with a color histogram, edge detectors and so forth. Air Z is the representation of text data, the variables detector more similar.

From the above conditions it will result in the diversity of feature detection, but can be overcome by using the parse tree-based object, such as the example below, figure 5.

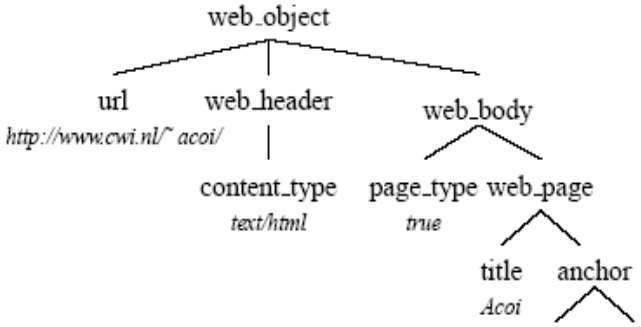


Figure 5. Parse Tree Based Object

From the parse tree approach will produce a tuple is unique, so it can be used as a variable index of the multimedia database.

B. Quality of Service

Quality of Service (QoS) is an important part that should be a reference, if the identification and transformation process multimedia data into serial stories do not have the same level of knowledge of each person. For the interpretation of a frame it is not to be of outline [8,9]. The transformation and identification process multimedia data into serial stories with standard level of knowledge into this process is easier and produce a collection of stories that fit with the video data. But if this done with a level of knowledge that is not standard, it can be curtained. The story is not obtained in accordance with the data video. It becomes very subjective, because each individual's knowledge becomes the size of the story.

For the process of transformation and identification of multimedia data into serial stories with a standard level of knowledge using only movement oriented method. While that does not make standard level knowledge required two additional processes, namely matching process and Quality of Service. Matching Process is to match the transformation of data with a database of knowledge and the video frame data Quality of Service is a service to provide a standard of knowledge of the data matching process before it is sent story board. With the quality of service can reduce the level subjectivities of each user.

Figure 6. shows a diagram that provides an illustration of the process of transformation and identification of multimedia data (video) into serial stories with the standard of knowledge and the non-standard of knowledge and the non-standard of knowledge.

The process that does not use a standard knowledge becomes more complex than it that uses a standard knowledge. In figure 6 it is described that there are three processes for transforming a multimedia into serial stories, 1) transformation process, 2) matching process and 3) quality of service process.

Figure 7. shows data process that is already in the matching process to carry out the quality of service. Each data derived from the matching process will be compared with a database of existing knowledge in order to become standard serial stories.

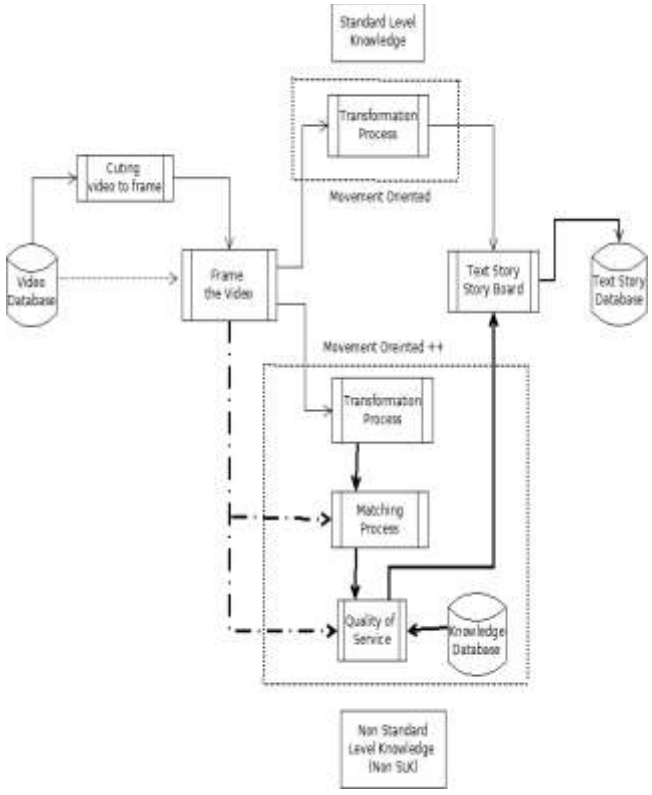


Figure 6. Movement Oriented Transformation with Quality of Service

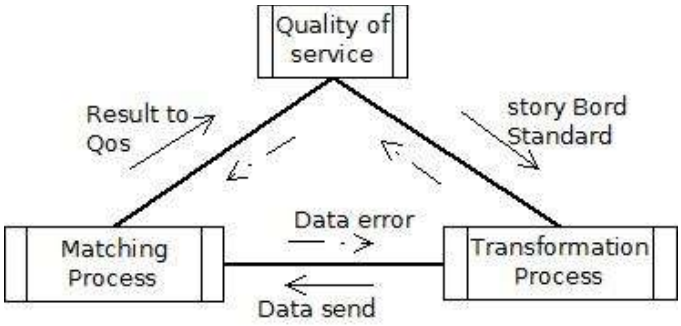


Figure 7. QoS on data Transformation and Matching Process

The result by using movement oriented with standard level knowledge is the following serial stories: robot with the background wall, swaying from side to side, motion in the legs and arms, to head downward and upward. The result of the story is very standard and does not include other element.

The result by using the movement oriented of non standard level knowledge is the following serial stories: Iron robot with eyes lit; there was a wall behind the robot, the robot moves to the left and right, rotating the robot head up and down. The transformation story approaches the desired results, but in addition there is the word "iron with his eyes lit", which is an element of user subjectivities.

And the result by using movement oriented and non standard level knowledge and plus quality of service is the following serial stories: Robot in front of the wall, moving to the left and move to the right, on leg and arm movements, head move up and down. The transformation produces a shorter story but all of it has the standard word and knowledge.

IV. Conclusion and Future Work

The division of a series of frames from video data movement cannot be done in a partial, to avoid errors in interpretation of the relationship frame with other frames. The process of change in the video frame into a set of stories that shapes the text is a very personal knowledge each of which translated, so that it is necessary a standard knowledge on the system that was built. The process of cutting the frame is the most crucial to the next step, because the cutting process is expected to be automated (currently still rely on manual with the personal ability). Identification of sound and image are in the process of development methods, which may be able to collaborate with other research. From the results of the experiment can be conclude that the use of movement-oriented methods with a standard scenario of knowledge provides a more straightforward and assertive story. While in non-standard knowledge and the quality of service words are more compact, dense and precise.

Development of QoS, a priority for further development, this standard provides for the cutting measurement the right. To measure the success of the design algorithm index, conducted the simulation / testing query data on the original database with a database that

Acknowledgment

Thanks to Prof. Kokou Yetongnon, from Le2i CNRS Université de Bourgogne, Dijon, France.

References

[1] F. Long, Fundamentals Of Content-Based Image Retrieval, research.microsoft.com, China, 2003

[2] N. Sharda**,** Movement Oriented Design a New Paradigm for Multimedia Design, International Journal of Lateral Computing, Vol. 1 No. 1, May 2005

[3] W. I. Grosky, Managing Multimedia Information in Database Sistem, Communication of ACM, 1997

[4] D. Berardi, D. Calvanese, D. Giacomo, G., Lenzerini, M., Macella, M., Automatic service composition based on behavioral descriptions. International Journal of Cooperative Information systems 14 (4), 333-376, 2005

[5] M.A. Phillip, C. Huntley, Dramatica: A New Theory of Story, 2001, Screenplay Systems Incorporated, California

[6] D.T. Altılar1, Y. Paker, Minimum Overhead Data Partitioning Algorithms for Parallel Video Processing, 12th International Conference on Domain Decomposition Methods, 2001

[7] 7. M. Windhouwer, A. Schmidt, M. Kersten, Acoi: A System for Indexing Multimedia Objects, CWI NL-1098 SJ Amsterdam Netherlands, 2000

[8] S. Chaudhuri, and L. Gravano, Optimizing queries over multimedia repositories. In *Proceedings of SIGMOD '96* (Montreal, Canada, June 1996). ACM Press, New York, 1996.

[9] J. L. Lemke, Multiplying Meaning: Visual and Verbal Semiotics in Scientific Text, in J. R. Martin & R. Veel (Eds.), *Reading Science*, London: Routledge. 1998

[10] Hanandi, M., & Grimaldi, M. (2010). Organizational and collaborative knowledge management : a Virtual HRD model based on Web2 . 0. International Journal of Advanced Computer Science and Applications.

AUTHORS PROFILE

A Muslim, Graduate from Master Program in Information System, Gunadarma Unviversity, 1997. He is a Ph.D-Student at Groupe Database Sistem Information et Image, Le2i, UMR CNRS 5158, Faculté de Science et L'Enginer, Université de Bourgogne, Dijon, France

A.B. Mutiara is a Professor of Computer Science at Faculty of Computer Science and Information Technology, Gunadarma University

A Survey on Attacks and Defense Metrics of Routing Mechanism in Mobile Ad hoc Networks

K.P.Manikandan
HOD/MCA
Chirala Engineering College
Chirala-523157.A.P. India
+919908847047
manikandankp@yahoo.com

Dr.R.Satyaprasad
CSE Department
Achariya Nagarjuna University
Nagarjuna Nager-522 510,A.P.,India
+919848487478
profrsp@gmail.com

Dr.K.Rajasekhararao
PRINCIPAL
KL University
Vadeshwaram-522502,A.P.,India
+919848452344
krr_it@yahoo.co.in

***Abstract*-A Mobile Ad hoc Network (MANET) is a dynamic wireless network that can be formed infrastructure less connections in which each node can act as a router. The nodes in MANET themselves are responsible for dynamically discovering other nodes to communicate. Although the ongoing trend is to adopt ad hoc networks for commercial uses due to their certain unique properties, the main challenge is the vulnerability to security attacks. In the presence of malicious nodes, one of the main challenges in MANET is to design the robust security solution that can protect MANET from various routing attacks. Different mechanisms have been proposed using various cryptographic techniques to countermeasure the routing attacks against MANET. As a result, attacks with malicious intent have been and will be devised to exploit these vulnerabilities and to cripple the MANET operations. Attack prevention measures, such as authentication and encryption, can be used as the first line of defense for reducing the possibilities of attacks. However, these mechanisms are not suitable for MANET resource constraints, i.e., limited bandwidth and battery power, because they introduce heavy traffic load to exchange and verifying keys. In this paper, we identify the existent security threats an ad hoc network faces, the security services required to be achieved and the countermeasures for attacks in routing protocols. To accomplish our goal, we have done literature survey in gathering information related to various types of attacks and solutions. Finally, we have identified the challenges and proposed solutions to overcome them. In our survey, we focus on the findings and related works from which to provide secure protocols for MANETs. However, in short, we can say that the complete security solution requires the prevention, detection and reaction mechanisms applied in MANET.**

Keywords- MANET; Routing Protocol; Security Attacks; Routing Attacks and Defense Metrics

I. INTRODUCTION

A Mobile Ad hoc Network (MANET) is an autonomous system that consists of a variety of mobile hosts forming temporary network without any fixed infrastructure. Since it is difficult to dedicate routers and other infrastructure in such network, all the nodes are self-organized and collaborated to each other. All the nodes as well as the routers can move about freely and thus the network topology is highly dynamic. Due to self-organize and rapidly deploy capability, MANET can be applied to different applications including battlefield communications, emergency relief scenarios, law enforcement, public meeting, virtual class room and other security-sensitive computing environments. There are 15 major issues and sub-issues involving in MANET [10] such as routing, multicasting/broadcasting, location service, clustering, mobility management, TCP/UDP, IP addressing, multiple access, radio interface, bandwidth management, power management, security, fault tolerance, QoS/multimedia, and standards/products. The routing protocol sets an upper limit to security in any packet network. If routing can be misdirected, the entire network can be paralyzed. The problem is enlarged by the fact that routing usually needs to rely on the trustworthiness of all the nodes that are participating in the routing process. It is hard to distinguish compromised nodes from nodes that are suffering from bad links.

The main objective of this paper is to discuss ad hoc routing security with respect to the area of application.

II. ROUTING PROTOCOLS IN MANETS

Many protocols have been proposed for MANETs. These protocols can be divided into three categories: *proactive, reactive, and hybrid*. Proactive methods maintain routes to all nodes, including nodes to which no packets are sent. Such methods react to topology changes, even if no traffic is affected by the changes. They are also called table-driven methods. Reactive methods are based on demand for data transmission. Routes between hosts are determined only when they are explicitly needed to forward packets. Reactive methods are also called on-demand methods. They can significantly reduce routing overhead when the traffic is lightweight and the topology changes less dramatically, since they do not need to update route information periodically and do not need to find and maintain routes on which there is no traffic. Hybrid methods combine proactive and reactive methods to find efficient routes, without much control overhead.

A. Proactive Routing Protocols

Proactive MANET protocols are table-driven and will actively determine the layout of the network. Table-driven routing protocols attempt to maintain consistent, up-to-date routing information from each node to every other node in the network. These protocols require each node to maintain one or more tables to store routing information, and they respond to changes in network topology propagating updates throughout the network in order to maintain a consistent network view.

The areas in which they differ are the number of necessary routing-related tables and the methods by which changes in network structure are broadcast. Examples of proactive MANET protocols include Optimized Link State Routing (OLSR)[11], Topology Broadcast based on Reverse Path Forwarding (TBRPF)[12], Fish-eye State Routing (FSR)[13], Destination-Sequenced Distance Vector (DSDV)[14], Landmark Routing Protocol (LANMAR)[15], Cluster head Gateway Switch Routing Protocol (CGSR)[16].

B. Reactive Routing Protocols

Reactive protocols are on- demand protocols, create routes only when desired by source nodes. When a node requires a route to destination, it initiates route discovery process within the network. This process is completed once a route is found or all possible route permutations are examined. Once a route is discovered and established, it is maintained by route maintenance procedure until either destination becomes inaccessible along every path from source or route is no longer desired. Examples of reactive MANET protocols include Ad Hoc On-Demand Distance Vector (AODV) [17], Dynamic Source Routing (DSR) [18], Temporally Ordered Routing Algorithm (TORA) [19], and Dynamic MANET On Demand (DYMO) [20].

C. Hybrid Routing Protocols

Since proactive and reactive routing protocols each work best in oppositely different scenarios, there is good reason to develop hybrid routing protocols, which use a mix of both proactive and reactive routing protocols. These hybrid protocols can be used to find a balance between the proactive and reactive protocols.

The basic idea behind hybrid routing protocols is to use proactive routing mechanisms in some areas of the network at certain times and reactive routing for the rest of the network. The proactive operations are restricted to a small domain in order to reduce the control overheads and delays. The reactive routing protocols are used for locating nodes outside this domain, as this is more bandwidth-efficient in a constantly changing network. Examples of hybrid routing protocols include Core Extraction Distributed Ad Hoc Routing Protocol (CEDAR) [21], Zone Routing Protocol (ZRP) [22], and Zone Based Hierarchical Link State Routing Protocol (ZHLS) [23].

D. Proactive vs. Reactive vs. Hybrid Routing

The tradeoffs between proactive and reactive routing strategies are quite complex. Which approach is better depends on many factors, such as the size of the network, the mobility, the data traffic and so on. Proactive routing protocols try to maintain routes to all possible destinations, regardless of whether or not they are needed. Routing information is constantly propagated and maintained. In contrast, reactive routing protocols initiate route discovery on the demand of data traffic. Routes are needed only to those desired destinations. This routing approach can dramatically reduce routing overhead when a network is relatively static and the active traffic is light. However, the source node has to wait until a route to the destination can be discovered, increasing the response time.

The hybrid routing approach can adjust its routing strategies according to a network's characteristics and thus provides an attractive method for routing in MANETs. However, a network's characteristics, such as the mobility pattern and the traffic pattern, can be expected to be dynamic. The related information is very difficult to obtain and maintain.

This complexity makes dynamically adjusting routing strategies hard to implement.

III. TYPES OF SECURITY ATTACKS

External attacks, in which the attacker aims to cause congestion, propagate fake routing information or disturb nodes from providing services. Internal attacks, in which the adversary wants to gain the normal access to the network and participate the network activities, either by some malicious impersonation to get the access to the network as a new node, or by directly compromising a current node and using it as a basis to conduct its malicious behaviors.

The security attacks in MANET can be roughly classified into two major categories, namely passive attacks and active attacks. The active attacks further divided according to the layers.

A. Passive Attacks

A passive attack does not disrupt the normal operation of the network; the attacker snoop's the data exchanged in the network without altering it. Here the requirement of confidentiality gets violated. Detection of passive attack is very difficult since the operation of the network itself doesn't get affected. One of the solutions to the problem is to use powerful encryption mechanism to encrypt the data being transmitted, and thereby making it impossible for the attacker to get useful information from the data overhead.

1) Eavesdropping

Eavesdropping is another kind of attack that usually happens in the mobile ad hoc networks. It aims to obtain some confidential information that should be kept secret during the communication. The information may include the location, public key, private key or even passwords of the nodes. Because such data are very important to the security state of the nodes, they should be kept away from the unauthorized access.

2) Traffic Analysis & Monitoring

Traffic analysis attack adversaries monitor packet transmission to infer important information such as a source, destination, and source-destination pair.

B. Active Attacks

An active attack attempts to alter or destroy the data being exchanged in the network there by disrupting the normal functioning of the network. Active attacks can be internal or external. External attacks are carried out by nodes that do not belong to the network. Internal attacks are from compromised nodes that are part of the network. Since the attacker is already part of the network, internal attacks are more severe and hard to detect than external attacks.

Active attacks, whether carried out by an external advisory or an internal compromised node involves actions such as impersonation, modification, fabrication and replication.

IV. Types Of Routing Attacks And Defense Metrics On Manet

A. Routing Attacks

An attacker can absorb network traffic, inject themselves into the path between the source and destination and thus control the network traffic flow. For example, as shown in the *fig 1(a) and (b)*, a malicious node ***M*** can inject itself into the routing path between sender ***S*** and receiver ***R***.

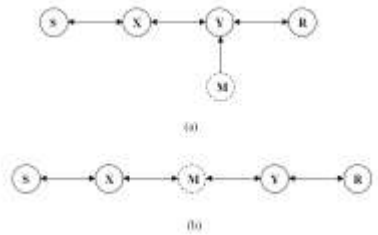


Figure 1: Routing attack

Network layer vulnerabilities fall into two categories: routing attacks and packet forwarding attacks [5]. The family of routing attacks refers to any action of advertising routing updates that does not follow the specifications of the routing protocols. The specific attack behaviors are related to the routing protocol used by the MANET.

The venomous routing nodes can attacks in MANET using dissimilar ways, so that, the following subsections are discussed various issues of routing attacks and its defense methods to mitigating route security attacks vulnerability on MANET.

B. Types of Routing Attacks

1) Routing Table Overflow Attack

This attack is basically happens to proactive routing algorithms, which update routing information periodically. To launch routing table overflow attack, the attacker tries to create routes to nonexistent nodes to the authorized nodes present in the network. It can simply send excessive route advertisements to overflow the target system's routing table. The goal is to have enough routes so that creation of new routes is prevented or the implementation of routing protocol is overwhelmed.

2) Routing Table Poisoning

Here, the compromised nodes in the networks send fictitious routing updates or modify genuine route update packets sent to other uncompromised nodes. Routing table poisoning may result in suboptimal routing, congestion in portions of the network, or even make some parts of the network inaccessible.

3) Packet Replication

In this attack, an adversary node replicates stale packets. This consumes additional bandwidth and battery power resources available to the nodes and also causes unnecessary confusion in the routing process.

4) Rushing Attack

On-demand routing protocols that use duplicate suppression during the route discovery process are vulnerable to this attack. An adversary node which receives a Route Request packet from the source node floods the packet quickly throughout the network before other nodes which also receive the same Route Request packet can react. Nodes that receive the legitimate Route Request packets assume those packets to be duplicates of the packet already received through the adversary node and hence discard those packets. Any route discovered by the source node would contain the adversary node as one of the intermediate nodes. Hence, the source node would not be able to find secure routes, that is, routes that do not include the adversary node. It is extremely difficult to detect such attacks in ad hoc wireless networks [3].

5) Routing Cache Poisoning Attack

Routing cache poisoning attack uses the advantage of the promiscuous mode of routing table updating. This occurs when information stored in routing tables is either deleted, altered or injected with false information. Suppose a malicious node ***M*** wants to poison routes node to ***X***. ***M*** could broadcast spoofed packets with source route to ***X*** via ***M*** itself, thus neighboring nodes that overhear the packet may add the route to their route caches [6].

C. Attacks on Specific Routing Protocol

There are many attacks in MANET that target the specific routing protocols. This is due to developing routing services without considering security issues. Most of the recent research suffers from this problem. In this section, we describe about the security threats, advantage and disadvantage of some common routing protocols.

1) AODV

The Ad-hoc On-demand Distance Vector (AODV) routing algorithm is a reactive algorithm that routes data across wireless mesh networks. The advantage of AODV is that it is simple, requires less memory and does not create extra traffic for communication along existing links. In AODV [2], the attacker may advertise a route with a smaller distance metric than the original distance or advertise a routing update with a large sequence number and invalidate all routing updates from other nodes.

2) DSR

Dynamic Source Routing (DSR) protocol is similar to AODV in that it also forms route on-demand. But the main difference is that it uses source routing instead of relying on the routing table at each intermediate node. It also provides functionality so that packets can be forwarded on a hop-by-hop basis. In DSR, it is possible to modify the source route listed in the RREQ or RREP packets by the attacker. Deleting a node from the list, switching the order or appending a new node into the list is also the potential dangers in DSR.

3) ARAN

Authenticated Routing for Ad-hoc Networks (ARAN) is an on-demand routing protocol that detects and protects against malicious actions carried out by third parties and peers in particular ad-hoc environment [4]. This protocol introduces authentication, message integrity and non-repudiation as a part of a minimal security policy. Though ARAN is designed to enhance ad-hoc security, still it is immune to rushing attack.

4) ARIADNE

ARIADNE is an on-demand secure ad-hoc routing protocol based on DSR that implements highly efficient symmetric cryptography. It provides point-to-point authentication of a routing message using a message authentication code (MAC) and a shared key between the two communicating parties. Although ARIADNE is free from a flood of RREQ packets and cache poisoning attack, but it is immune to the wormhole attack and rushing attack [3].

5) SEAD

Specifically, SEAD builds on the DSDV-SQ version of the DSDV (Destination Sequenced Distance Vector) protocol. It deals with attackers that modify routing information and also with replay attacks and makes use of one-way hash chains rather than implementing expensive asymmetric cryptography operations. Two different approaches are used for message authentication to prevent the attackers. SEAD does not cope with wormhole attacks [3].

D. Advanced Attacks

1) Wormhole attack

An attacker records packets at one location in the network and tunnels them to another location. Routing can be disrupted when routing control messages are tunneled. This tunnel between two colluding attackers is referred as a wormhole. Wormhole attacks are severe threats to MANET routing protocols. For example, when a wormhole attack is used against an on-demand routing protocol such as DSR or AODV, the attack could prevent the discovery of any routes other than through the wormhole [1].

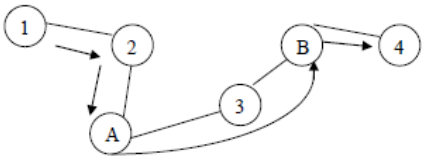


Figure 2: Wormhole attack

2) Blackhole attack

The blackhole attack has two properties. First, the node exploits the mobile ad hoc routing protocol, such as AODV [1], to advertise itself as having a valid route to a destination node, even though the route is spurious, with the intention of intercepting packets. Second, the attacker consumes the intercepted packets without any forwarding. However, the attacker runs the risk that neighboring nodes will monitor and expose the ongoing attacks. There is a more subtle form of these attacks when an attacker selectively forwards packets. An attacker suppresses or modifies packets originating from some nodes, while leaving the data from the other nodes unaffected, which limits the suspicion of its wrongdoing.

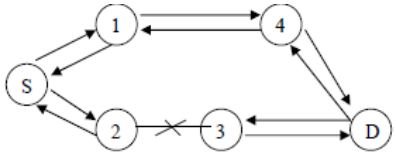


Figure 3: Blackhole attack [1]

3) Byzantine attack

A compromised intermediate node works alone, or a set of compromised intermediate nodes works in collusion and carry out attacks such as creating routing loops, forwarding packets through non-optimal paths, or selectively dropping packets, which results in disruption or degradation of the routing services [3].

4) Rushing Attack

In wormhole attack, two colluded attackers form a tunnel to falsify the original route. If luckily the transmission path is fast enough (e.g. a dedicated channel) then the tunneled packets can propagate faster than those through a normal multi-hop route, and result in the rushing attack. Basically, it is another form of denial of service (DoS) attack that can be launched against all currently proposed on-demand MANET routing protocols such as ARAN and Ariadne [3] [8].

5) Resource Consumption Attack

Energy is a critical parameter in the MANET. Battery-powered devices try to conserve energy by transmitting only when absolutely necessary [7]. The target of resource consumption attack is to send request of excessive route discovery or unnecessary packets to the victim node in order to consume the battery life. An attacker or compromised node thus can disrupt the normal functionalities of the MANET. This attack is also known as sleep deprivation attack.

6) Location Disclosure Attack

Location disclosure attack is a part of the information disclosure attack. The malicious node leaks information regarding the location or the structure of the network and uses the information for further attack. It gathers the node location information such as a route map and knows which nodes are situated on the target route. Traffic analysis is one of the unsolved security attacks against MANETs.

E. Defense Metrics against Routing Attacks in MANET

Network layer is more vulnerable to attacks than all other layers in MANET. A variety of security threats is imposed in this layer. Use of secure routing protocols provides the first line of defense. The active attack like modification of routing messages can be prevented through source authentication and message integrity mechanism. For example, digital signature, message authentication code (MAC), hashed MAC (HMAC), one-way HMAC key chain is used for this purpose. By an unalterable and independent physical metric such as time delay or geographical location can be used to detect wormhole attack. For example, packet leashes are used to combat this attack [9]. IPSec is most commonly used on the network layer in internet that could be used in MANET to provide certain level of confidentiality. The secure routing protocol named ARAN protects from various attacks like modification of sequence number, modification of hop counts, modification of source routes, spoofing, fabrication of source route etc [4]. The research by Deng [7], et al presents a solution to overcome blackhole attack. The solution is to disable the ability to reply in a message of an intermediate node, so all reply messages should be sent out only by the destination node.

V. RELATED WORK

The major task of the routing protocol is to discover the topology to ensure that each node can acquire a recent map of the network to construct routes to its destinations. Several efficient routing protocols have been proposed for MANET.

These protocols generally fall into one of the two major categories: reactive routing protocols and proactive routing protocols. In reactive routing protocols, such as Ad hoc On Demand Distance Vector (AODV) protocol [2], nodes find routes only when they must send data to the destination node whose route is unknown. In contrast, in proactive routing protocols, such as OLSR [11], nodes obtain routes by periodic exchange of topology information with other nodes and maintain route information all the time. Based on the behavior of attackers, attacks against MANET can be classified into passive or active attacks. Attacks can be further categorized as either outsider or insider attacks. With respect to the target, attacks could be also divided into data packet or routing packet attacks. In routing packet attacks, attackers could not only prevent existing paths from being used, but also spoof non-existing paths to lure data packets to them. Several studies [1],[2],[24], [25], [26] have been carried out on modeling MANET routing attacks. Typical routing attacks include black-hole, fabrication, and modification of various fields in routing packets (route request message, route reply message, route error message, etc.). Some research efforts have been made to seek preventive solutions [8], [27],[28] for protecting the routing protocols in MANET. Although these approaches can prevent unauthorized nodes from joining the network, they introduce a significant overhead for key exchange and verification with the limited intrusion elimination. Besides, prevention-based techniques are less helpful for defending from malicious insiders who possess the credentials to communicate in the network.

Numerous intrusion detection systems (IDS) for MANET have been recently introduced. Due to the nature of MANET, most IDS are structured to be distributed and have a cooperative architecture. Similar to signatured-based and anomaly based IDS models for wired network; IDS for MANET use specification-based approaches and statistics-based approaches. Specification-based approaches, for example DEMEM [30], C. Tseng et al. [29] and M. Wang et al. [32], monitor network activities and compare them with known attack features, which are impractical to cope with new attacks. On the other hand, statistics-based approaches, such as Watchdog and Lipad , compare network activities with normal behavior patterns, which result in higher false positives rate than specification-based ones. Because of the existence of false positives in both MANET IDS models, intrusion alerts from these systems always accompany with alert confidence, which indicates the possibility of attack occurrence. Intrusion response systems (IRS) for MANET are inspired by MANET IDS [31] isolate malicious nodes based on their reputations. Their work fails to take advantage of IDS alerts and simple isolation of nodes may cause unexpected network partition [33] brings the concept of cost-sensitive into MANET intrusion response which considers topology dependency and attack damage. The advantage of our solution is that we integrate evidences from IDS, local routing table with expert knowledge to estimate risk of attacks, and countermeasures with a mathematical reasoning approach.

VI. CONCLUSION

Mobile Ad Hoc Networks have the ability to setup networks on the fly in a harsh environment where it may not possible to deploy a traditional network infrastructure. Whether ad hoc networks have vast potential, still there are many challenges left to overcome. Security is an important feature for deployment of MANET. In this paper, we have overviewed the challenges and solutions of the routing security threats in mobile ad hoc networks. Of these attacks, the passive attacks do not disrupt the operation of a protocol, but is only information seeking in nature whereas active attacks disrupt the normal operation of the MANET as a whole by targeting specific node(s). In this survey, we reviewed the current state of the art routing attacks and countermeasures MANETs. The advantages as well as the drawbacks of the countermeasures have been outlined. It has been observed that although active research is being carried out in this area, the proposed solutions are not complete in terms of effective and efficient routing security. There are limitations on all solutions. They may be of high computational or communication overhead (in case of cryptography and key management based solutions) which is detrimental in case of resource constrained MANETS, or of the ability to cope with only single malicious node and ineffectiveness in case of multiple colluding attackers. Furthermore, most of the proposed solutions can work only with one or two specific attacks and are still vulnerable to unexpected attacks. A number of challenges like the Invisible Node Attack remain in the area of routing security of MANETs. Future research efforts should be focused not only on improving the effectiveness of the security schemes but also on minimizing the cost to make them suitable for a MANET environment.

REFERENCES

[1] Pradip M. Jawandhiya et. al. / International Journal of Engineering Science and Technology Vol. 2(9), 2010, 4063-4071.

[2] Nishu Garg and R.P.Mahapatra, "MANET Security Issues ," IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.8, August 2009.

[3] Ping Yi, Yue Wu and Futai Zou and Ning Liu, "A Survey on Security in Wireless Mesh Networks", Proceedings of IETE Technical Review, Vol. 27, Issue 1, Jan-Feb 2010.

[4] K. Sanzgiri, B. Dahill, B.N. Levine, C. Shields, E.M. Belding-Royer, "*Secure routing protocol for ad hoc networks,"* In Proc. of 10th IEEE International Conference on Network Protocols, Dept. of Comput. Sci., California Univ., Santa Barbara, CA, USA. 12-15 Nov. 2002, Page(s): 78- 87, ISSN: 1092-1648.

[5] H. Yang, H. Luo, F. Ye, S. Lu, L. Zhang, "*Security in mobile ad hoc networks: challenges and solutions,"* In proc. IEE Wireless Communication, UCLA, Los Angeles, CA, USA; volume- 11, Page(s):38- 47, ISSN: 1536-1284

[6] B. Wu, J. Chen, J. Wu, M. Cardei, "*A Survey of Attacks and Countermeasures in Mobile Ad Hoc Networks,"* Department of Computer Science and Engineering, Florida Atlantic University, http://student.fau.edu/jchen8/web/papers/SurveyBookchapter.pdf

[7] H. Deng, W. Li, Agrawal, D.P., "*Routing security in wireless ad hoc networks,"* Cincinnati Univ., OH, USA; IEEE Communications Magazine, Oct. 2002, Volume: 40, page(s): 70- 75, ISSN: 0163-6804

[8] Y. Hu, A. Perrig, and D. Johnson, "*Ariadne: A Secure On-Demand Routing for Ad Hoc Networks,"* Proc. of MobiCom 2002, Atlanta, 2002.

[9] Y. Hu, A. Perrig, and D. Johnson, *"Packet Leashes: A Defense Against Wormhole Attacks in Wireless Ad Hoc Networks,"* Proc. of IEEE INFORCOM, 2002.

[10] C. R. Dow, P. J. Lin, S. C. Chen*, J. H. Lin*, and S. F. Hwang. A Study of Recent Research Trends and Experimental Guidelines in Mobile. Ad-hoc Networks. 19th International Conference on *Advanced Information Networking and Applications, 2005. AINA 2005***,** *Volume: 1, On page(s): 72- 77 vol.1*.

[11] T.H.Clausen, G.Hansen, L.Christensen, and G.Behrmann, "The Optimized Link State Routing Protocol, Evaluation Through Experiments and Simulation," *Proceedings of IEEE Symposium on Wireless Personal Mobile Communications 2001*, September 2001.

[12] R. Ogier, F. Templin, M. Lewis, "Topology Dissemination Based on Reverse-Path Forwarding (TBRPF)", *IETF Internet Draft*, v.11, October 2003.

[13] A.Iwata, C.C.Chiang, G.Pei, M.Gerla and T.W.Chen, "Scalable Routing Strategies for Ad Hoc Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 8, pp. 1369-1379, August 1999.

[14] C.E.Perkins and P.Bhagwat, "Highly Dynamic Destination- Sequenced Distance-Vector Routing (DSDV) For Mobile Computers," *Proceedings of ACM SIGCOMM 1994*, pp. 233-244, August 1994.

[15] M.Gerla, X.Hong, L.Ma and G.Pei, "Landmark Routing Protocol (LANMAR) for Large Scale Ad Hoc Networks", *IETF Internet Draft*, v.5, November 2002.

[16] C.C.Chiang, H.K.Wu, W.Liu and M.Gerla, "Routing in Clustered Multi Hop Mobile Wireless Networks with Fading Channel," *Proceedings of IEEE SICON 1997*, pp. 197-211, April 1997.

[17] C.E.Perkins and E.M.Royer, "Ad Hoc On-Demand Distance Vector Routing," *Proceedings of IEEE Workshop on Mobile Computing Systems and Applications 1999*, pp. 90-100, February 1999.

[18] D.B.Jhonson and D.A.Maltz, "Dynamic Source Routing in Ad Hoc Wireless Networks," *Mobile Computing*, Kluwer Academic Publishers, vol.353, pp. 153-181, 1996.

[19] V.D.Park and M.S.Corson, "A Highly Adaptive Distributed Routing Algorithm for Mobile Ad Hoc Networks," *Proceedings of IEEE INFOCOM 1997*, pp. 1405-1413, April 1997.

[20] I. Chakeres and C. Perkins, "Dynamic MANET On-demand (DYMO) Routing Rrotocol", *IETF Internet Draft*, v.15, November 2008, (Work in Progress).

[21] P.Sinha, R.Sivakumar and V.Bharghavan, "CEDAR: A Core Extraction Distributed Ad Hoc Routing Algorithm," *IEEE Journal on Selected Areas in Communications*, vol.17, no.8, pp. 1454-1466, August 1999.

[22] Z.J.Haas, "The Routing Algorithm for the Reconfigurable Wireless Networks," *Proceedings of ICUPC 1997*, vol. 2, pp. 562-566, October 1997.

[23] M.Joa-Ng and I.T.Lu, "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, pp. 1415-1425, August 1999.

[24] B. Kannhavong, H. Nakayama, N. Kato, A. Jamalipour, and Y. Nemoto.

A study of a routing attack in OLSR-based mobile ad hoc networks. *International Journal of Communication Systems*, 20(11):1245–1261, 2007.

[25] B. Kannhavong, H. Nakayama, N. Kato, Y. Nemoto, and A. Jamalipour.

A collusion attack against olsr-based mobile ad hoc networks. In *GLOBECOM*, 2006.

[26] B. Kannhavong, H. Nakayama, Y. Nemoto, N. Kato, and A. Jamalipour.

A Survey of routing attacks in mobile ad hoc networks. *IEEE Wireless Communications*, page 86, 2007.

[27] Y. Hu, D. Johnson, and A. Perrig. SEAD: secure efficient distance vector routing for mobile wireless ad hoc networks. *Ad Hoc Networks*, 1(1):175–192, 2003.

[28] Y. Hu, A. Perrig, and D. Johnson. Ariadne: A Secure On-Demand Routing Protocol for Ad Hoc Networks. *Wireless Networks*, 11(1):21–38, 2005.

[29] C. Tseng, T. Song, P. Balasubramanyam, C. Ko, and K. Levitt. A Specification-Based Intrusion Detection Model for OLSR.

LECTURE NOTES IN COMPUTER SCIENCE, 3858:330, 2006.

[30] C. Tseng, S. Wang, C. Ko, and K. Levitt. Demem: Distributed evidence driven message exchange intrusion detection model for manet.

In *Recent Advances in Intrusion Detection*, pages 249–271. Springer, 2006.

[31] T. View. Information theoretic framework of trust modeling and evaluation for ad hoc networks. *Selected Areas in Communications, IEEE Journal on*, 24(2):305–317, 2006.

[32] M.Wang, L. Lamont, P. Mason, and M. Gorlatova. An effective intrusion

detection approach for OLSR MANET protocol. In *Secure Network Protocols, 2005.(NPSec). 1st IEEE ICNP Workshop on*, pages 55–60, 2005.

[33] S. Wang, C. Tseng, K. Levitt, and M. Bishop. Cost-Sensitive Intrusion Responses for Mobile Ad Hoc Networks. *LECTURE NOTES IN COMPUTER SCIENCE*, 4637:127, 2007.

[34] Ramasamy, U. (2010). Classification of Self-Organizing Hierarchical Mobile Adhoc Network Routing Protocols - A Summary. International Journal of Advanced Computer Science and Applications - IJACSA, 1(4).

[35] Nayak, C. K., Dash, G. K. A. K., & Das, S. (2011). Detection of Routing Misbehavior in MANETs with 2ACK scheme. International Journal of Advanced Computer Science and Applications - IJACSA, 2(1), 126-129.

Advanced Steganography Algorithm using Encrypted secret message

Joyshree Nath
A.K.Chaudhuri School of IT
University of Calcutta
Kolkata, India
joyshreenath@gmail.com

Asoke Nath
Department of Computer Science
St. Xavier's College(Autonomous)
Kolkata, India
asokejoy@gmail.com

***Abstract*—In the present work the authors have introduced a new method for hiding any encrypted secret message inside a cover file. For encrypting secret message the authors have used new algorithm proposed by Nath et al(1). For hiding secret message we have used a method proposed by Nath et al (2). In MSA (1) method we have modified the idea of Play fair method into a new platform where we can encrypt or decrypt any file. We have introduced a new randomization method for generating the randomized key matrix to encrypt plain text file and to decrypt cipher text file. We have also introduced a new algorithm for encrypting the plain text multiple times. Our method is totally dependent on the random text_key which is to be supplied by the user. The maximum length of the text_key can be of 16 characters long and it may contain any character (ASCII code 0 to 255). We have developed an algorithm to calculate the randomization number and the encryption number from the given text_key. The size of the encryption key matrix is 16x16 and the total number of matrices can be formed from 16 x 16 is 256! which is quite large and hence if someone applies the brute force method then he/she has to give trail for 256! times which is quite absurd. Moreover the multiple encryption method makes the system further secured. For hiding secret message in the cover file we have inserted the 8 bits of each character of encrypted message file in 8 consecutive bytes of the cover file. We have introduced password for hiding data in the cover file. We propose that our new method could be most appropriate for hiding any file in any standard cover file such as image, audio, video files. Because the hidden message is encrypted hence it will be almost impossible for the intruder to unhide the actual secret message from the embedded cover file. This method may be the most secured method in digital water marking.**

Keywords- Steganography; MSA algorithm; Encryption; Decryption;

I. INTRODUCTION

In the present work we have used two (2) methods: (i) We encrypt the secret message(SM) using a method MSA(Meheboob,Saima and Asoke) proposed by Nath et al.(1). (ii) We insert the encrypted secret message inside the standard cover file(CF) by changing the least significant bit(LSB). Nath et al(2) already proposed different methods for embedding SM into CF but there the SF was inserted as it is in the CF and hence the security of steganography was not very high. In the present work we have basically tried to make the steganography method more secured. It means even if someone can extract SM from CF but he cannot be able to decrypt the message as he has to know the exact decryption method. In our present work we try to embed almost any type of file inside some standard cover file (CF) such as image file(.JPEG or .BMP) or any image file inside another image file. Here first we will describe our steganography method for embedding any type of file inside any type of file and then we will describe the encryption method which we have used to encrypt the secret message and to decrypt the extracted data from the embedded cover file.

(i) Least Significant Bit Insertion method: Least significant bit (LSB) insertion is a common, simple approach for embedding information in a cover image. The LSB or in other words 8-th bit of some or all the bytes inside an image is changed to a bit of the secret message. Let us consider a cover image contains the following bit patterns:

Byte-1	Byte-2	Byte-3	Byte-4
00101101	00011100	11011100	10100110

Byte-5	Byte-6	Byte-7	Byte-8
11000100	00001100	11010010	10101101

Suppose we want to embed a number 200 in the above bit pattern. Now the binary representation of 200 is 11001000. To embed this information we need at least 8 bytes in cover file. We have taken 8 bytes in the cover file. Now we modify the LSB of each byte of the cover file by each of the bit of embed text 11001000.

Now we want to show what happens to cover file text after we embed 11001000 in the LSB of all 8 bytes:

TABLE 1 CHANGING LSB

Before Replacement	After Replacement	Bit inserted	Remarks
00101101	0010110**1**	1	No change in bit pattern
00011100	0001110**1**	1	Change in bit pattern(i)
11011100	1101110**0**	0	No change in bit pattern
10100110	1010011**0**	0	No change in bit pattern
11000100	1100010**1**	1	Change in bit pattern(ii)
00001100	0000110**0**	0	No change in bit pattern
11010010	1101001**0**	0	No change in bit pattern

10101101	10101100	0	Change in bit pattern(iii)

Here we can see that out of 8 bytes only 3 bytes get changed only at the LSB position. Since we are changing the LSB hence we are either changing the corresponding character in forward direction or in backward direction by only one unit and depending on the situation there may not be any change also as we have seen in the above example. As our eye is not very sensitive so therefore after embedding a secret message in a cover file our eye may not be able to find the difference between the original message and the message after inserting some secret text or message on to it. To embed secret message we have to first skip 600 bytes from the last byte of the cover file. After that according to size of the secret message (say n bytes) we skip 8*n bytes. After that we start to insert the bits of the secret file into the cover file. Under no circumstances the size of the cover should not be less the 10*sizeof(secret message) then our method will fail. For extracting embedded file from the cover file we have to perform the following:

One has to enter the password while embedding a secret message ile. If password is correct then the program will read the file size from the cover file. Once we get the file size we follow simply the reverse process of embedding a file in the cover file. We read LSB of each byte and accumulate 8 bits to form a character and we immediately write that character on to a file.

We made a long experiment on different types of host files and also the secret messages and found the following combinations are most successful:

Table-2 COVER FILE TYPE AND SECRET MESSAGE FILE TYPE

Sl.No.	Cover file type	Secret file type used
1	.BMP	.BMP,.DOC,.TXT,.WAV,.MP3,.XLS,.PPT, .AVI,.JPG,.EXE..COM
2.	.JPG	.JPG,.BMP,.TXT,.WAV,.MP3,.XLS,,.PPT, .EXE,.COM
3.	.DOC	.TXT
4.	.WAV	.BMP,.JPG,.TXT,.DOC
5.	.AVI	.TXT,.WAV,.JPEG
6.	.PDF	.TXT

Only in case of .PDF and .JPEG file to insert secret message is a bit difficult job as those files are either compressed or encrypted. Even then we got success for inserting a small text file in .PDF file. So we can conclude that .PDF file is not a good cover file. On the other hand .BMP file is the most appropriate file which can be used for embedding any type of file without facing any problem.

(ii) Meheboob, Saima and Asoke(MSA) Symmetric key Cryptographic method:

Symmetric key cryptography is well known concept in modern cryptography. The plus point of symmetric key cryptography is that we need one key to encrypt a plain text and the same key can be used to decrypt the cipher text and the main problem is that the same key is used for encryption as well as decryption process. Hence the key must be secured. Because of this problem we have introduced public key cryptography such as RSA public key method. RSA method has got both merits as well as demerits. The problem of Public key cryptosystem is that we have to do massive computation for encrypting any plain text. Sometimes these methods may not be also suitable such as in sensor networks. However, there are quite a number of encryption methods have come up in the recent past appropriate for the sensor nodes. Nath et al.(1) proposed a symmetric key method where they have used a random key generator for generating the initial key and that key is used for encrypting the given source file. MSA method is basically a substitution method where we take 2 characters from any input file and then search the corresponding characters from the random key matrix and store the encrypted data in another file. In our work we have the provision for encrypting message multiple times. The key matrix contains all possible characters (ASCII code 0 to 255) in a random order. The pattern of the key matrix will depend on text_key entered by the user. Nath et al. proposed algorithm to obtain randomization number, encryption number and the shift parameter from the initial text_key. We have given a long trial run on text_key and we found that it is very difficult to match the three above parameters for 2 different Text_key which means if someone wants to break our encryption method then he/she has to know the exact pattern of the text_key otherwise it will not be possible to obtain two sets of identical parameters from two different text_key. We have given several trial runs to break our encryption method but we found it is almost unbreakable. For pure text file we can apply brute force method to decrypt small text but for any other file such as binary file we cannot apply any brute force method.

II. RANDOM KEY GENERATION AND MSA ENCRYPTION ALGORITHM

Before we embed the secret message in a cover file we first encrypt the secret message using MSA method. The detail method is given in our previous publication (1). Here we will describe the MSA algorithm in brief:

To create Random key Matrix of size(16x16) we have to take any key. The size of key must be less than or equal to 16 characters long. These 16 characters can be any of the 256 characters(ASCII code 0 to 255). The relative position and the character itself is very important in our method to calculate the randomization number , the encryption number and the relative shift of characters in the starting key matrix. We take an example how to calculate randomization number, the encryption number and relative shift from a given key. Here we are demonstrating our method:

Suppose key=AB

Choose the following table for calculating the place value and the power of characters of the incoming key:

TABLE-3

Length of key(n)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Base value(b)	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2

$$\text{Step-1: Sum}=\sum_{m=1}^{n} \text{ASCII Code} * b^{m} \quad \text{----(1)}$$

Example-1:

Now we calculate the sum for key="**AB**" using equation(1)

$\text{Sum}=65*16^1 + 66 * 16^2 = 17936$

Now we have to calculate 3 parameters from this sum (i) Randomization number(n1), (ii) Encryption number(n2) and (iii)Relative shift(n3) using the following method:

a) Randomization number(n1):

num1=1*1+7*2+9*3+3*4+6*5=84

n1=sum mod num1=17936 mod 84=**44**

Note: if n1=0 then n1=num1 and n1<=128

b) Encryption number(n2):

num2=6*1+3*2+9*3+7*4+1*5=72

n2=sum mod num2 =17936 mod 72 =**8**

Note: if n2=0 then n2=num2 and n2<=64

c) Relative shift(n3):

n3= Σall digits in sum=1+7+9+3+6=**26**

Now we show the original key matrix(16 x 16) which contains all characters(ASCII code 0-255):

TABLE –4 : THE ORIGINAL MATRIX:

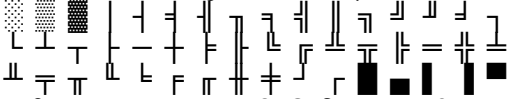
☺ ☻ ♥ ♦ ♣ ♠
♫ ☼
► ◄ ↕ ‼ ¶ § ▬ ↨ ↑ ↓ ← ∟ ↔ ▲ ▼
! " # $ % & ' ( ) * + , - . /
0 1 2 3 4 5 6 7 8 9 : ; < = > ?
@ A B C D E F G H I J K L M N O
P Q R S T U V W X Y Z [\ ] ^ _
` a b c d e f g h i j k l m n o
p q r s t u v w x y z { | } ~ ⌂
Ç ü é â ä à å ç ê ë è ï î ì Ä Å
É æ Æ ô ö ò û ù ÿ Ö Ü ¢ £ ¥ Pts *f*
á í ó ú ñ Ñ ª º ¿ ⌐ ¬ ½ ¼ ¡ « »

α ß Γ π Σ σ µ τ Φ Θ Ω δ ∞ φ ε ∩
≡ ± ≥ ≤ ⌠ ⌡ ÷ ≈ ° · · √ ⁿ ²

TABLE-5 : THE MATRIX AFTER RELATIVE SHIFT(N3=26) IS:

4 N h é £ ╣ ╩ Ω ☺ ← 5 O i â
¥ ╖ ╤ δ ☻ ∟ 6 P j ä Pts ╕ ╥ ∞ ♥ ↔
7 Q k à *f* ╣ ╚ φ ♦ ▲ 8 R l å á ╠
└ ε ♣ ▼ 9 S m ç í ╖ ╒ ∩ ♠ : T
n ê ó ╝ ╓ ≡ ! ; U o ë ú ╩ ╬ ±
" < V p è ñ ╛ ╪ ≥ # = W q ï
Ñ ╖ ┘ ≤
$ > X r î ª └ ┌ ⌠ ♂ %
? Y s ì º ┴ ▉ ⌡ ♀ & @ Z t Ä ¿ ┬
' A [u Å ⌐ ├ ▌ ≈ ♫ (B \
v É ¬ ─ ▐ ° ☼) C] w æ ½ ┼ ■ ·
► * D ^ x Æ ¼ ╞ α · ◄ + E _ y ô
¡ ╠ ß √ ↕ , F ` z ö « ╚ Γ ⁿ ‼ -
G a { ò » ╓ π ² ¶ . H b | û ░ ╩
Σ ■ § / I c } ù ▒ ╦ σ ═ 0 J d
~ ÿ ▓ ╠ µ ↨ 1 K e ⌂ Ö ▏ ═ τ ↑ 2
L f Ç Ü ┤ ╬ Φ ↓ 3 M g ü ¢ ╡ ╧ Θ

Now we apply the following randomization methods one after another in a serial manner:

Step-1: Function cycling()

Step-2: Function upshift()

Step-3: Function downshift()

Step-4:Function leftshift()

Step-5:Function rightshift()

Step-6:Function random()

Step-7:Function random_diagonal_right()

Step-8:Function random_diagonal_left()

For detail randomization methods we refer to our previous work(1).

After finishing above shifting process we perform

a) column randomization
b) row randomization and
c) diagonal rotation and
d) reverse diagonal rotation

Each operation will continue for n3 number of times.

Now we apply encryption process on any text file. Our encryption process is as follows:

We choose a 4X4 simple key matrix:

TABLE-6

A	B	C	D
E	F	G	H
I	J	K	L
M	N	O	P

Case-I: Suppose we want to encrypt **FF** then it will take as **GG** which is just one character after F in the same row.

Case –II: Suppose we want to encrypt **FK** where **F** and **K** appears in two different rows and two different columns. **FK** will be encrypted to **KH (FK→GJ→HK→KH).**

Case-III: Suppose we want to encrypt **EF** where **EF** occurs in the same row. Here **EF** will be converted to **HG**

III. CHANGING LSB BITS OF COVER FILE USING ENCRYPTED SECRET MESSAGE FILE

In the present work we have made an exhaustive study on embedding (i) text, (ii)sound, (iii)image in different cover files such as image file, sound file, word document file, .PDF file. The size of the cover file must be at least 10-times more than secret message file which is to be embedded within the cover file. The last 500 bytes of the cover file we reserved for storing the password and the size of the secret message file. After that we subtract n*(size of the secret message file) from the size of the cover file. Here n=8 depending on how many bytes we have used to embed one byte of the secret message file in the cover file. For strong password we have used a simple algorithm as follows: We take XOR operation with each byte of the password with 255 and insert it into the cover file. To retrieve the password we read the byte from the cover file and apply XOR operation with 255 to get back original password. To embed any any secret message we have to enter the password and to extract message we have to enter the same password. The size of the secret message file we convert into 32 bits binary and then convert it into 4 characters and write onto cover file. When we want to extract encrypted secret message from a cover file then we first extract the file size from the cover file and extract the same amount of bytes from cover file. Now we will describe the algorithms which we have used in our present study:

We read one byte at a time from the encrypted secret message file(ESMF) and then we extract 8 bits from that byte. After that we read 8 consecutive bytes from the cover file(CF). We check the LSB of each byte of that 8 byte chunk whether it is different from the bits of ESMF. If it different then we replace that bit by the bit we obtain from the ESMF. Our program also counts how many bits we change and how many bytes we change and then we also calculate the percentage of bits changed and percentage of bytes changed in the CF. Now we will demonstrate in a simple case.:

Suppose we want to embed "A" in the cover text "BBCDEFGH". Now we will show how this cover text will be modified after we insert "A" within it.

TABLE -7 CHANGING LSB

Original Text	Bit string	Bit to be inserted in LSB	Changed Bit string	Changed Text
B	01000010	0	0100001**0**	B
B	01000010	1	0100001**1**	C
C	01000011	0	0100001**0**	B
D	01000100	0	0100010**0**	D
E	01000101	0	0100010**0**	D
F	01000110	0	0100011**0**	F
G	01000111	0	0100011**0**	F
H	01001000	1	0100100**1**	I

Here we can see that to embed "A" we modify 5 bits out of 64 bits. After embedding "A" in cover text "BBCDEFGH" the cover text converts to "BCBDDFFI". We can see that the change in cover text is prominent as we are trying to embed text within text which is actually not possible using LSB method. But when we do it in some image or audio file then it will not be so prominent.

To extract byte from the cover file we follow the reverse process which we apply in case of encoding the message. We simply extract serially one by one from the cover file and then we club 8 bits and convert it to a character and then we write it to another file. But this extracted file is now in encrypted form and hence we apply decryption process which will be the reverse of encryption process to get back original secret message file.

IV. RESULTS AND DISCUSSION

Case-1: Cover File type=.jpg Secret File type=.jpg

Fig_1:Cover file name: sxcn.jpg Size=1155378 Bytes

+

Fig_2:Secret message File:joy1.jpg Size=1870 Bytes

=

Fig_3: Embedded Cover file name :sxcn.jpg Size=1155378 Bytes

(secret message encrypted before embedding)

Case-2: Cover File type=.BMP secret message file =.doc

Fig_4: Cover File name : tvshow.bmp Size=688856 Bytes.

Fig_5: Embedded Cover File name : tvshow.bmp Size=688856 Bytes
In this file an encrypted word file xxfile2.doc(size=19456B) is embedded

Case-3: Cover File type=.BMP secret message file =.jpg

Fig_6: Cover file name = tvshow1.bmp (size=688856B)

+

Fig_7: Secret message file= tuktuk1.jpg(size=50880B) (The secret message file was Encrypted while embedding)

=

Fig_8: Embedded cover file name=tvshow1.bmp (size=688856B)

Case-4: Cover File type=..AVI(Movie File) secret message file =.jpg

Fig_9: Cover File Name= Name=rhinos.avi(size=76800B)

+

Fig_10: Secret message File name = tuktuk_bw.jpg (Size=1870B)

=

Fig_11:Embedded Cover File Name=rhionos.avi. (Size=76800B)
(The encrypted secret message file is embedded)

V. CONCLUSION

In the present work we try to embed some secret message inside any cover file in encrypted form so that no one will be able to extract actual secret message. Here we use the standard steganographic method i.e. changing LSB bits of the cover file. Our encryption method can use maximum encryption number=64 and maximum randomization number=128. The key matrix may be generated in 256! Ways. So in principle it will be difficult for anyone to decrypt the encrypted message without knowing the exact key matrix. Our method is essentially stream cipher method and it may take huge amount of time if the files size is large and the encryption number is also large. The merit of this method is that if we change the key_text little bit then the whole encryption and decryption process will change. This method may most suitable for water marking. The steganography method may be further secured if we compress the secret message first and then encrypt it and then finally embed inside the cover file.

VI. ACKNOWLEDGEMENT

AN sincerely expresses his gratitude to Department of Computer Science for providing necessary help and assistance. AN is also extremely grateful to University Grants Commission for providing fund for continuing minor research project on Data encryption using symmetric key and public key crypto system. JN is grateful to A.K. Chaudhury School of I.T. for giving inspiration for research work.

REFERENCES

[1] Symmetric key cryptography using random key generator, A.Nath , S.Ghosh, M.A.Mallik, Proceedings of International conference on SAM-2010 held at Las Vegas(USA) 12-15 July,2010, Vol-2,P-239-244

[2] Data Hiding and Retrieval, A.Nath, S.Das, A.Chakrabarti, Proceedings of IEEE International conference on Computer Intelligence and Computer Network held at Bhopal from 26-28 Nov, Page-392-397, 2010.

[3] Advanced steganographic approach for hiding encrypted secret message in LSB ,LSB+1,LSB+2 and LSB+3 bits in non standard cover files, Joyshree Nath, Sankar Das, Shalabh Agarwal and Asoke Nath , to be published in IJCA(USA),Vol 14-No.7, P-31-35, February 2011.

[4] Cryptography and Network , William Stallings , Prectice Hall of India

[5] Modified Version of Playfair Cipher using Linear Feedback Shift Register, P. Murali and Gandhidoss Senthilkumar, UCSNS International journal of Computer Science and Network Security, Vol-8 No.12, Dec 2008.

[6] Jpeg20000 Standard for Image Compression Concepts algorithms and VLSI Architectures by Tinku Acharya and Ping-Sing Tsai, Wiley Interscience.

[7] Steganography and Seganalysis by Moerland, T , Leiden Institute of Advanced Computing Science.

[8] SSB-4 System of Steganography using bit 4 by J.M.Rodrigues Et. Al.

[9] An Overview of Image Steganography by T.Morkel, J.H.P. Eloff and M.S.Oliver.

[10] An Overview of Steganography by Shawn D. Dickman

[11] Hamdy, S., El-messiry, H., Roushdy, M., & Kahlifa, E. (2010). Quantization Table Estimation in JPEG Images. International Journal of Advanced Computer Science and Applications - IJACSA, 1(6), 17-23.

[12] Lalith, T. (2010). Key Management Techniques for Controlling the Distribution and Update of Cryptographic keys. International Journal of Advanced Computer Science and Applications - IJACSA, 1(6), 163-166.

[13] Hajami, A., & Elkoutbi, M. (2010). A Council-based Distributed Key Management Scheme for MANETs. International Journal of Advanced Computer Science and Applications - IJACSA, 1(3).

[14] Meshram, C. (2010). Modified ID-Based Public key Cryptosystem using Double Discrete Logarithm Problem. International Journal of Advanced Computer Science and Applications - IJACSA, 1(6). Retrieved from http://ijacsa.thesai.org/.

Transparent Data Encryption- Solution for Security of Database Contents

Dr. Anwar Pasha Abdul Gafoor Deshmukh[1]
[1]College of Computers and Information Technology, University of Tabuk, Tabuk 71431, Kingdom of Saudi Arabia
[1]mdanwardmk@yahoo.com

Dr. Riyazuddin Qureshi [2]
[2]Community College, Jazan University, Jazan, Kingdom of Saudi Arabia
[2]riaz_quraysh21@yahoo.com

***Abstract*— The present study deals with Transparent Data Encryption which is a technology used to solve the problems of security of data. Transparent Data Encryption means encrypting databases on hard disk and on any backup media. Present day global business environment presents numerous security threats and compliance challenges. To protect against data thefts and frauds we require security solutions that are transparent by design. Transparent Data Encryption provides transparent, standards-based security that protects data on the network, on disk and on backup media. It is easy and effective protection of stored data by transparently encrypting data. Transparent Data Encryption can be used to provide high levels of security to columns, table and tablespace that is database files stored on hard drives or floppy disks or CD's, and other information that requires protection. It is the technology used by Microsoft SQL Server 2008 to encrypt database contents. The term encryption means the piece of information encoded in such a way that it can only be decoded read and understood by people for whom the information is intended. The study deals with ways to create Master Key, creation of certificate protected by the master key, creation of database master key and protection by the certificate and ways to set the database to use encryption in Microsoft SQL Server 2008.**

Keywords- Transparent Data Encryption; TDE; Encryption; Decryption; Microsoft SQL Server 2008;

I. INTRODUCTION

Present day life is vastly driven by Information Technology. Extensive use of IT is playing vital role in decision-making in all commercial and non-commercial organizations. All activities are centered on data, its safe storage and manipulation. In present scenario organizations' survival is at stake if its data is misused. Data is vulnerable to a wide range of threats like, Weak Authentication, Backup Data Exposure, Denial of Service, etc.

This study is aimed at to deal with the most critical of those threats to which database is vulnerable. Transparent Data Encryption (TDE) shields database up to considerable extent against such threats. TDE is used to prevent unauthorized access to confidential database, reduce the cost of managing users and facilitate privacy managements. This latest technology arms users' i.e. database administrators to solve the possible threats to security of data. This technology allows encrypting databases on hard disk and on any backup media. TDE nowadays, is the best possible choice for bulk encryption to meet regulatory compliance or corporate data security standards.

Main purpose of Transparent Data Encryption is to provide security to columns, tables, Tablespace of database. In section 1, we explain about the Encryption, Plaintext, Cipher text (Encrypted Text) with simple examples and Types of Encryption. Section 2 explains Transparent Data Encryption, its Scope, Uses and its Limitations. Section 3 deals with Transparent Data Encryption in Microsoft Server 2008 with its architecture. Section 4 addresses issues like Create Database Using Microsoft SQL Server 2008 with coding and its description; Section 5 is for Creation of Encryption and Decryption of Database with coding and its descriptions respectively. Section 6 deals with conclusion.

II. ENCRYPTION

Encryption is said to occur when data is passed through a series of mathematical operations that generate an alternate form of that data; the sequence of these operations is called an algorithm. To help distinguish between the two forms of data, the unencrypted data is referred to as the plaintext and the encrypted data as cipher text. Encryption is used to ensure that information is hidden from anyone for whom it is not intended, even those who can see the encrypted data. The process of reverting cipher text to its original plaintext is called decryption. This process is illustrated in the Figure 1 below.

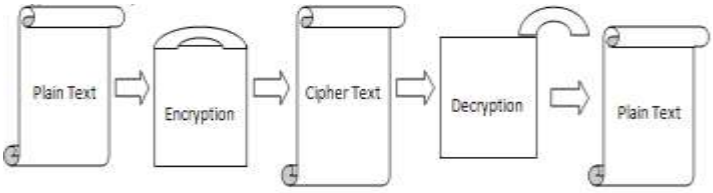


Figure 1 Process of Encryption and Decryption

A. Simple Example

- Encryption of the word "abcd" could result in "wxyz" Reversing the order of the letters in the plaintext generates the cipher text.
- It is Simple encryption and quit easy for attacker to retrieve the original data. A better solution of

encrypting this message is to create an alternative alphabet by shifting each letter by arbitrary number.

- Example: The word "abcd" is encrypted by shifting the alphabet by 3 letters to right then the result will be "defg". The "Ceasar Cipher" means exchange of letters or words with another.
- Normal Alphabet: abcdefghijklmnopqrstuvwxyz
- Alphabet shifted by 3: defghijklmnopqrstuvwxyzabc

B. Types of Encryptions

There are two general categories for key-based encryption

- Symmetric
- Asymmetric

1) Symmetric

Symmetric encryption uses a single key to encrypt and decrypt the message. To encrypt the message provide key to the recipient before decrypt it. To use symmetric encryption, the sender encrypts the message and, if the recipient does not already have a key, sends the key and cipher text separately to the recipient. The message then decrypt by recipient key. This method is easy and fast to implement but has weaknesses; for instance, if a hacker intercepts the key, he can also decrypt the messages. Single key encryptions tend to be easier for hacker /cracker. This means that the algorithm that is used to encode the message is easier for attackers to understand, enabling them to more easily decode the message.

2) Asymmetric encryption

Asymmetric encryption, also known as Public-Key encryption, uses two different keys - a public key to encrypt the message, and a private key to decrypt it. The public key is used to encrypt and private key to decrypt. One can easily distribute the public key to communicate because only with private key one can decrypt it. To protect the message between users the sender encrypts it by public key. Then receiver decrypts it by private key. Only recipient can decrypt the message in this type of encryptions.

III. TRANSPARENT DATA ENCRYPTION

Transparent data encryption is used for encryption and decryption of the data and log files. The encryption uses a Database Encryption Key (DEK), which is stored in the database boot record for availability during recovery. It is Asymmetric key secured by using a certificate stored in the master database. Transparent data encryption protects data and log files. It is a technology used to solve the problems of security of data means encrypting databases on hard disk and on any backup media. Transparent Data Encryption can be used to provide high levels of security to columns, table and Tablespace that is database files stored on hard drives or floppy disks or CD's, and other information that requires protection. It is the technology used by Microsoft SQL Server 2008 to encrypt database contents.

Transparent data encryption encrypts data before it's written to disk and decrypts data before it is returned to the application. The encryption and decryption process is performed at the SQL layer, completely transparent to applications and users. Subsequent backups of the database files to disk or tape will have the sensitive application data encrypted.

A. Scope

Extensive use of IT is playing vital role in decision-making in commercial as well as non-commercial organizations. In present scenario operations of any organization are badly affected if; data is misused. Data is vulnerable to a wide range of threats like, Excessive Privilege Abuse, Legitimate Privilege Abuse, Weak Authentication, Backup Data Exposure, etc.

The crucial role data plays in any organization itself explains its importance. So it is exposed to grave threats of theft, misuse or loss. This study is aimed to deal with the most critical of those threats to which database is vulnerable by focusing on Transparent Data Encryption (TDE). TDE is used to prevent unauthorized access to confidential database, reduce the cost of managing users, and facilitate privacy managements. This latest technology enables users' i.e. database administrators to counter the possible threats to security of data. TDE facilitates encrypting databases on hard disk and on any backup media. TDE nowadays, is the best possible choice for bulk encryption to meet regulatory compliance or corporate data security standards.

B. Use of Transparent data encryption

There are three important uses of Transparent data encryption as below

1) Authentication
2) Validation
3) Data Protection

1) Authentication

Unauthorized access to information is a very old problem. Business decisions today are driven by information gathered from mining terabytes of data. Protecting sensitive information is key to a business's ability to remain competitive. Access to key data repositories such as the Microsoft SQL Server 2008 that house valuable information can be granted once users are identified and authenticated accurately. Verifying user identity involves collecting more information than the usual user name and password. Microsoft SQL Server 2008 Advanced Security provides the ability for businesses to leverage their existing security infrastructures such as encrypt Master key, Database Master Key and Certificate.

2) Validation

Validation describes the ability to provide assurance that a senders identity is true and that a Column, Tablespace or file has not been modified. Encryption can be used to provide validation by making a digital Certificate of the information contained within a database. Upon validation, the user can be reasonably sure that the data came from a trusted person and that the contents of the data have not been modified.

3) Data protection

Probably the most widely-used application of transparent encryption is in the area of data protection. The information that a business owns is invaluable to its productive operation; consequently, the protection of this information is very important. For people working in small offices and home

offices, the most practical uses of transparent encryption for data protection are column, Tablespace and files encryption. This information protection is vital in the event of theft of the computer itself or if an attacker successfully breaks into the system. The encryption and decryption process is performed at the SQL layer, completely transparent to applications and users.

C. Limitation of Transparent data encryption

- Transparent data Protection does not provide encryption across communication channels.
- When enabling Transparent data Protection, you should immediately back up the certificate and the private key associated with the certificate. If the certificate ever becomes unavailable or if you must restore or attach the database on another server, you must have backups of both the certificate and the private key or you will not be able to open the database.
- The encrypting certificate or Asymmetric should be retained even if Transparent data Protection is no longer enabled on the database. Even though the database is not encrypted, the database encryption key may be retained in the database and may need to be accessed for some operations.
- Altering the certificates to be password-protected after they are used by Transparent data Protection will cause the database to become inaccessible after a restart.

IV. Transparent Data Encryption in Microsoft SQL Server 2008

In Microsoft SQL Server 2008 Transparent Data Encryption of the database file is performed at the page level. The pages in an encrypted database are encrypted before they are written to disk and decrypted when read into memory. Transparent data Protection does not increase the size of the encrypted database.

A. Architecture of Transparent Data Encryption

In Microsoft SQL Server 2008 Transparent Data Encryption first upon we need to Create a master key, then obtain a certificate protected by the master key after that Create a database encryption key and protect it by the certificate and Set the database to use encryption. The following Figure 2 shows the steps of Transparent Data Encryption

In Microsoft SQL Server 2008 Transparent Data Encryption of the database file is performed at the page level. The pages in an encrypted database are encrypted before they are written to disk and decrypted when read into memory. Service Master key is created at a time of SQL Server setup, DPAPI encrypts the Service Master key. Service Master key encrypts Database Master key for the Master Database. The Database Master key of the master Database Creates the Certificate then the certificate encrypts the database encryption key in the user database. The entire database is secured by the Database Master key of the user Database by using Transparent Database encryption.

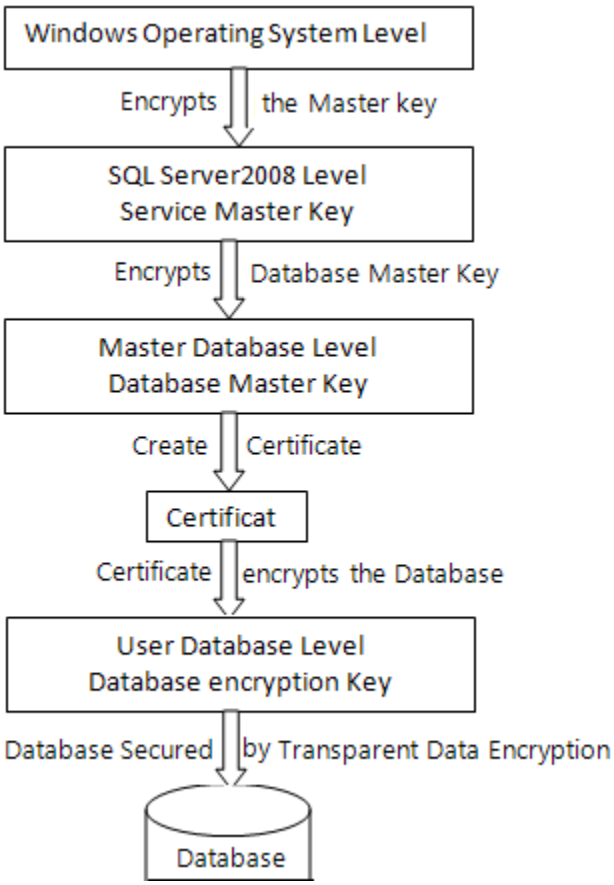


Figure 2 Microsoft SQL Server 2008 Transparent Data Encryption

V. Create Database Using Microsoft SQL Server 2008

Creating a database that specifies the data and transaction log files, the following code 1 to create database name "Sales"

A. Code 1

```
USE master;
GO
        CREATE DATABASE Sales
        ON
        ( NAME = Sales_dat,
        FILENAME = 'C:\Program Files\Microsoft SQL
      Server\MSSQL10_50.MSSQLSERVER\MSSQL\DA
      TA\saledat.mdf',
        SIZE = 10,
        MAXSIZE = 50,
        FILEGROWTH = 5 )
        LOG ON
        ( NAME = Sales_log,
         FILENAME = 'C:\Program Files\Microsoft SQL
      Server\MSSQL10_50.MSSQLSERVER\MSSQL\DA
      TA\salelog.ldf',
         SIZE = 5MB,
        MAXSIZE = 25MB,
        FILEGROWTH = 5MB ) ;
GO
```

B. Description of above Code 1

CREAT DATABASE command is used to create database in SQL Server 2008, Sales is a name of Database. In these example we have created one data file name "Sales_dat' and one lof file named "Sales_log" with specified size

VI. Creation of Encryption and Decryption of Database

The following code 2 illustrates encrypting and decrypting the Sales database using a certificate installed on the server named MySalesCert.

A. Code 2

```
USE master;
GO
        CREATE MASTER KEY ENCRYPTION BY
PASSWORD = '<writeanypasswordhere>';
        go
                CREATE CERTIFICATE
        MySalesCert WITH SUBJECT = 'It is my
        Certificate';
        go
        USE Sales;
        GO
                CREATE DATABASE ENCRYPTION
KEY
                WITH ALGORITHM = AES_128
                ENCRYPTION BY SERVER
CERTIFICATE MySalesCert;
        GO
        ALTER DATABASE Sales
        SET ENCRYPTION ON;
GO
```

B. Description of Code 2

The Command CREATE MASTER KEY ENCRYPTION BY PASSWORD is used to create Master key encryption with password here user can assign any password. CREATE CERTIFICATE is used to create certificate as MySalesCert, WITH SUBJECT is used to assign any subject as "It is my Certificate" for the database created in code 1 "Sales". The Command CREATE DATABASE ENCRYPTION KEY WITH ALGORITHM = AES_128 it is a encryption key. By using ENCRYPTION BY SERVER CERTIFICATE command assigning the certificate "MySalesCert" to the Server. By using the command SET ENCRYPTION keeping it ON.

VII. Conclusions

Transparent Data Encryption plays an especially important role in safeguarding data in transit. Microsoft SQL Server 2008 Transparent Data Encryption protects sensitive data on disk drives and backup media from unauthorized access, helping reduce the impact of lost or stolen media. The Transparent Data Encryption are developed under the present work has been successfully, created, implemented and thoroughly tested. We used this Transparent Data Encryption on few computer we found it to be very effective. The Transparent Data Encryption based on the study of various security measures were planned, designed and developed using techniques and tools described earlier. Transparent Data Encryption provides a highly configurable environment for application development. User can use Transparent Data Encryption to create both single-machine and networked environments in which developers can safely try out. Such a setup do not requires additional infrastructure and easily caters to the needs of beginners as well as advanced learners on one premise.

References

[1] Rob Walters, Christian Kirisch, "Database Encryption and key management for Microsoft SQL Server 2008", by Create Space, January 2010

[2] Micheal Otey, "Microsoft SQL Server 2008 New Features", McGraw Hill Osborn Media, 2 Edition , December 2008.

[3] Denny Cherry, "Securing SQL Server: Protecting your Database from Attackers", Syngress 1 Edition, Feburary 2011

[4] Carl Hamacher, Zvonko Vranesic and Safwat Zaky, " Computer Organization," McGraw hill, August 2001.

[5] Andrew S. Tennanbaum, "Computer Network," 5th Edition, Prentice Hall, October 2010.

[6] Micheal Coles, Rodney Landrum, "Expert SQL Server 2008 Encryption", Apress 1 Edition , October 2009

[7] Oracle Advanced Security ,"An Oracle White Paper", http://www.oracle.com/technetwork/database/security/advanced-security-11g-whitepaper-133635.pdf, June 2007

[8] MSDN Library, "Encryption Connection to SQL Server", http://msdn.microsoft.com/en-us/library/ms189067.aspx

Author profile

Dr. Anwar Pasha Deshmukh: Working as a Lecturer, College of Computer and Information Technology, University of Tabuk, Tabuk, Saudi Arabia. PhD (Computer Science) July 2010, MPhil (Information Technology) 2008, M Sc (Information Technolgy) 2003 from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad. India.

Dr Riyazuddin Qureshi: Working as a Assistant Professor, Community College, Jazan University,

Jazan, Kingdom of Saudi Arabia. PhD(Management Science) February 2008, M. Com May 1993, MCA May 1998, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad.

Knowledge discovery from database Using an integration of clustering and classification

Varun Kumar
Department of Computer Science and Engineering
ITM University
Gurgaon, India
kumarvarun333@gmail.com

Nisha Rathee
Department of Computer Science and Engineering
ITM University
Gurgaon, India
nisharathee29@gmail.com

***Abstract*— Clustering and classification are two important techniques of data mining. Classification is a supervised learning problem of assigning an object to one of several pre-defined categories based upon the attributes of the object. While, clustering is an unsupervised learning problem that group objects based upon distance or similarity. Each group is known as a cluster. In this paper we make use of a large database 'Fisher's Iris Dataset' containing 5 attributes and 150 instances to perform an integration of clustering and classification techniques of data mining. We compared results of simple classification technique (using J48 classifier) with the results of integration of clustering and classification technique, based upon various parameters using WEKA (Waikato Environment for Knowledge Analysis), a Data Mining tool. The results of the experiment show that integration of clustering and classification gives promising results with utmost accuracy rate and robustness even when the data set is containing missing values**.

Keywords- Data Mining; J48; KMEANS; WEKA; Fisher's Iris dataset;

I. INTRODUCTION

Data mining is the process of automatic classification of cases based on data patterns obtained from a dataset. A number of algorithms have been developed and implemented to extract information and discover knowledge patterns that may be useful for decision support [2]. Data Mining, also popularly known as Knowledge Discovery in Databases (KDD), refers to the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases [1]. Several data mining techniques are pattern recognition, clustering, association, classification and clustering [7]. The proposed work will focus on challenges related to integration of clustering and classification techniques. Classification has been identified as an important problem in the emerging field of data mining [5]. Given our goal of classifying large data sets, we focus mainly on decision tree classifiers [8] [9]. Decision tree classifiers are relatively fast as compared to other classification methods. A decision tree can be converted into simple and easy to understand classification rules [10].

Finally, tree classifiers obtained similar and sometimes better accuracy when compared with other classification methods [11]. Clustering is the unsupervised classification of patterns into clusters [6].The community of users has played lot emphasis on developing fast algorithms for clustering large datasets [14].It groups similar objects together in a cluster (or clusters) and dissimilar objects in other cluster (or clusters) [12]. In this paper WEKA (Waikato Environment for knowledge analysis) machine learning tool [13][18] is used for performing clustering and classification algorithms. The dataset used in this paper is Fisher's Iris dataset, consists of 50 samples from each of three species of Iris flowers (Iris setosa, Iris virginica and Iris versicolor). Four features were measured from each sample; they are the length and the width of sepal and petal, in centimeters. Based on the combination of the four features, Fisher developed a linear discriminant model to distinguish the species from each other.

A. Organisation of the paper

The paper is organized as follows: Section 2 defines problem statement. Section 3 describes the proposed classification method to identify the class of Iris flower as Iris-setosa, Iris-versicolor or Iris-virginica using data mining classification algorithm and an integration of clustering and classification technique of data mining. Experimental results and performance evaluation are presented in Section 4 and finally, Section 5 concludes the paper and points out some potential future work.

II. PROBLEM STATEMENT

The problem in particular is a comparative study of classification technique algorithm J48 with an integration of SimpleKMeans clusterer and J48 classifier on various parameters using Fisher's Iris Dataset containing 5 attributes and 150 instances.

III. PROPOSED METHOD

Classification is the process of finding a set of models that describe and distinguish data classes and concepts, for the purpose of being able to use the model to predict the class whose label is unknown. Clustering is different from classification as it builds the classes (which are not known in advance) based upon similarity between object features. Fig. 1 shows a general framework of an integration of clustering and classification process. Integration of clustering and classification technique is useful even when the dataset contains missing values. Fig. 2 shows the block diagram of

steps of evaluation and comparison. In this experiment, object corresponds to Iris flower, and object class label corresponds to species of Iris flower. Every Iris flower consists of length and width of petal and sepal, which are used to predict the species of Iris flower. Apply classification technique (J48 classifier) using WEKA tool. Classification is a two step process, first, it build classification model using training data. Every object of the dataset must be pre-classified i.e. its class label must be known, second the model generated in the preceding step is tested by assigning class labels to data objects in a test dataset.

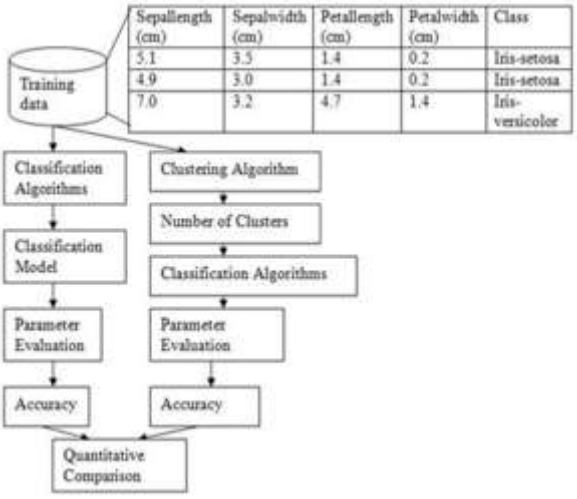


Figure 1. Proposed Classification Model

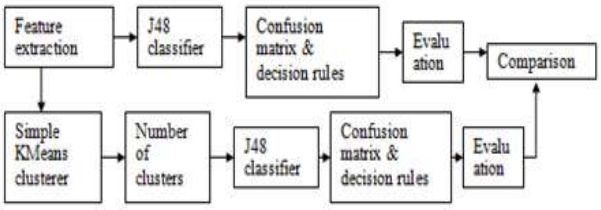


Figure 2. Block Diagram

The test data may be different from the training data. Every element of the test data is also preclassified in advance. The accuracy of the classification model is determined by comparing true class labels in the testing set with those assigned by the model. Apply clustering technique on the original data set using WEKA tool and now we are come up with a number of clusters. It also adds an attribute 'cluster' to the data set. Apply classification technique on the clustering result data set. Then compare the results of simple classification and an integration of clustering and classification. In this paper, we identified the finest classification rules through experimental study for the task of classifying Iris flower type in to Iris setosa, Iris versicolor, or Iris virginica species using Weka data mining tool.

A. Iris Dataset Preprocessing

We make use of a database 'Fisher's Iris dataset' containing 5 attributes and 150 instances to perform comparative study of data mining classification algorithm namely J48(C4.5) and an integration of Simple KMeans clustering algorithm and J48 classification algorithm. Prior to indexing and classification, a preprocessing step was performed. The Fisher's Iris Database is available on UCI Machine Learning Repository website http://archive.ics.uci.edu:80/ml/datasets.html in Excel Format i.e. .xls file. In order to perform experiment using WEKA [20], the file format for Iris database has been changed to .arff or .csv file.

The complete description of the of attribute value are presented in Table 1. A sample training data set is also given in Table 2 .During clustering technique we add an attribute i.e. 'cluster' to the data set and use filtered clusterer with SimpleKMeans algorithms which removes the use of 5,6 attribute during clustering and add the resulting cluster to which each instance belongs to, along with classes to the dataset.

TABLE I. COMPLETE DESCRIPTION OF VARIABLES

Variable/Attributes	Category	Possible Values
Sepallength	Numeric	4-8
Sepalwidth	Numeric	1-5
Petallength	Numeric	1-8
Petalwidth	Numeric	0-3
Class	Nominal	Three, Iris-setosa, Iris-virginica & Iris versicolor

TABLE II. SAMPLE OF INSTANCES FROM

sepal length	Sepal Width	Petal Length	Petal width	class
5.1	3.5	1.4	0.2	Iris-setosa
4.9	3.0	1.4	0.2	Iris-setosa
4.7	3.2	1.3	0.2	Iris-setosa
4.6	3.1	1.5	0.2	iris-setosa
5.0	3.6	1.4	0.2	Iris-setosa
7.0	3.2	4.7	1.4	Iris-versicolor
6.4	3.2	4.5	1.5	iris-versicolor
7.6	3.0	6.6	2.1	Iris-virginica

B. Building Classifiers

1) J48(C4.5): J48 is an implementation of C4.5[17] that builds decision trees from a set of training data in the same way as ID3, using the concept of Information Entropy. . The training data is a set S = s1, s2... of already classified samples. Each sample si = x1, x2... is a vector where x1, x2… represent attributes or features of the sample. Decision tree are efficient to use and display good accuracy for large amount of data. At each node of the tree, C4.5 chooses one attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other.

2) KMeans clusterer: Simple KMeans is one of the simplest clustering algorithms [4].KMeans algorithm is a classical clustering method that group large datasets in to clusters[15][16]. The procedure follows a simple way to

classify a given data set through a certain number of clusters. It select k points as initial centriods and find K clusters by assigning data instances to nearest centroids. Distance measure used to find centroids is Euclidean distance.

3) Measures for performance evaluation: To measure the performance, two concepts sensitivity and specificity are often used; these concepts are readily usable for the evaluation of any binary classifier. TP is true positive, FP is false positive, TN is true negative and FN is false negative. TPR is true positive rate, it is equivalent to Recall.

$$sensitivity = TPR = \frac{TP}{TP + FN} \quad (1)$$

$$specificity = \frac{TN}{FP + TN} \quad (2)$$

a) Confusion Matrix: Fig. 3 shows the confusion matrix of three class problem .If we evaluate a set of objects, we can count the outcomes and prepare a confusion matrix (also known as a contingency table), a three-three (as Iris dataset contain three classes) table that shows the classifier's correct decisions on a major diagonal and the errors off this diagonal.

		predicted		
		positive link	negative link	non-existent link
actual	positive link	TP	FP	FN
	negative link	FP	TP	FN
	non-existent link	FP	FP	TN

Figure 3. Confusion Matrix

The columns represent the predictions and the rows represent the actual class [3]. An edge is denoted as true positive (TP), if it is a positive or negative link and predicted also as a positive or negative link, respectively. False positives (FP) are all predicted positive or negative links which are not correctly predicted, i.e., either they are non-existent or they have another sign in the reference network. As true negatives (TN) we denote correctly predicted non-existent edges and as false negatives (FN) falsely predicted non-existent edges are defined i.e., an edge is predicted to be non-existent but it is a positive or a negative link in the reference network.

b) Precision: In information retrieval positive predictive value is called precision. It is calculated as number of correctly classified instances belongs to X divided by number of instances classified as belonging to class X; that is, it is the proportion of true positives out of all positive results. It can be defined as:

$$precision = \frac{TP}{TP + FP} \quad (3)$$

c) Accuracy: It is simply a ratio of ((no. of correctly classified instances) / (total no. of instances)) *100). Technically it can be defined as:

$$accuracy = \frac{TP + TN}{(TP + FN) + (FN + TN)} \quad (4)$$

d) False Positive Rate : It is simply the ratio of false positives to false positives plus true negatives. In an ideal world we want the FPR to be zero. It can be defined as:

$$FPR = \frac{FP}{FP + TN} \quad (5)$$

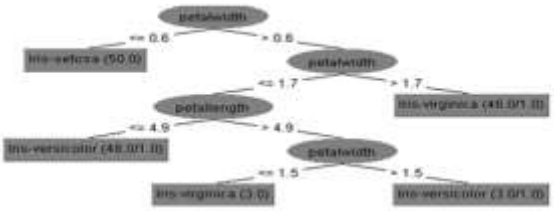
e) F-Meaure: F-measure is a way of combining recall and precision scores into a single measure of performance. The formula for it is:

$$\frac{2*recall*precision}{recall+precision} \quad (6)$$

IV. Experiment Results And Performance Evaluation

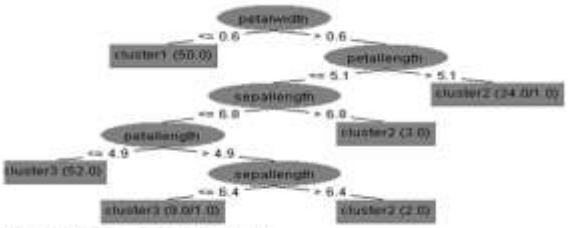
In this experiment we present a comparative study of classification technique of data mining with an integration of clustering and classification technique of data mining on various parameters using Fisher's Iris dataset containing 150 instances and 5 attributes. During simple classification, the training dataset is given as input to WEKA tool and the classification algorithm namely C4.5 (implemented in WEKA as J48) was implemented. During an integration of clustering and classification techniques of data mining first, Simple KMeans clustering algorithm was implemented on the training data set by removing the class attribute from the data set as clustering technique is unsupervised learning and then J48 classification algorithm was implemented on the resulting dataset.

The results of the experiment show that integration of clustering and classification technique gives a promising result with utmost accuracy rate and robustness among the classification and clustering algorithms (Table 3). An experiment measuring the accuracy of binary classifier based on true positives, false positives, false negatives, and true negatives (as per Equation 4), decision trees and decision tree rules are shown in Table 3 and Fig. 4 &5.



```
petalwidth <= 0.6: Iris-setosa (50.0)
petalwidth > 0.6
|   petalwidth <= 1.7
|   |   petallength <= 4.9: Iris-versicolor (48.0/1.0)
|   |   petallength > 4.9
|   |   |   petalwidth <= 1.5: Iris-virginica (3.0)
|   |   |   petalwidth > 1.5: Iris-versicolor (3.0/1.0)
|   petalwidth > 1.7: Iris-virginica (46.0/1.0)
```

Figure 4. Decision tree and Rules during classification of Iris data



```
petalwidth <= 0.6: cluster1 (50.0)
petalwidth > 0.6
|   petallength <= 5.1
|   |   sepallength <= 6.8
|   |   |   petallength <= 4.9: cluster3 (52.0)
|   |   |   petallength > 4.9
|   |   |   |   sepallength <= 6.4: cluster3 (9.0/1.0)
|   |   |   |   sepallength > 6.4: cluster2 (2.0)
```

Figure 7. Decision tree and Rules of integration of clustering and classification technique

TABLE III. PERFORMANCE EVALUATION

Parameters	C4.5 (J48)	Simple KMeans+J48
TP	96	88
FP	4	1
TN	47	60
FN	3	1
Size of tree	9	11
Leaves in tree	5	6
Error Rate	0.1713	0.0909
Accuracy	95. 33%	98.6667%

A. *Observations and Analysis*

- It may be observed from Table 3 that the error rate of binary classifier J48 with Simple KMeans Clusterer is lowest i.e. 0.0909 in comparison with J48 classifier without clusterer i.e. 0.1713, which is most desirable.
- Accuracy of J48 classifier with KMeans clusterer is high i.e. 98.6667% (Table 3), which is highly required.
- Sensitivity (TPR) of clusters (results of integration of classification and clustering technique) is higher than that of classes (Table 4, 5 &6).
- In an ideal world we want the FPR to be zero. Considering results presented in Table 4, 5&6, FPR is lowest of integration of clustering and classification technique, in other words closet to the zero as compared with simple classification technique with J48 classifier.
- In an ideal world we want precision value to be 1. Precision value is the proportion of true positives out of all positive results. Precision value of integration of classification and clustering technique is higher than that of simple classification with J48 classifier (Table 4, 5&6).

TABLE IV. IRIS SETOSA CLASS AND CLUSTER 1

Parameters	J48 (iris-setosa)	SimpleKMeans+ J48 (cluster1)
Precision	1	1
Recall/Sensitivity (TPR)	0.98	1
Specificity (TNR)	0.9215	0.9836
F-measure	0.99	1
FPR	0	0

TABLE V. IRIS VERGINICA CLASS AND CLUSTER 2

Parameters	J48 (iris-setosa)	SimpleKMeans+ J48 (cluster1)
Precision	1	1
Recall/Sensitivity (TPR)	0.98	1
Specificity (TNR)	0.9215	0.9836
F-measure	0.99	1
FPR	0	0

TABLE VI. IRIS VERSICOLOR CLASS AND CLUSTER 3

Parameters	J48 (iris-versicolor)	SimpleKMeans+ J48 (cluster2)
Precision	0.922	0.987
Recall/Sensitivity (TPR)	0.94	0.987
F-measure	0.931	0.987
FPR	0.04	0.007

According to the experiments and result analysis presented in this paper, it is observed that an integration of classification and clustering technique is better to classify datasets with better accuracy.

V. CONCLUSION AND FUTURE WORK

A comparative study of data mining classification technique and an integration of clustering and classification technique helps in identifying large data sets. The presented experiments shows that integration of clustering and

classification technique gives more accurate results than simple classification technique to classify data sets whose attributes and classes are given to us. It can also be useful in developing rules when the data set is containing missing values. As clustering is an unsupervised learning technique therefore, it build the classes by forming a number of clusters to which instances belongs to, and then by applying classification technique to these clusters we get decision rules which are very useful in classifying unknown datasets. We can then assigns some class names to the clusters to which instance belongs to. This integrated technique of clustering and classification gives a promising classification results with utmost accuracy rate and robustness. In future we will perform experiments with other binary classifiers and try to find the results from the integration of classification, clustering and association technique of data mining.

REFERENCES

[1] J. Han and M. Kamber, "Data Mining: Concepts and Techniques," Morgan Kaufmann, 2000.

[2] Desouza, K.C. (2001) ,Artificial intelligence for healthcare management In Proceedings of the First International Conference on Management of Healthcare and Medical Technology Enschede, Netherlands Institute for Healthcare Technology Management.

[3] http://www.comp.leeds.ac.uk/andyr

[4] I. K. Ravichandra Rao , "Data Mining and Clustering Techniques," DRTC Workshop on Semantic Web 8th – 10th December, 2003,DRTC, Bangalore ,Paper: K

[5] Rakesh Agrawal,Tomasz Imielinski and Arun Swami," Data mining : A Performance perspective ". IEEE Transactions on Knowledge and Data Engineering , 5(6):914-925, December 1993.

[6] Jain, A.K., Murty M.N., and Flynn P.J. (1999):" Data Clustering: A Review".

[7] Ritu Chauhan, Harleen Kaur, M.Afshar Alam, "Data Clustering Method for Discovering Clusters in Spatial Cancer Databases", International Journal of Computer Applications (0975 – 8887) Volume 10– No.6, November 2010.

[8] NASA Ames Res.Ctr. INTRO. IND Version 2.1, GA23-2475-02 edition, 1992

[9] J.R Quinlan and R..L. Rivest. Inferring decision tress using minium description length principle. Information and computation ,1989.

[10] J.R Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufman, 1993.

[11] D.Michie , D.J Spiegehalter, and C.C Taylor. Machine Learning, Neural and Statistical Classification. Ellis horword, 1994.

[12] Paul Agarwal, M.Afsar Alam, Ranjit Biswas. Analysing the agglomeratve hierarchical Clustering Algorithm for Categorical Attributes. International Journal of Innovation, Management and Technology , vol.1,No.2 , June 2010, ISSN: 2010- 0248.

[13] Weka 3- Data Mining with open source machine learning software available from :- http://www.cs.waikato. ac.nz/ml/ weka/

[14] U.M Fayyad and P. Smyth. Advances in Knowledge Discovery and Data Mining. AAAI/MIT Presss, Menlo Park, CA, 1996.

[15] L.Kaufinan, and P.J Rousseeuw, Finding groups in Data: an Introduction to Cluster Analysis, John Wiley & Sons 1990.

[16] M.S Chen, J.Han, and P.S.Yu. Data mining: an overview from database perspective . IEEE Trans. On Knowledge and Data Engineering , 5(1) : 866-833, Dec 1996.

[17] Shuyan wang , Mingquan Zhou and guohua geng ," Application of Fuzzy cluster analysis for Medical Image Data Mining " proceedings of the IEEE International Conference on Mechatronics & Automation Niagra falls ,Canada, July 2005.

[18] Holmes, G.,Donkin,A.,Witten,I.H.: "WEKA a machine learning workbench". In: Proceeding second Australia and New Zeeland Conference on Intelligent Information System, Brisbane , Australia, pp.357-361 (1994).

[19] Kim, H. and Loh , W. –Y.2001, Classification trees with unbiased multiway splits, Journal of the American Stastical Association, vol. 96, pp. 589- 604.

[20] Garner S.R. (1995) "WEKA: The Waikato Environment for Knowledge Analysis Proc" New Zealand Computer Science Research Students Conference, University of Waikato, Hamilton, New Zealand, pp 57-64.

[21] The Result Oriented Process for Students Based On Distributed Data Mining. (2010). International Journal of Advanced Computer Science and Applications - IJACSA, 1(5), 22-25.

[22] Saritha, S. J., Govindarajulu, P. P., Prasad, K. R., Rao, S. C. V. R., Lakshmi, C., Prof, A., et al. (2010). Clustering Methods for Credit Card using Bayesian rules based on K-means classification. International Journal of Advanced Computer Science and Applications - IJACSA, 1(4), 2-5.

[23] Saritha, S. J., Govindarajulu, P. P., Prasad, K. R., Rao, S. C. V. R., Lakshmi, C., Prof, A., et al. (2010). Clustering Methods for Credit Card using Bayesian rules based on K-means classification. International Journal of Advanced Computer Science and Applications - IJACSA, 1(4), 2-5.

[24] Firdhous, M. F. M. (2010). Automating Legal Research through Data Mining. International Journal of Advanced Computer Science and Applications - IJACSA, 1(6), 9-16.

[25] Takale, S. A. (2010). Measuring Semantic Similarity between Words Using Web Documents. International Journal of Advanced Computer Science and Applications - IJACSA, 1(4), 78-85.

AUTHORS PROFILE

Dr. Varun Kumar, Professor & Head, Department of CSE, School of Engg. & Tech., ITM University, Gurgaon, completed his PhD in Computer Science. He received his M. Phil. in Computer Science and M. Tech. in Information Technology. He has 13 years of teaching experience. He is recipient of Gold Medal at his Master's degree. His area of interest includes Data Warehousing, Data Mining, and Object Oriented Languages like C++, JAVA, C# etc. He is an approved Supervisor for Ph.D., M. Phil., and M. Tech. Programme of various universities and currently he is guiding Ph.D. & M. Phil. scholars and M. Tech. students for their major project work in his area of research interest. He has published more than 35 research papers in Journals/Conferences/Seminars at international/national levels. He is working as an Editorial Board Member / Reviewer of various International Journals and Conferences.

Ms. Nisha Rathee is perusing her MTech in Computer Science and Engineering from ITM University Gurgaon, Haryana, India. Her area of interest includes Data Warehousing and Data Mining.

A Fuzzy Decision Support System for Management of Breast Cancer

Ahmed Abou Elfetouh Saleh
Dept. of Information Systems,
Faculty of computer and information system
Mansoura University
Mansoura, Egypt
elfetouh@gmail.com

Sherif Ebrahim Barakat
Dept. of Information Systems,
Faculty of computer and information system
Mansoura University
Mansoura, Egypt
sherifiib@yahoo.com

Ahmed Awad Ebrahim Awad
Dept. of Information Systems,
Faculty of computer and information system
Mansoura University
Mansoura, Egypt
ahmedaweb@yahoo.com

***Abstract*— In the molecular era the management of cancer is no more a plan based on simple guidelines. Clinical findings, tumor characteristics, and molecular markers are integrated to identify different risk categories, based on which treatment is planned for each individual case.**

This paper aims at developing a fuzzy decision support system (DSS) to guide the doctors for the risk stratification of breast cancer, which is expected to have a great impact on treatment decision and to minimize individual variations in selecting the optimal treatment for a particular case.

The developed system was based on clinical practice of Oncology Center Mansoura University (OCMU)

This system has six input variables (Her2, hormone receptors, age, tumor grade, tumor size, and lymph node) and one output variable (risk status). The output variable is a value from 1 to 4; representing low risk status, intermediate risk status and high risk status. This system uses Mamdani inference method and simulation applied in MATLAB R2009b fuzzy logic toolbox.

Keywords: Decision Support System; Breast Cancer; Fuzzy Logic; Mamdani Inference;

I. INTRODUCTION

In recent years, the methods of Artificial Intelligence have largely been used in the different areas including the medical applications. In the medical field, many decision support systems (DSSs) were designed, as Aaphelp, Internist I, Mycin, Emycin, Casnet/Glaucoma, Pip, Dxplain, Quick Medical Reference, Isabel, Refiner Series System and PMA [1,2,3,4,5,6,7] which assist physicians in their decisions for diagnosis and treatment of different diseases.

In cancer management many DSSs have been developed as ONCOCIN [1], OASIS, Lisa [8, 9].

The diagnosis of disease involves several levels of uncertainty and imprecision [10]. According to Aristotelian logic, for a given proposition or state we only have two logical values: true-false, black-white, 1-0. In real life, things are not either black or white, but most of the times are grey. Thus, in many practical situations, it is convenient to consider intermediate logical values. Uncertainty is now considered essential to science and fuzzy logic is a way to model and deal with it using natural language. We can say that fuzzy logic is a qualitative computational approach. Fuzzy logic is a method to render precise what is imprecise in the world of medicine.

Many medical applications use fuzzy logic as CADIAG [11], MILORD [11], DOCTORMOON [12], TxDENT [13], MedFrame/CADIAG-IV [14], FuzzyTempToxopert [14] and MDSS [15].

In the field of breast cancer, DSS is very important, as breast cancer is the most common cause of cancer death among women worldwide, in Egypt, breast cancer is the most common cancer among women; representing 18.9% of total cancer cases [16]. The National Cancer Institute (NCI) reported a series of 10556 patients with breast cancer during the year 2001.

The diagnoses have a lot of confounding alternatives, some of them are uncertain as Her2-neu positivity, hormone receptor status and age. Therefore the treatment planning is based on the interaction of a lot of compound variables with complex outcomes.

We planned to use fuzzy logic to deal with uncertainty for diagnosis risk status of breast cancer.

This paper is organized as follows; general structure of fuzzy logic system is introduced in section II, design of the system is presented in section III and test system and discussion are presented in section IV.

II. GENERAL STRUCTURE OF FUZZY LOGIC SYSTEM

Fuzzy logic system as seen in Fig. 1 consists of the following modules [17]:

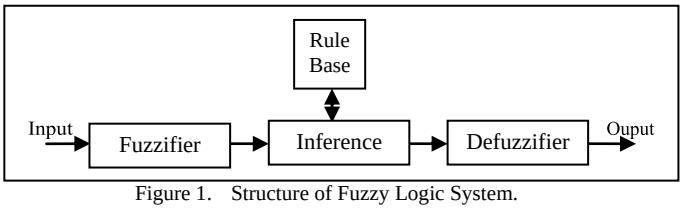


Figure 1. Structure of Fuzzy Logic System.

1. *Fuzzification:* - is the operation of transforming a crisp set to a fuzzy set. The operation translates crisp

input or measured values into linguistic concepts by using suitable membership functions.

2. *Inference Engine and Rule base:-* Once the inputs are fuzzified, the corresponding inputs fuzzy sets are passed to the inference engine that processes current inputs using the rules retrieved from the rule base.
3. *Defuzzifcation:-* At the output of the fuzzy inference there will always be a fuzzy set that is obtained by the composition of the fuzzy sets output by each of the rules. In order to be used in the real world, the fuzzy output needs to be interfaced to the crisp domain by the defuzzifier by using suitable membership functions.

III. DESIGN OF THE SYSTEM

In this section, we show the fuzzy decision support system designing, membership functions, fuzzy rule base, fuzzification and defuzzification.

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 DSS designing is determination of input and output variables. There are six input variables and one 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. In next section, paper shows the rules of system and Fuzzification , Defuzzification process.

A. Input Variables Are:

1) HER2: Stands for "Human Epidermal growth factor Receptor 2" and is a protein giving higher aggressiveness in breast cancers [18].This input variable has two fuzzy sets are "Negative" and "Positive". Membership functions of them are trapezoidal. Fuzzy sets Range of HER2 are identified in table I and membership functions for fuzzy sets are identified in Fig. 2

TABLE I. FUZZY SETS OF HER2 FACTOR

Input Field	Range	Fuzzy set
Her2	<=1.5	Negative
	1.5 - 3	May be Negative or Positive
	>=3	Positive

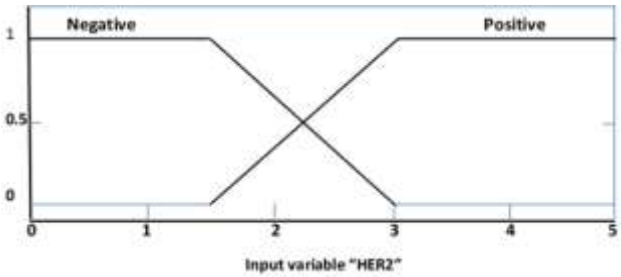


Figure 2. Membership Functions for HER2

$$\mu_{Negative}(x) = \begin{cases} 1 & x<=1.5 \\ (3-x)/1.5 & 1.5<x<3 \\ 0 & x >= 3 \end{cases}$$

$$\mu_{Positive}(x) = \begin{cases} 0 & x<=1.5 \\ (x-1.5)/1.5 & 1.5<x<3 \\ 1 & x >= 3 \end{cases}$$

2) Hormone Receptor: Identifies sensitivity of breast to hormone [19]. This input variable has four fuzzy sets are Negative, Weak Positive, Moderate Positive and Strong Positive. Membership functions of Negative and Strong Positive fuzzy sets are trapezoidal, membership functions of Weak Positive and Moderate Positive are triangle. Table II identifies fuzzy sets range and Fig. 3 identifies membership functions of fuzzy sets.

TABLE II. FUZZY SETS OF HORMONE RECEPTORS

Input Field	Range	Fuzzy set
Hormone	<=10	Negative
	10 - 15	May be Negative or Weak Positive
	15 - 20	May be Weak Positive or Moderate Positive
	20 - 35	Moderate Positive
	35 - 40	May be Moderate Positive or Strong Positive
	>=40	Strong Positive

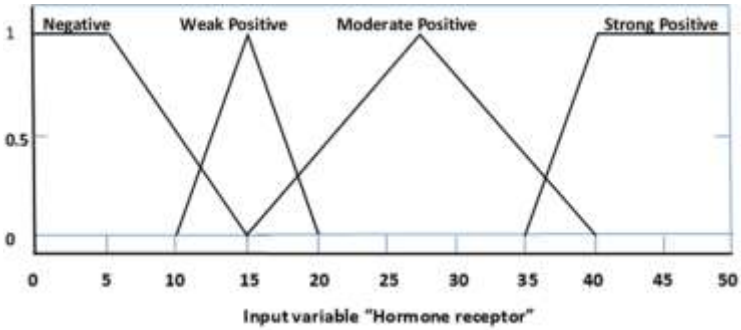


Figure 3. Membership Functions for Hormone Receptors

$$\mu\text{Negative}(x) = \begin{cases} 1 & x<=5 \\ (15-x)/10 & 5<x<15 \\ 0 & x >= 15 \end{cases}$$

$$\mu\text{Weak Positive}(x) = \begin{cases} 0 & x<=10 \\ (x-10)/5 & 10<x<15 \\ 1 & x = 15 \\ (20-x)/5 & 15<x<20 \\ 0 & x>=20 \end{cases}$$

$$\mu\text{Moderate Positive}(x) = \begin{cases} 0 & x<=15 \\ (x-15)/12.5 & 15<x<27.5 \\ 1 & x = 27.5 \\ (40-x)/12.5 & 27.5<x<40 \\ 0 & x>=40 \end{cases}$$

$$\mu\text{Strong Positive}(x) = \begin{cases} 0 & x<=35 \\ x-35)/\ 5 & 35<x<40 \\ 1 & x >=40 \end{cases}$$

3) Risk Age: This input variable has three fuzzy sets, are Very High, High And Low Risk age. Membership functions of these fuzzy sets are trapezoidal. Table III identifies fuzzy sets range and Fig. 4 identifies membership functions of them

TABLE III. FUZZY SETS OF RISK AGE

Input Field	Range	Fuzzy set
Age	<=20	Very High Risk
	20 - 30	May be Very High Risk or High Risk
	30 - 35	High Risk
	35 - 45	May be High Risk or Low Risk
	>=45	Low Risk

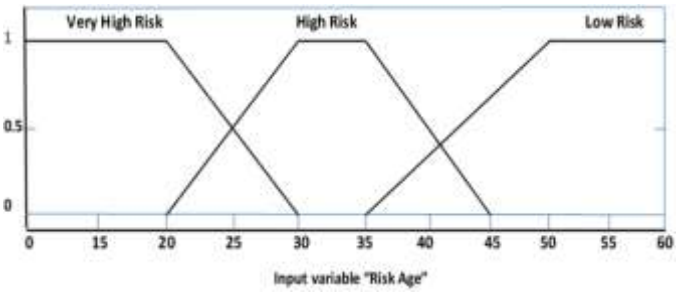


Figure 4. Membership Functions for Risk Age

$$\mu\text{VeryhighRisk}(x) = \begin{cases} 1 & x<=20 \\ (30-x)/10 & 20<x<30 \\ 0 & x >= 30 \end{cases}$$

$$\mu\text{HighRisk}(x) = \begin{cases} 0 & x<=20 \\ (x-20)/10 & 20<x<30 \\ 1 & 30=<x<=35 \\ (45-x)/10 & 35<x<45 \\ 0 & x>=45 \end{cases}$$

$$\mu\text{LowRisk}(x) = \begin{cases} 0 & x<=35 \\ (x-35)/15 & 35<x<50 \\ 1 & x >= 50 \end{cases}$$

4) Tumor Grade: A description of a tumor based on how abnormal the cancer cells look under a microscope and how quickly the tumor is likely to grow and spread [20]. Grading systems are different for each type of cancer. This input variable has three fuzzy sets Grade1, Grade2 and Grade3. Membership functions of these fuzzy sets are trapezoidal. Table IV identifies fuzzy sets range and Fig. 5 identifies membership functions of them

TABLE IV. FUZZY SETS OF TUMOR GRADE

Input Field	Range	Fuzzy set
Grade	<=4	Grade1
	4 - 5.5	May be Grade1 or Grade2
	5.5 - 6	Grade2
	6 - 7.5	May be Grade2 or Grade3
	>=7.5	Grade3

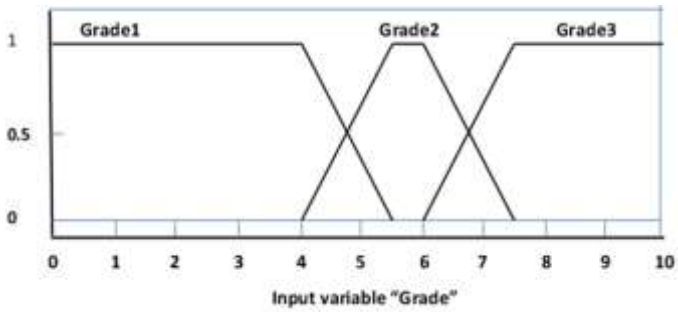


Figure 5. Membership Functions for Tumor Grade

$$\mu\text{Grade1}(x) = \begin{cases} 1 & x<=4 \\ (5.5-x)/1.5 & 4<x<5.5 \\ 0 & x >= 5.5 \end{cases}$$

$$\mu\text{Grade2}(x) = \begin{cases} 0 & x<=4 \\ (x-4)/1.5 & 4<x<5.5 \\ 1 & 5.5=<x<=6 \\ (7.5-x)/1.5 & 6<x<7.5 \\ 0 & x>=7.5 \end{cases}$$

$$\mu\text{Grade3}(x) = \begin{cases} 0 & x<=6 \\ (x-6)/1.5 & 6<x<7.5 \\ 1 & x >= 7.5 \end{cases}$$

5) Lymph Node: A lymph node is part of the body's lymphatic system, in the lymphatic system, a network of lymph vessels carries clear fluid called lymph, lymph vessels lead to lymph nodes; with it cancer cells are likely to spread from the primary tumor [21]. This input variable is Zero or has two fuzzy sets are Intermediate Number(Intermediate No.) and High Number(High No.), membership functions for these fuzzy sets are trapezoidal.
Table V identifies fuzzy sets range of lymph node variable and Fig. 6 identifies membership functions of them.

TABLE V. FUZZY SETS OF LYMPH NODE

Input Field	Range	Fuzzy set
Lymph Node	1 - 2	Intermediate No.
	2 - 10	May be Intermediate No. or High No.
	>=10	High No.

$$\mu\text{IntermediateNo.}(x) = \begin{cases} 1 & 1<=x<=2 \\ (10-x)/8 & 2<x<10 \\ 0 & x>=10 \end{cases}$$

$$\mu\text{HighNo.}(x) = \begin{cases} 0 & x<=2 \\ (x-2)/10 & 2<x<12 \\ 1 & x>=12 \end{cases}$$

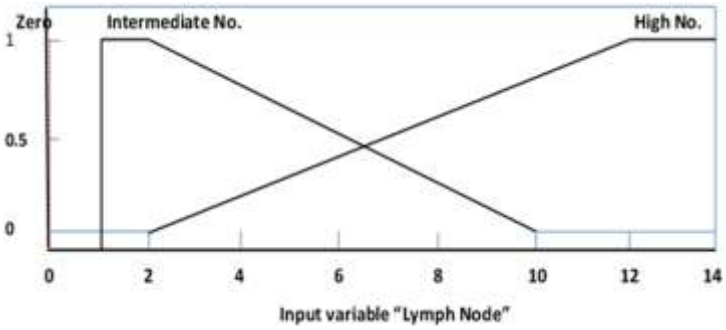


Figure 6. Membership Functions for Lymph Node

6) Tumor Size: This input variable has two fuzzy sets Small Size and Intermediate Size, membership functions for these fuzzy sets are trapezoidal, Table VI identifies fuzzy sets range of tumor size variable and Fig. 7 identifies membership functions of them.

TABLE VI. FUZZY SETS OF TUMOR SIZE

Input Field	Range	Fuzzy set
Tumor Size	<=2	Small Size
	2 - 4	May be Small Size or Intermediate Size
	>=4	Intermediate Size

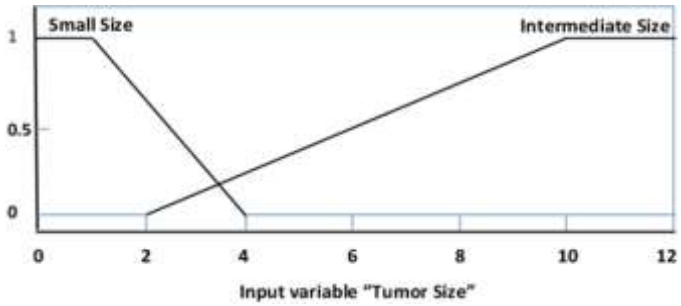


Figure 7. Membership Functions for Tumor Size

$$\mu_{SmallSize}(x) = \begin{cases} 1 & x<=1 \\ (4-x)/3 & 1<x<4 \\ 0 & x>=4 \end{cases}$$

$$\mu_{IntermediateSize}(x) = \begin{cases} 0 & x<=2 \\ (x-2)/8 & 2<x<10 \\ 1 & x>=10 \end{cases}$$

B. Output Variable Is:

The "goal" of the system is to identify risk status of breast cancer recurrence or mortality in early diagnosed patients. The output variable is a value from 1 to 4; representing Low Risk status, Intermediate Risk status and High Risk status. By increasing the value, tumor risk increases. This output has three fuzzy sets Low Risk, Intermediate Risk And High Risk; table VII identifies these fuzzy sets and its range.

The membership functions of these fuzzy sets are triangle as shown in Fig. 8.

TABLE VII. FUZZY SETS OF OUTPUT VARIABLE RISK STATUS

Output	Range	Fuzzy set
Risk Status	0 - 2	Low Risk
	1 - 3	Intermediate Risk
	2 - 4	High Risk

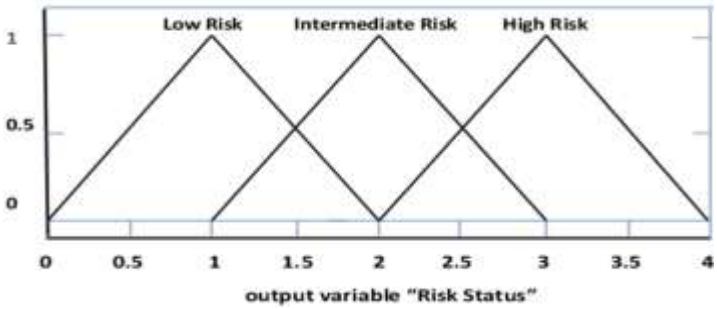


Figure 8. Membership Function For Output Variable Risk Status

C. Fuzzy Rule Base

The Rules Base is determined by the help of consultant doctors of the OCMU Center.

The rule base consists of 14 rules that determine the Risk status (High Risk, Intermediate Risk and Low Risk) by evaluation of the input variables mentioned above. Hormone receptor positivity level has no value in risk status characterization; however it plays an important role in further treatment decision. The rule base is shown in Table VIII

D. Fuzzification and Defuzzification

This system depends on Mamdani model for inference mechanism, in it and method is minimum (this system doesn't contains or operator), Implication method is minimum which involves defining the consequence as an output fuzzy set.

This can only be achieved after each rule has been evaluated and is allowed contribute its 'weight' in determining the output fuzzy set, Aggregation method between rules is maximum to combine output fuzzy set, so Fuzzification method here is max-min and Defuzzification method is centroid .

IV. TEST SYSTEM AND DISCUSSION

System has been tested by consultant oncologists and here is one of tested values as shown in Table IX and Fig. 9

TABLE VIII. RULE BASE OF THE SYSTEM

Rule NO	Her2	Hormone Receptors	Risk Age	Grade	Tumor Size	Lymph Node	Risk Status
1	Negative	Weak Positive	Low	Grade1	Small	Zero	Low Risk
2	Negative	Weak Positive	High	Grade1	Small	Zero	Low Risk
3	Negative	Moderate Positive	Low	Grade1	Small	Zero	Low Risk
4	Negative	Moderate Positive	High	Grade1	Small	Zero	Low Risk
5	Negative	Strong Positive	Low	Grade1	Small	Zero	Low Risk
6	Negative	Strong Positive	High	Grade1	Small	Zero	Low Risk
7	Negative	Any	Any	Grade2	Any	Zero	Intermediate Risk
8	Negative	Any	Any	Grade3	Any	Zero	Intermediate Risk
9	Negative	Any	Any	Any	Intermediate	Zero	Intermediate Risk
10	Negative	Any	Very High	Any	Any	Zero	Intermediate Risk
11	Negative	Any	Any	Any	Any	Intermediate No.	Intermediate Risk
12	Positive	Any	Any	Any	Any	Zero	Intermediate Risk
13	Positive	Any	Any	Any	Any	Intermediate No.	High Risk
14	Any	Any	Any	Any	Any	High No.	High Risk

TABLE IX. TESTED VALUES

Her2	Hormone Receptors	Age	Grade	Tumor Size	Lymph Node	Risk Status
4	35	35	4	3	5	3

Figure 9. Result Of Tested Values

And Fig. 10, 11, 12 and 13 shown surface viewer of some fields as follow:-

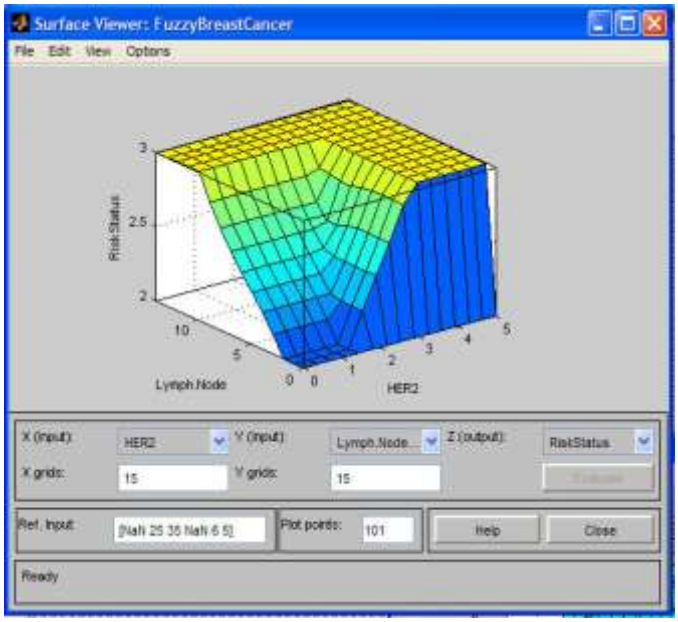


Figure 10. Surface Viewer of HER2 and Lymph Node

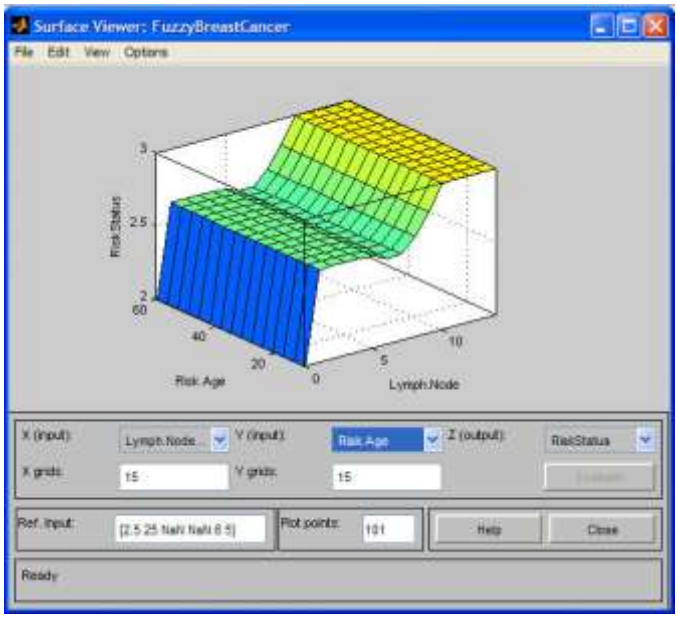


Figure 11. Surface Viewer of Hormone Receptor and Lymph Nod

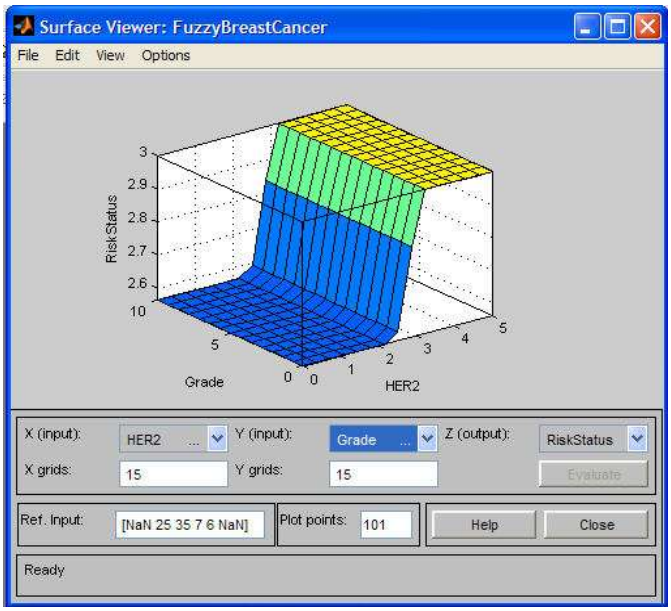


Figure 12. Surface Viewer of HER 2 and Tumor Grade

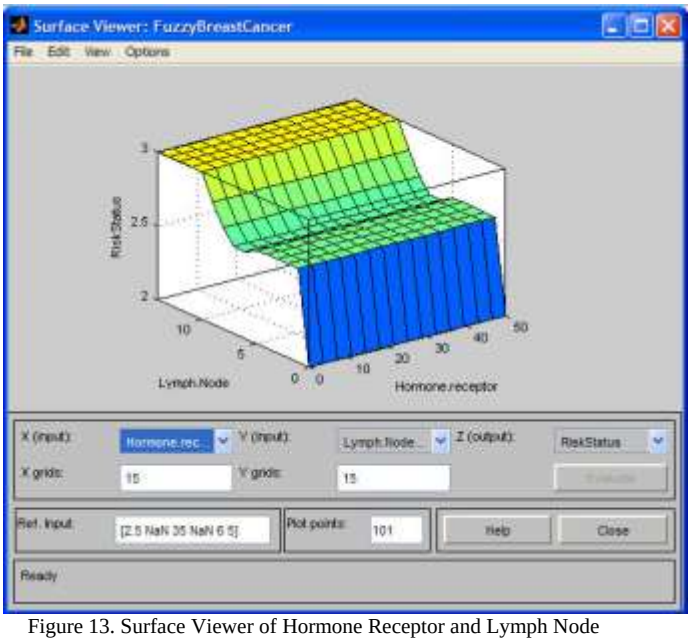


Figure 13. Surface Viewer of Hormone Receptor and Lymph Node

V. Conclusion and Future Work

This paper describes design of fuzzy decision support system for identification of breast cancer risk status in situations of data diversity and imprecision, which can be used by specialized doctors for cancer treatment.

The system design with based on membership functions, input variables, output variables and rule base. This system has been tested and approved by consultant oncologists in OCMU (Oncology Center Mansoura University).

In future this system can be applied for other types of cancers. In addition it can integrate the more complex evolving molecular data in cancer diagnosis.

Acknowledgments

We acknowledge Dr. Mohamed Awad Ebrahim Lecturer in Faculty of Medicine, Oncology Center Mansoura University for his efforts in this research.

REFERENCES

[1] "Clinical decision support system" available on http://www.openclinical.org/dss.html.

[2] "Using decision support to help explain clinical manifestations of disease", available on http://lcs.mgh.harvard.edu/projects/dxplain.html.

[3] A. Aiken, D. Sleeman "AKT-R4: "Supporting the Exploration of a Knowledge Web" , Scotland, 2005.

[4] C. Goodman, "A Pilot Reference to organizations, assessments and information resources Care Technology" , National Academies Press, 1988.

[5] M. L. Graber and A. Mathew, "Performance of a Web-Based Clinical Diagnosis Support System for Internists" , Society of General Internal Medicine, USA, 2007.

[6] "Free Patient online search "available on :http://www.freepatentsonline.com/.

[7] J. H. Knab, M.S. Wallace, R. L. Wagner, J. Tsoukatos, and M. B. Weinger, "The Use of a Computer-Based Decision Support System Facilitates Primary Care Physicians Management of Chronic Pain" , ANESTH ANALG journal, vol. 93, pp 712–720, September 2001 .

[8] P. Hammond and M. Sergot "Computer Support For Protocol-Based Treatment Of Cancer", Journal of Logic Programming, vol. 26, pp 93-111, 1996.

[9] J. P. Bury, C. Hurt, C. Bateman, S. Atwal, K. Riddy, J. Fox and V. Saha, "LISA: A Clinical Information and Decision Support System for Childhood Acute Lymphoblastic Leukaemia" Proceedings of the AMIA Annual Symposium, UK London, pp. 988, 2002.

[10] A. Torres and J. J. Nieto, " Fuzzy Logic in Medicine and Bioinformatics", Hindawi Publishing Corporation , Journal of Biomedicine and Biotechnology, Article ID 91908, pp 1–7, 2006.

[11] Ahmad, Y., & Husain, S. (2010). Applying Intuitionistic Fuzzy Approach to Reduce Search Domain in an Accidental Case. *International Journal of Advanced Computer Science and Applications - IJACSA*, *1*(4).

[12] Khan, A. R. (2011). Knowledge-Based Systems Modeling for Software Process Model Selection. *International Journal of Advanced Computer Science and Applications - IJACSA*, *2*(2), 20-25.

[13] N. H. Phuong, "Fuzzy set theory and medical expert systems: Survey and model" , Theory and Practice of Informatics book Springer Berlin / Heidelberg publisher, vol. 1012, pp 431-436,1995.

[14] N. H. Phuong and V. Kreinovich, "Fuzzy logic and its applications in medicine ", International Journal of Medical Informatics, vol. 62, no. 2-3, pp. 165-173, 2001.

[15] N. Kawahata and M. I. MacEntee, "A Measure of Agreement Between Clinicians and a Computer-Based Decision Support System for Planning Dental Treatment", Journal of Dental Education, vol. 66, Issue. 9, pp. 1031-1037, 2002.

[16] K. P. Adlassnig, "Fuzzy Systems in Medicine", vol. 21, Issue 1, pp 139-146, 2001.

[17] A. R. Filho, "MDSS, Medical Diagnosis Support System",2008

[18] S. Omar, H. Khaled, R. Gaafar, A.R. Zekry, S. Eissa and O. El-Khatib, "Breast cancer in Egypt: a review of disease presentation and detection strategies", Eastern Mediterranean Health Journal, vol. 9, no. 3, pp. 448-463, 2003.

[19] J.C. Bezdek, "Editorial: Fuzzy Models? What Are They, and Why?" IEEE Transactions on Fuzzy Systems, vol. 1, pp 1-5, February 1993.

[20] "what is her2" available on http://www.herceptin.com/her2-breast-cancer/testing-education/what-is.jsp

[21] "Are Hormone Receptors Present?" Available on http://www.breastcancer.org/symptoms/diagnosis/horm_receptors.jsp

[22] "Dictionary of Cancer Terms" available on http://www.cancer.gov/dictionary/?CdrID=45702

[23] "National Cancer Institute Sentinel Lymph Node Biopsy" available on http://www.cancer.gov/cancertopics/factsheet/therapy/sentinel-node-biopsy

Effective Implementation of Agile Practices

Ingenious and Organized Theoretical Framework

Veerapaneni Esther Jyothi
Lecturer, Department of Computer Applications,
V.R. Siddhartha Engineering College,
Kanuru, Vijayawada – 520 007, Andhra Pradesh, India.
nejyothi@hotmail.com

K. Nageswara Rao
Professor and Head, Department of Computer Science and Engineering,
P.V.P.Siddhartha Institute of Technology,
Kanuru, Vijayawada – 520 007, Andhra Pradesh, India.
drknrao@ieee.org

***Abstract*—Delivering software in traditional ways is challenged by agile software development to provide a very different approach to software development. Agile methods aim at fast, light and efficient than any other vigorous method to develop and support customers business without being chaotic. Agile software development methods claim to be people-oriented rather than process-oriented and adaptive rather than predictive. Solid Determination and Dedicated efforts are required in agile development to overcome the disadvantages of predefined set of steps and changing requirements to see the desirable outcome and to avoid the predictable results. These methods reach the target promptly by linking developers and stakeholders.**

The focus of this research paper is two fold. The first part is to study different agile methodologies, find out the levelheaded difficulties in agile software development and suggests possible solutions with a collaborative and innovative framework. The second part of the research paper concentrates on the importance of handling traceability in agile software development and finally proposes an ingenious and organized theoretical framework with a systematic approach to agile software development.

Keywords-Traceability; requirements; agile manifesto; framework

I. INTRODUCTION

Over the years many different software methodologies have been introduced and used by the software engineering community. Developers and users of these methods have invested significant amount of time and energy in improving and refining them. The method they choose for software development depends on the type of organization, the type of project and the type of people involved. Agile software development methodologies have been greeted with enthusiasm by many software developers because work is done at different levels in parallel [9]. We can use our creativity on the design level you can also have the fun of implementing the design in a smart way.

Handling unstable and volatile requirements throughout the development life cycles and delivering products in short time frames and under budget constraints when compared with traditional development methods are the two most significant characteristics of the agile approaches. Successful agile traceability processes incorporates the carefully devised planning stage to determine when, how, where and why each traceability link will be created.

This paper is organized as follows; section II presents the comparative study of different agile methodologies. Section III Surveys the major worthwhile risks associated with agile software development and suggests possible solutions. Section IV explains the importance of traceability in agile. Finally section V proposes ingenious and organized theoretical framework.

II. COMPARITIVE STUDY OF AGILE METHODS

Agile methods are a family of development techniques designed to deliver products on time, on budget, with high quality and customer satisfaction [15]. This family includes several and very different methods. The most popular include:

- eXtreme Programming (XP)
- Scrum
- Dynamic Systems Development Method (DSDM)
- Adaptive Software Development (ASD)
- The Crystal family

XP is a deliverable and disciplined approach to agile software development [1]. XP is successful because it stresses customer satisfaction and allows the software developers to confidently respond to changing software requirements even late in the lifecycle. The business culture affecting the development unit is another focal issue in XP.

Scrum approach has been developed for managing the systems development process [8]. It is an empirical approach applying the ideas of industrial process control theory to software development resulting in an approach that reintroduces the ideas of flexibility, adoptability & productivity.

Scrum concentrates on how the team members should function in order to produce the system flexibly in a constantly changing environment [8].

The fundamental idea behind DSDM is that instead of fixing the amount of functionality in a product, and then adjusting time and resources to reach that functionality, it is preferred to fix time and resources, and then adjust the amount of functionality accordingly. DSDM is a non-profit and non – proprietary framework for rapid application development (RAD) maintained by the DSDM consortium.

Adaptive software development is a lightweight software development method that accepts continuous change as the norm. The method follows a dynamic lifecycle, Speculate-Collaborate-Learn, instead of the traditional, static lifecycle, Plan-Design-Build. ASD emphasizes continuous learning. It is characterized by constant change, reevaluation, peering into an uncertain future, and intense collaboration among developers, testers, and customers [16]. ASD was designed for projects that are characterized with high-speed, high-change, and uncertainty.

The crystal family of methodologies includes a number of different methodologies for selecting the most suitable methodology for each individual project. Besides the methodologies, the crystal approach also includes principles for tailoring the methodologies to fit the raring circumstances of different projects. A team using crystal clear should be located in a shared office-space due to the limitations in its communication structure. Both the crystal clear as well as crystal orange suggest the following policy standards:

- incremental delivery on a regular basis
- progress tracking
- direct user involvement
- automated regression testing of functionality
- two user reviewing per release

TABLE I. TEAM SIZE IN CRYSTAL FAMILY

Methodology	Team (n° people)
Crystal Clear	2-6
Crystal Yellow	6-20
Crystal Orange	20-40
Crystal Red	40-80

TABLE II. COMPARISON OF AGILE METHODS

Concept	XP	SCRUM	DSDM	CRYSTAL	FDD
Team size	3-16	5-9	2-6	4-8	6-15
Number of teams	1	1-4	1-6	1-10	1-3
Volatility	high	high	low	high	low
Team distribution	no	no	yes	yes	yes

III. LEVELHEADED DIFFICULTIES IN AGILE SOFTWARE DEVELOPMENT AND THE SUGGESTED SOLUTIONS

Collaboration between customers and development team is important factor of agile software development. Even though the agile methods are meant for fast delivery under budget constraints there are certain levelheaded difficulties in agile software development.

Below are such difficulties which are worthwhile to be considered:

- Need high-quality collaboration between customers and agile development team.
- Need a high-level of customer involvement.
- Lack of long-term detailed plans.
- Producing a lower level documentation.
- Misinterpreted as unplanned and undisciplined.

Requirements are the base of all software products and their elicitation, management, and understanding are very common problems for all development methodologies [2]. In particular, the requirements variability is a major challenge for all commercial software projects. Five of the eight main factors for project failure deal with requirements incomplete requirements, low customer involvement, unrealistic expectations, changes in the requirements, and useless requirements.

TABLE III. MAIN CAUSES OF PROJECT FAILURE

Problem	%
Incomplete requirements	13.1
Low customer involvement	12.4
Lack of resources	10.6
Unrealistic expectations	9.9
Lack of management support	9.3
Changes in the requirements	8.7
Lack of planning	8.1
Useless requirements	7.5

A. Quality facilitator

Figure1 portrays the responsibilities of the Quality Facilitator in different aspects of the agile software development process. QF should be responsible of the effective management of changing requirements which is an important factor to be concentrated on in order to maximize stakeholder ROI. QF along with the agile team should as well address the issues of maintenance and support so as to maintain high discipline and good engineering principles. QF is responsible in the following aspects of different issues which are discussed.

Agile Project Management is an iterative method of determining requirements for software and for delivering projects in a highly flexible and interactive manner [4]. It requires empowered individuals from the relevant business with supplier and customer input. Agile project management takes the ideas from agile software development and applies them to project management. Agile methodologies generally promote a project management process that encourages stakeholder involvement, feedback, objective metrics and effective controls [11].

Software configuration management is a process for developing and maintaining concurrency of the product's performance, functional, and physical attributes with its requirements, design and operational information throughout its life [9]. Hence both the agile teams and the quality facilitator should focus on the configuration management.

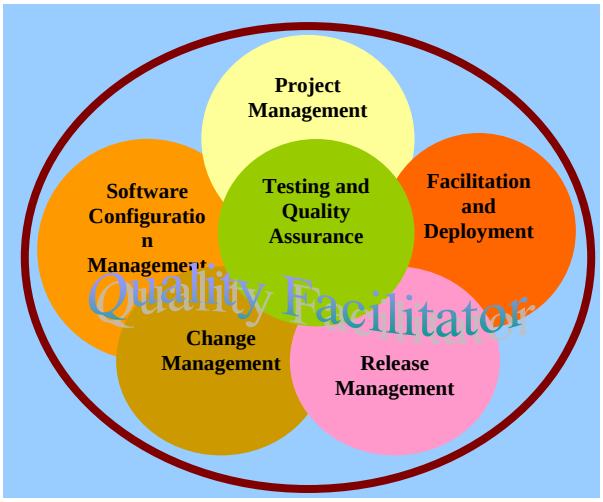


Figure 1. Responsibilities of Quality Facilitator

Agilists want to develop software which is both high-quality and high-value, and the easiest way to develop high-value software is to implement the highest priority requirements first. The activities contributing to effective change management are motivating change, creating vision of change, developing political support, managing the transition of change and sustaining momentum [11].

Teams practicing Agile Software Development value working software over other artifacts. A feature from the release plan is not complete until you can demonstrate it to your customer, ideally in a shippable state. Agile teams strive to have a working system ("potentially shippable") ready at the end of each iteration [11]. Thus Release Management should be easy for an ideal agile team, as agile teams, in theory are ready to release at regular intervals, and the release management aspect is the customer saying "ship it!."

Deployment is the point where an application starts to provide a return on the development investment. Delivering software doesn't stop once the application is written. It needs to be built, assembled and deployed, and historically this has been a non-trivial, manual and an error-prone task.

Agile testing does not emphasize testing procedures and focuses on ongoing testing against newly developed code until quality software from an end customer's perspective results. Agile testing is built upon the philosophy that testers need to adapt to rapid deployment cycles and changes in testing patterns [13].

Agile development processes place both added importance as well as special demands on your software quality assurance (SQA) practices. As an integrated part of the agile team, testers participate in the full life-cycle from requirements through release [9]. While many of the principles and practices of software quality assurance apply, an agile approach requires some new ways of viewing testing activities in the development process. Successful outsourcing projects are the ones that strike a good balance between testing and quality assurance throughout the lifecycle of the project [13].

B. Collaborative and innovative framework

To generate a significant impact on the productivity of the project, companies should adopt a hybrid approach that is using scrum and adding XP engineering practices.

Unlike XP, projects are wrapped by scrum, they becomes scalable and can be run simultaneously by distributed teams[1]. XP can be gradually implemented within scrum framework. Scrum is a project management approach, whereas XP is a methodology for project development [6]. Scrum only focuses on the managing side that is on what needs to be done rather than how to do it. XP on the developer side uses test driven development, pair programming, refactoring, etc. which are very essential to build good quality software but are missing in scrum [5].

Since Scrum doesn't have any engineering practices and XP doesn't have any management practices, XP with scrum projects allows better value metrics process for measuring and managing initiative ROI [10].

Planning includes the definition of the system being developed and the definition of the project team, tools and other resources, risk assessment, controlling issues, training needs and verification management approval.

The requirements can originate from the customer, sales and marketing division, customer support or software developers. The requirements are prioritized and the effort needed for their implementation is estimated. The product backlog list is constantly updated with new and more detailed items, as well as with more accurate estimations and new priority orders. Sprint backlog lists product backlog items selected to be implemented in the next iteration.

Unlike the product backlog, sprint backlog is stable until the sprint is completed. A new iteration of the system is delivered after all the items in the sprint backlog are completed.

Figure 2 is the framework, a hybrid approach in which XP engineering practices are implemented in the scrum sprint [5]. Sprints are iterative cycles where the functionality is developed or enhanced to produce new increments. Each Sprint includes the traditional phases of software development: requirements, analysis, design, evolution and delivery phases. These phases are implemented using extreme programming methodology. The functional tests created by the customer are run at the end of each iteration.

Here the key characteristics of XP are included such as refactoring - restructuring the system by removing duplication, improving communication, simplifying and adding flexibility without changing its functionality, pair programming – two people write the code at one computer which is great for complex and critical logic [11], collective code ownership – code belongs to the project not to any individual engineer, continuous integration – a new piece of code is integrated into the code-base as soon as it is ready. Thus the system is integrated and built many times a day [7]. All tests are run and passed for the changes in the code to be accepted.

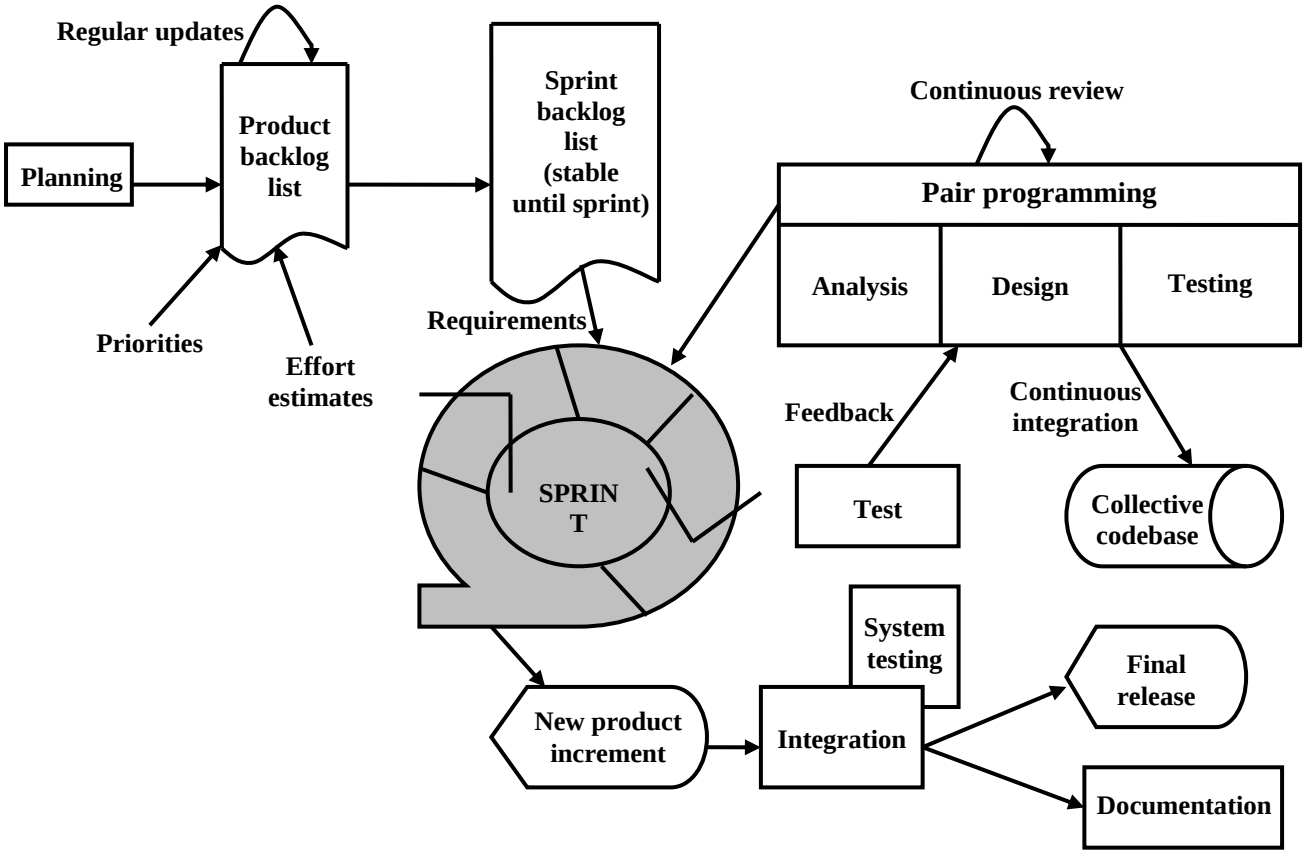


Figure 2. A Collaborative and innovative framework

Extra testing and checking of the performance of the system before the system can be released to the customer. At this stage, new changes may still be found and the decision has to be made if they are included in the current release. The postponed ideas and suggestions should be documented for later implementation [9].

Communication and coordination between project members should be enabled at all times. Any resistance against XP practices from project members, management or customers may be enough to fail the process. Ultimately better results can be obtained by tailoring some of the scrum principles such as the daily scrum meeting to keep track of the progress of the scrum team continuously and they also serve as planning meetings [8] [13].

Developing software on time and within budget is not good enough if the product developed is full of defects and customers today are demanding higher quality software than ever before. Now-a-days the software market is mature enough and users want to be assured of quality [5]. Due to time-to-market pressures or cost considerations, the developer may limit the software quality assurance function and not choose to conduct independent reviews.

IV. TRACEABILITY IN AGILE

This section of the research paper explains the importance of traceability in agile methods, how much level of traceability should be added and how to add traceability to agile methods.

A. The Importance of Traceability

Traceability is an important part in agile software development. Today there are many conflicts on whether the traceability is important in agile methods. During the study we have considered several opinions of different people of agile teams. Majority of people articulated that if traceability is supposed to be added in agile methods it should see that it will not place much administrative overhead to the team and also it should be considered on how the team will be benefited from the gathered information.

Traceability done in agile methods should help keep all the information gathered, organized and easy to find. It should ensure that the teams involved in development should find a way to look at how different artifacts are linked. Also it is important for teams to be capable of tracing the information and the decisions that were made during the entire process [14].

To generate a significant impact on the productivity of the project, companies should adopt a hybrid approach that is using scrum and adding XP engineering practices as discussed in section III. The tracing practices that apply to adding traceability to Scrum can be applied to XP and vice-versa.

B. Levels of AddingTtraceability

Some level of tracing is always considered to be useful for all the projects to be developed. There are several situations where it is good to add traceability in the agile methods. Some

projects are suitable to add traceability completely where as some are suitable to add traceability to the minimum.

Adding traceability to a project depends on the life span of the project. The organization should consult the customer or stake holder on how long the product will be used. The level of traceability to be added depends on the life span of the product. If the life span is too short it is better to avoid traceability in agile methods.

If the life span is about one year there is no need of adding all the traceability and the documentation. If the life span for the project is about to change then the traceability should be added from the beginning of the development itself. If the life span last long it is better to add full tracing. The customer or stake holder is uncertain about the life span of the project it is better to add traceability to a certain level.

Adding traceability to a project depends on the size of the project [12]. If the size of the project is small it is difficult to know how much the level of traceability is added. When Scrum or XP agile methods are used to develop small projects the level of traceability can be quite minimum. If the size of the project is large and complex, the team should document their work from the beginning of the development. Hence the level of traceability to be added should be maximum for such projects.

C. *The Tracing Practices in Adding Traceability*

In order to add traceability in agile methods there are several tracing practices which can be used individually or combined with each other. Below are several tracing practices:

1) *Trace the requirements to stakeholder.*
2) *Trace the requirements to problem description.*
3) *Trace the requirements to product backlog.*
4) *Trace the requirements to sprint backlog.*
5) *Trace the requirements to code.*
6) *Trace the requirements to database tables.*
7) *Trace the requirements to test.*
8) *Documentation.*

V. INGENIOUS AND ORGANIZED THEORETICAL FRAMEWORK

Agile Alliance formulated their ideas into values and further to twelve principles that support those values. Values of Agile Manifesto are as follows:

- Individuals and interactions over processes and tools.
- Working software over comprehensive documentation.
- Customer collaboration over contract negotiation.
- Responding to change over following a plan.

The above values are realized in the principles of Agile Manifesto[3].

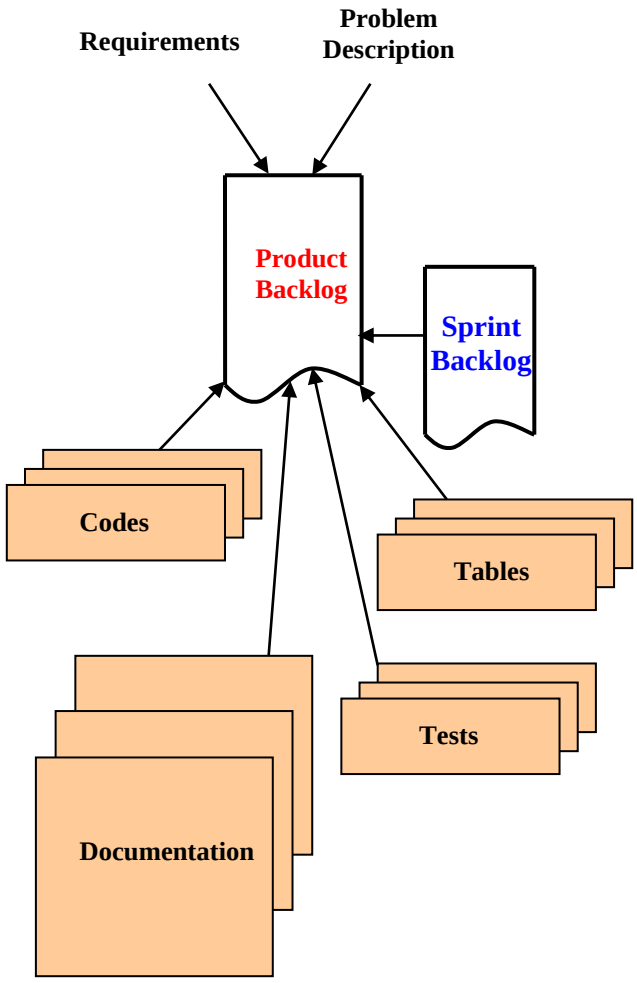


Figure 3. Traceability practices

The eleven principles among the twelve are chosen to model ingenious and organized theoretical frameworks which are as follows:

1) *Our highest priority is to satisfy the customer through early and continuous delivery of valuable software.*
2) *Welcome changing requirements, even late in development. Agile processes harness change for the customer's competitive advantage.*
3) *Deliver working software frequently, from a couple of weeks to a couple of months, with a preference to the shorter timescale.*
4) *Business people and developers must work together daily throughout the project.*

5) *Build projects around motivated individuals. Give them the environment and support they need, and trust them to get the job done.*
6) *The most efficient and effective method of conveying information to and within a development team is face-to-face conversation.*
7) *Working software is the primary measure of progress.*
8) *Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.*
9) *Continuous attention to technical excellence and good design enhances agility.*
10) *Simplicity--the art of maximizing the amount of work not done--is essential.*
11) *At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behavior accordingly.*

To improve software development and to maximize the ROI (Return On Investment), it is not merely sufficient if the organization just understand the principles of Agile manifesto but indeed implementing them in right way and in right situations is more important. In order to facilitate agile software development, all the aspects of the organization are to be considered. The four different aspects that lead the organization are internal, external, technical and social aspects. Each aspect should be handled very carefully.

The ingenious and theoretical framework given below clearly shows which of the agile manifesto principles are to be followed in each aspect of the system. By doing so working software can be produced using agile software development with faster delivery and within budget. Not all the principles are relevant enough to be followed in all the aspects of the organization. With careful examination and with experience the framework given suggests what principles are to be followed in each and every aspect of the organization and their interdependencies through which working software can be produced which satisfies both customers and the development team.

Internal aspects in the organization relate to the development team. External aspects relate to the customers or stakeholders. Technical aspects relate to different stages of development. Social aspects relate to the people, their working style, communication, job satisfaction.

The first principle which tells team to show highest priority as mentioned above relates to all the four aspects of the organization. The second principle relates to external - social aspects and also external – Technical aspects since it deals with the basic goal of agile software development. The third principle relates to technical aspects with external implementation and also to technical aspects with internal implementation. Principle four relates to social aspects with internal and external implementation since it deals with coordination and collaboration between people and developers.

The fifth principle relates to internal –social aspects of the organization, principle six relates to social aspects in coordination with to internal and external aspects since communication is the most important issue. Principle seven relates to technical aspects with external implementation and also to technical aspects with internal implementation which is the ultimate goal of agile software development. The eighth principle relates to social aspects with external implementation and also to social aspects with internal implementation

The ninth principle which is the key strength and success factor of the organization relates to technical and internal aspects, principle 10 also relates to technical and internal aspects, the eleventh principle relates to internal and social aspects.

The involvement of the subject matter expert is very important for smooth running of the project. Since the subject matter experts have the domain knowledge he can help the development team in following the principles and practices.

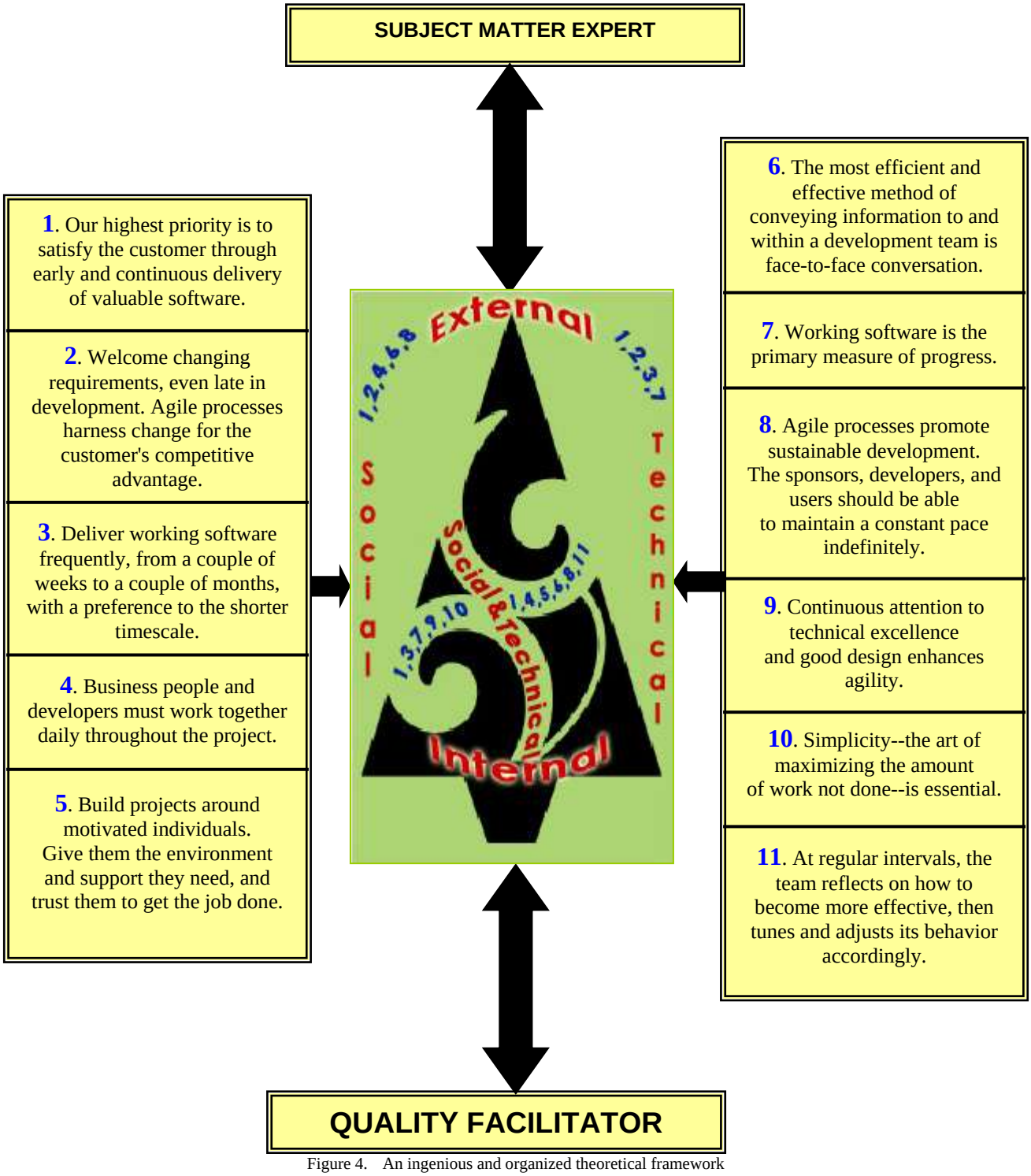


Figure 4. An ingenious and organized theoretical framework

CONCLUSION

Most of the agile techniques are repacking of well-known techniques like ignoring documentation of meetings, maintaining reviews, writing test cases before writing code, etc. This research paper suggests that small and mid-sized companies can have a Quality Facilitator who should play an important role to eliminate all the discussed worthwhile risks to the extent possible to produce a defect free product. Since certifications like ISO, CMMI emphasize more on documentation of the activities/work items, we need to concentrate on the documentation work for agile also [10]. So

our research suggests a QF role for ensuring the process is followed and the documentation is done properly so that there won't be any quandary for the project while facing the internal/external audits.

The proposed collaborative and innovative framework can be implemented along with the suggested possible solutions for the mentioned levelheaded difficulties. Also the organizations have to take steps in implementing the suggested theoretical framework which can lead in developing successful projects.

REFERENCES

[1] Weisert, C., (2002), The #1 Serious Flaw in Extreme Programming (XP), Information Disciplines, Inc., Chicago. [Online]. Available from: http://www.idinews.com/Xtreme1.html.

[2] B. Boehm and R. Turner, "Using Risk to Balance Agile and Plan-Driven Methods," IEEE Computer, vol. 36, no. 6, pp. 57-66, June 2003.

[3] Agile Alliance.2002. Agile Manifesto. http://www.agilealliance.org/

[4] Coram, M., and Bohner, S., "The Impact of Agile Methods on Software Project Management", In Proceedings of the 12th IEEE International Conference and Workshops on the Engineering of Computer-Based Systems (ECBS), April 2005, pp. 363-370.

[5] V. Esther Jyothi and K. Nageswara Rao, "Effective implementation of agile practices – A collaborative and innovative framework", published in CiiT International Journal of Software Engineering and Technology, Vol 2, No 9, September 2010.

[6] Maurer, F. and Martel, S. (2002a). Extreme programming: Rapid development for Web-based applications. IEEE Internet Computing 6(1): 86-90.

[7] Highsmith, J. and Cockburn, A. (2001). Agile Software Development: The Business of Innovation. Computer 34(9): 120-122.

[8] Schwaber, K. and Beedle, M. (2002). Agile Software Development with Scrum. Upper Saddle River, NJ, Prentice-Hall.

[9] Robert C. Martin, Agile Software Development, Principles, Patterns, and Practices, Prentice Hall, 2002

[10] Baker, Steven W. Formalizing Agility: An Agile Organization's Journey toward CMMI Accreditation. Agile 2005 Proceedings, IEEE Press, 2005.

[11] Alleman, Glen. Agile Program Management: Moving from Principles to Practice. Agile Product & Project Management, Vol. 6 No. 9, Cutter Consortium, September, 2005.

[12] Leffingwell, Dean, Scaling Software Agility: Best Practices for Large Enterprises, Addison-Wesley Professional, 2007.

[13] Anderson, David J., and Eli Schragenheim, Agile Management for Software Engineering: Applying the Theory of Constraints for Business Results, Prentice Hall, 2003.

[14] Poppendieck, Mary and Tom Poppendieck, Lean Software Development: An Agile Toolkit. Addison Wesley Professional, 2003.

[15] Ming Huo, June Verner, Liming Zhu, Mohammad Ali Babar, "Software Quality and Agile Methods", COMPSAC '04, IEEE 2004.

[16] Craig Larman, Victor R. Basili. Iterative and Incremental Development: A Brief History, 0018-9162/03/ © 2003 IEEE

[17] Salem, A. M. (2010). A Model for Enhancing Requirements Traceability and Analysis. International Journal of Advanced Computer Science and Applications - IJACSA, 1(5), 14-21.

AUTHORS PROFILE

Mrs. Veerapaneni Esther Jyothi is a Microsoft certified professional, currently pursuing Ph.D Computer Science from Raylaseema University, Kurnool. She is working as a Lecturer in the department of Computer Applications, Velagapudi Siddhartha Engineering college since 2008 and also has Industrial experience. She has published papers in reputed international conferences recently. Her areas of interest include Software Engineering, Object Oriented Analysis and Design and DOT NET.

Dr. K. Nageswara Rao is currently working as Professor and Head in the Department of Computer Science Engineering, Prasad V. Potluri Siddhartha Institute of Technology, Kanuru, Vijayawada-7. He has an excellent academic and research experience. He has contributed various research papers in the journals, conferences of International/national. His area of interest includes Artificial Intelligence, Software Engineering, Robotics, & Datamining.

Wavelet Based Image Denoising Technique

Sachin D Ruikar
Department of Electronics and Telecommunication Engineering
Shri Guru Gobind Singhji Institute of Engineering and Technology,
Nanded, India
ruikarsachin@gmail.com

Dharmpal D Doye
Department of Electronics and Telecommunication Engineering
Shri Guru Gobind Singhji Institute of Engineering and Technology,
Nanded, India
dddoye@yahoo.com

***Abstract*— This paper proposes different approaches of wavelet based image denoising methods. The search for efficient image denoising methods is still a valid challenge at the crossing of functional analysis and statistics. In spite of the sophistication of the recently proposed methods, most algorithms have not yet attained a desirable level of applicability. Wavelet algorithms are useful tool for signal processing such as image compression and denoising. Multi wavelets can be considered as an extension of scalar wavelets. The main aim is to modify the wavelet coefficients in the new basis, the noise can be removed from the data. In this paper, we extend the existing technique and providing a comprehensive evaluation of the proposed method. Results based on different noise, such as Gaussian, Poisson's, Salt and Pepper, and Speckle performed in this paper. A signal to noise ratio as a measure of the quality of denoising was preferred.**

Keywords: Image; Denoising; Wavelet Transform; Signal to Noise ratio; Kernel;

I. INTRODUCTION

The image usually has noise which is not easily eliminated in image processing. According to actual image characteristic, noise statistical property and frequency spectrum distribution rule, people have developed many methods of eliminating noises, which approximately are divided into space and transformation fields The space field is data operation carried on the original image, and processes the image grey value, like neighborhood average method, wiener filter, center value filter and so on. The transformation field is management in the transformation field of images, and the coefficients after transformation are processed. Then the aim of eliminating noise is achieved by inverse transformation, like wavelet transform [1], [2]. Successful exploitation of wavelet transform might lessen the noise effect or even overcome it completely [3]. There are two main types of wavelet transform - continuous and discrete [2]. Because of computers discrete nature, computer programs use the discrete wavelet transform. The discrete transform is very efficient from the computational point of view. In this paper, we will mostly deal with the modeling of the wavelet transform coefficients of natural images and its application to the image denoising problem. The denoising of a natural image corrupted by Gaussian noise is a classic problem in signal processing [4]. The wavelet transform has become an important tool for this problem due to its energy compaction property [5]. Indeed, wavelets provide a framework for signal decomposition in the form of a sequence of signals known as approximation signals with decreasing resolution supplemented by a sequence of additional touches called details [6][7]. Denoising or estimation of functions, involves reconstituting the signal as well as possible on the basis of the observations of a useful signal corrupted by noise [8] [9] [10] [11]. The methods based on wavelet representations yield very simple algorithms that are often more powerful and easy to work with than traditional methods of function estimation [12]. It consists of decomposing the observed signal into wavelets and using thresholds to select the coefficients, from which a signal is synthesized [5]. Image denoising algorithm consists of few steps; consider an input signal $\mathbf{x}(\mathbf{t})$and noisy signal $\mathbf{n}(\mathbf{t})$. Add these components to get noisy data $\mathbf{y}(\mathbf{t})$ i.e.

$$\mathbf{y}(\mathbf{t}) = \mathbf{x}(\mathbf{t}) + \mathbf{n}(\mathbf{t}). \quad (1)$$

Here the noise can be Gaussian, Poisson's, speckle and Salt and pepper, then apply wavelet transform to get $\mathbf{w}(\mathbf{t})$.

$$\mathbf{y}(\mathbf{t}) \xrightarrow{\textbf{Wavelet Transform}} \mathbf{w}(\mathbf{t}). \quad (2)$$

Modify the wavelet coefficient $\mathbf{w}(\mathbf{t})$ using different threshold algorithm and take inverse wavelet transform to get denoising image $\hat{\mathbf{x}}(\mathbf{t})$.

$$\mathbf{w}(\mathbf{t}) \xrightarrow{\textbf{Inverse Wavelet Transform}} \hat{\mathbf{x}}(\mathbf{t}). \quad (3)$$

The system is expressed in Fig. 1.

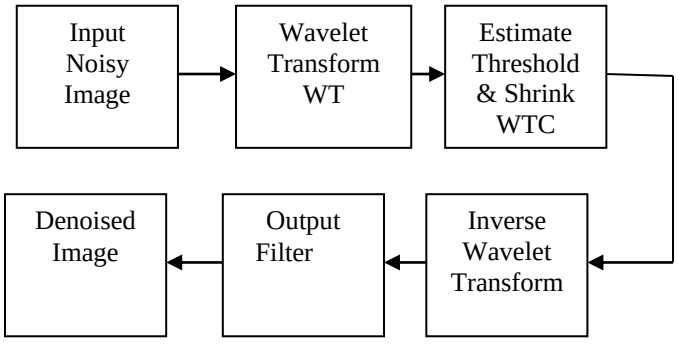


Figure 1: Block diagram of Image denoising using wavelet transform.

Image quality was expressed using signal to noise ratio of denoised image.

II. WAVELET TRANSFORM

The wavelet expansion set is not unique. A wavelet system is a set of building blocks to construct or represents a signal or function. It is a two dimensional expansion set, usually a basis, for some class one or higher dimensional signals. The wavelet expansion gives a time frequency localization of the signal. Wavelet systems are generated from single scaling function by

scaling and translation. A set of scaling function in terms of integer translates of the basic scaling function by

$$\boldsymbol{\varphi_k(t) = \varphi(t-k) \quad k \in Z \quad \varphi \in L^2.} \tag{4}$$

The subspaces of $L^2(R)$ spanned by these functions is defined as $\mathbf{v_0 = \overline{Span}_k\{\varphi_k(t)\}}$ for all integers k from minus infinity to infinity. A two dimensional function is generated from the basic scaling function by scaling and translation by

$$\boldsymbol{\varphi_{j,k}(t) = 2^{j/2}\varphi(2^j t - k)}. \tag{5}$$

Whose span over k is

$$\mathbf{v_j = \overline{Span}_k\{\varphi_k(2^j t)\} = v_j = \overline{Span}_k\{\varphi_{j,k}(t)\}, \; k \in Z}. \tag{6}$$

for all integer. The multiresolution analysis expressed in terms of the nesting of spanned spaces as $\boldsymbol{\ldots \ldots \subset v_{-2} \subset v_{-1} \subset v_0 \subset v_1 \subset v_2 \subset \cdots \subset L^2}$. The spaces that contain high resolution signals will contain those of lower resolution also. The spaces should satisfy natural scaling condition $\boldsymbol{f(t) \in v_j \Leftrightarrow f(2t) \in v_{j+1}}$ which ensures elements in space are simply scaled version of the next space. The nesting of the spans of $\boldsymbol{\varphi(2^j t - k)}$denoted by $\mathbf{v_j}$ i.e. $\boldsymbol{\varphi(t)}$is in $\mathbf{v_0}$, it is also in $\mathbf{v_1}$, the space spanned by $\boldsymbol{\varphi(2t)}$. This $\boldsymbol{\varphi(t)}$can be expressed in weighted sum of shifted $\boldsymbol{\varphi(2t)}$as

$$\boldsymbol{\varphi(t) = \sum_n h(n)\sqrt{2}\varphi(2t-n), \quad n \in Z} \tag{7}$$

Where the h (n) is scaling function. The factor $\boldsymbol{\sqrt{2}}$ used for normalization of the scaling function. The important feature of signal expressed in terms of wavelet function $\boldsymbol{\psi_{j,k}(t)}$not in scaling function $\boldsymbol{\varphi_{j,k}(t)}$. The orthogonal complement of $\mathbf{v_j}$ in $\mathbf{v_{j+1}}$ is defined asW_j, we require,

$$\boldsymbol{\langle \varphi_{j,k}(t)\psi_{j,l}(t)\rangle = \int \varphi_{j,k}(t)\psi_{j,l}(t)dt} \tag{8}$$

For all appropriate $\boldsymbol{j, k, l \in Z}$. The relationship of the various subspaces is $\boldsymbol{v_0 \subset v_1 \subset v_2 \subset \cdots \subset L^2}$. The wavelet spanned subspaces $\mathbf{W_0}$such that $\boldsymbol{v_1 = v_0 \otimes w_0}$, which extends to $\boldsymbol{v_2 = v_0 \otimes w_0 \otimes w_1}$.In general this $\boldsymbol{L^2 = v_0 \otimes w_0 \otimes w_1 \otimes \ldots}$ where $\mathbf{v_0}$ is in the space spanned by the scaling function $\boldsymbol{\varphi(t-k)}$, at $\boldsymbol{j = -\infty}$, equation becomes

$$\boldsymbol{L^2 = \ldots \otimes w_{-2} \otimes w_{-1} \otimes w_0 \otimes w_1 \otimes w_2 \ldots} \tag{9}$$

eliminating the scaling space altogether . The wavelet can be represented by a weighted sum of shifted scaling function $\boldsymbol{\varphi(2t)}$ as,

$$\boldsymbol{\psi(t) = \sum_n h_1(n)\sqrt{2}\varphi(2t-n), \quad n \in Z} \tag{10}$$

For some set of coefficient $\mathbf{h_1(n)}$, this function gives the prototype or mother wavelet $\boldsymbol{\psi(t)}$ for a class of expansion function of the form,

$$\boldsymbol{\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)} \tag{11}$$

Where $\boldsymbol{2^j}$the scaling of t is, $\boldsymbol{2^{-j}k}$ is the translation in t, and $\boldsymbol{2^{j/2}}$ maintains the $\mathbf{L^2}$ norms of the wavelet at different scales. The construction of wavelet using set of scaling function $\boldsymbol{\varphi_k(t)}$ and $\boldsymbol{\psi_{j,k}(t)}$ that could span all of $\mathbf{L^2(R)}$, therefore function $g(t) \in L^2(R)$ can be written as

$$\boldsymbol{g(t) = \sum_{k=-\infty}^{\infty} c(k)\varphi_k(t) + \sum_{j=0}^{\infty}\sum_{k=-\infty}^{\infty} d(j,k)\psi_{j,k}(t)} \tag{12}$$

First summation in above equation gives a function that is low resolution of g (t), for each increasing index j in the second summation, a higher resolution function is added which gives increasing details. The function d (j, k) indicates the differences between the translation index k, and the scale parameter j. In wavelet analysis expand coefficient at a lower scale level to higher scale level, from equation (10), we scale and translate the time variable to given as

$$\boldsymbol{\varphi(2^j t - k) = \sum_n h(n)\sqrt{2}\varphi(2^{j+1}t - 2k - n)}. \tag{13}$$

After changing variables m=2k+n, above equation becomes

$$\boldsymbol{\varphi(2^j t - k) = \sum_m h(m-2k)\sqrt{2}\varphi(2^{j+1}t - m)}. \tag{14}$$

If we denote v_j as $\boldsymbol{v_j = \overline{Span}_k\{2^j\varphi(2^j t - k)\}}$ then $\boldsymbol{f(t) \in v_{j+1} \Rightarrow f(t) = \sum_k c_{j+1}(k)2^{\frac{j+1}{2}}\varphi(2^{j+1}t - k)}$. is expressible at scale j+1, with a scaling function only not wavelets. At one scale lower resolution, wavelets are necessary for the detail not available at a scale of j. We have

$$\boldsymbol{f(t) = \sum_k c_j(k)2^{j/2}\varphi(2^j t - k) + \sum_k d_j(k)2^{\frac{j}{2}}\varphi(2^j t - k)}. \tag{15}$$

Where the 2^j terms maintain the unity norm of the basis functions at various scales. If $\varphi_{j,k}(t)$ and $\psi_{j,k}(t)$ are orthonormal, the j level scaling coefficients are found by taking the inner product

$$\boldsymbol{c_j(k) = \langle f(t), \varphi_{j,k}(t)\rangle = \int f(t)\, 2^{j/2}\varphi(2^j t - k)dt}. \tag{16}$$

By using equation (14) and interchanging the sum and integral, can be written as

$$\boldsymbol{c_j(k) = \sum_m h(m-2k)\int f(t)2^{\frac{j+1}{2}}\varphi(2^{j+1}t - m)dt}. \tag{17}$$

But the integral is inner product with the scaling function at a scale j+1giving

$$\boldsymbol{c_j(k) = \sum_m h(m-2k)c_{j+1}(m)}. \tag{18}$$

The corresponding wavelet coefficient is

$$\boldsymbol{d_j(k) = \sum_m h_1(m-2k)c_{j+1}(m)}. \tag{19}$$

Fig. 2 shows the structure of two stages down sampling filter banks in terms of coefficients.

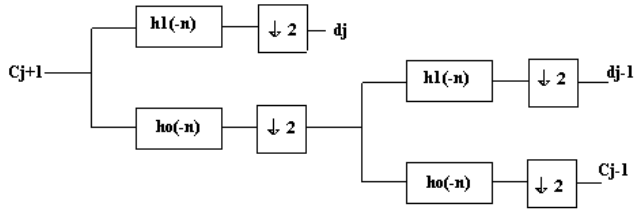


Figure 2: Two stages down sampling filter bank

A reconstruction of the original fine scale coefficient of the signal made from a combination of the scaling function and wavelet coefficient at a course resolution which is derived by

considering a signal in the $\mathbf{j+1}$ scaling function space $\mathbf{f(t)} \in \mathbf{v_{j+1}}$. This function written in terms of the scaling function as

$$\mathbf{f(t)} = \sum_{\mathbf{k}} \mathbf{c_{j+1}(k)} \mathbf{2^{\frac{j+1}{2}}} \boldsymbol{\varphi}(\mathbf{2^{j+1}t - k}). \quad (20)$$

In terms of next scales which requires wavelet as

$$\mathbf{f(t)} = \sum_{\mathbf{k}} \mathbf{c_j(k)} \mathbf{2^{j/2}} \boldsymbol{\varphi}(\mathbf{2^j t - k}) + \sum_{\mathbf{k}} \mathbf{d_j(k)} \mathbf{2^j} \boldsymbol{\psi}(\mathbf{2^j t - k}) \quad (21)$$

Substituting equation (15) and equation (11) into equation (20), gives

$$\mathbf{f(t)} = \sum_{\mathbf{k}} \mathbf{c_j(k)} \sum_{\mathbf{n}} \mathbf{h(n)} \mathbf{2^{\frac{j+1}{2}}} \boldsymbol{\varphi}(\mathbf{2^{j+1}t - 2k - n}) + \sum_{\mathbf{k}} \mathbf{d_j(k)} \sum_{\mathbf{n}} \mathbf{h_1(n)} \mathbf{2^{\frac{j+1}{2}}} \boldsymbol{\varphi}(\mathbf{2^{j+1}t - 2k - n}). \quad (22)$$

Because all of these function are orthonormal, multiplying equation (20) and equation (21) by $\boldsymbol{\varphi}(\mathbf{2^{j+1}t - k})$ and integrating evaluates the coefficients as

$$\mathbf{c_{j+1}(k)} = \sum_{\mathbf{m}} \mathbf{c_j(m)h(k - 2m)} + \sum_{\mathbf{m}} \mathbf{d_j(m)h_1(k - 2m)}. \quad (23)$$

Fig. 3 shows the structure of two stages up sampling filter banks in terms of coefficients i.e. synthesis from coarse scale to fine scale [5] [6] [7].

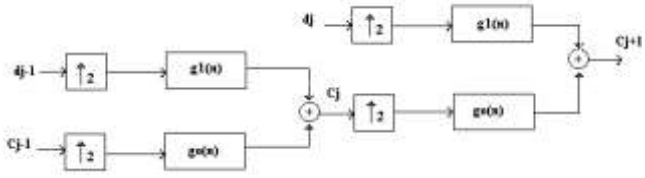


Figure 3: Two stages up sampling filter

In filter structure analysis can be done by apply one step of the one dimensional transform to all rows, then repeat the same for all columns then proceed with the coefficients that result from a convolution with in both directions[6][7][8][9][10][12]. The two level wavelet decomposition as shown in fig 4.

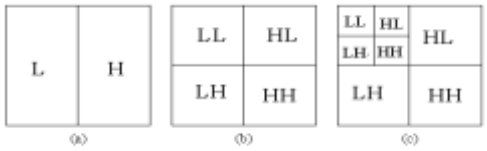


Figure 4: Two-dimensional wavelet transform.

III. Denoising Technique With Existing Threshold

Noise is present in an image either in an additive or multiplicative form. An additive noise follows the rule, $\mathbf{w(x,y)} = \mathbf{s(x,y)} + \mathbf{n(x,y)}$

While the multiplicative noise satisfies

$$\mathbf{w(x,y)} = \mathbf{s(x,y)} \times \mathbf{n(x,y)}.$$

Where $\mathbf{s(x,y)}$ is the original signal, $\mathbf{n(x,y)}$denotes the noise. When noise introduced into the signal it produces the corrupted image $\mathbf{w(x,y)}$.14]. 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. Salt and Pepper Noise is an impulse type of noise, which is also referred to as intensity spikes. This is caused generally due to errors in data transmission. 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. The source of this noise is attributed to random interference between the coherent returns [7], [8], [9] [10]. Fully developed speckle noise has the characteristic of multiplicative noise.

A. Universal Threshold

The universal threshold can be defined as,

$$\mathbf{T} = \boldsymbol{\sigma}\sqrt{\mathbf{2log(N)}} \quad (24)$$

N being the signal length, $\boldsymbol{\sigma}$ being the noise variance is well known in wavelet literature as the Universal threshold. It is the optimal threshold in the asymptotic sense and minimizes the cost function of the difference between the function. One can surmise that the universal threshold may give a better estimate for the soft threshold if the number of samples is large [13] [14].

B. Visu Shrink

Visu Shrink was introduced by Donoho [13]. It uses a threshold value t that is proportional to the standard deviation of the noise. It follows the hard threshold rule. An estimate of the noise level σ was defined based on the median absolute deviation given by

$$\widehat{\boldsymbol{\sigma}} = \frac{\mathbf{Median}(\{|\mathbf{g_{j-1,k}}|:\mathbf{k=0,1,\ldots..2^{j-1}-1}\})}{\mathbf{0.6745}} \quad (25)$$

Where$\mathbf{g_{j-1,k}}$ corresponds to the detail coefficients in the wavelet transform. VisuShrink does not deal with minimizing the mean squared error. Another disadvantage is that it cannot remove speckle noise. It can only deal with an additive noise. VisuShrink follows the global threshold scheme, which is globally to all the wavelet coefficients [9].

C. Sure Shrink

A threshold chooser based on Stein's Unbiased Risk Estimator (SURE) was proposed by Donoho and Johnstone and is called as Sure Shrink. It is a combination of the universal threshold and the SURE threshold [15] [16]. This method specifies a threshold value $\mathbf{t_j}$ for each resolution level $\mathbf{j}$ in the wavelet transform which is referred to as level dependent threshold. The goal of Sure Shrink is to minimize the mean squared error [9], defined as,

$$\mathbf{MSE} = \frac{\mathbf{1}}{\mathbf{n^2}} \sum_{\mathbf{x,y-1}}^{\mathbf{n}} (\mathbf{Z(x,y) - S(x,y)})^{\mathbf{2}} \quad (26)$$

Where $\mathbf{z(x,y)}$is the estimate of the signal, $\mathbf{s(x,y)}$ is the original signal without noise and n is the size of the signal. Sure Shrink suppresses noise by threshold the empirical wavelet coefficients. The Sure Shrink threshold t* is defined as

$$\mathbf{t^*} = \mathbf{min(t, \sigma\sqrt{2logn})} \quad (27)$$

Where t denotes the value that minimizes Stein's Unbiased Risk Estimator, σ is the noise variance computed from

Equation, and n is the size of the image. It is smoothness adaptive, which means that if the unknown function contains abrupt changes or boundaries in the image, the reconstructed image also does [17] [18].

D. Bayes Shrink

Bayes Shrink was proposed by Chang, Yu and Vetterli. The goal of this method is to minimize the Bayesian risk, and hence its name, Bayes Shrink [19]. The Bayes threshold, t_B, is defined as

$$\mathbf{t_B} = \boldsymbol{\sigma^2}/\boldsymbol{\sigma_s} \quad (28)$$

Where σ^2 is the noise variance and σ_s is the signal variance without noise. The noise variance σ^2 is estimated from the sub band HH by the median estimator shown in equation (28). From the definition of additive noise we have $\mathbf{w(x,y) = s(x,y) + n(x,y)}$. Since the noise and the signal are independent of each other, it can be stated that

$$\boldsymbol{\sigma_w^2 = \sigma_s^2 + \sigma^2}. \quad (29)$$

σ_w^2 can be computed as shown below:

$$\boldsymbol{\sigma_w^2 = \frac{1}{n^2}\sum_{x,y-1}^{n} w^2(x,y)}. \quad (30)$$

The variance of the signal, σ^2_s is computed as

$$\boldsymbol{\sigma_s = \sqrt{max(\sigma_w^2 - \sigma^2, 0)}}. \quad (31)$$

With σ^2 and σ_s^2, the Bayes threshold is computed from Equation (31). Using this threshold, the wavelet coefficients are threshold at each band [20].

E. Normal Shrink:

The threshold value which is adaptive to different sub band characteristics

$$\mathbf{TN} = \boldsymbol{\beta}\, \boldsymbol{\sigma^2}/\boldsymbol{\sigma_y}.$$

Where the scale parameter β has computed once for each scale, using the following equation.

$$\boldsymbol{\beta} = \sqrt{\mathbf{log}\left(\frac{L_k}{J}\right)}. \quad (32)$$

L_k means the length of the sub band at $\mathbf{K^{th}}$ scale. $\boldsymbol{\sigma^2}$means the noise variance, [13] which can be estimated from the sub band HH using equation (32).

IV. Proposed Denoising Scheme

There are different denoising scheme used to remove noise while preserving original information and basic parameter of the image. Contrast, brightness, edges and background of the image should be preserved while denoising in this technique. Wavelet transform tool used in denoising of image. Multi resolution analysis structure consider for denoising scheme. Actually, the performance of our algorithm is very close, and in some cases even surpasses, to that of the already published denoising methods. Performance measured in terms of signal to noise ratio

A. New threshold function:

This function is calculated by

$$\mathbf{newth} = \sqrt{\mathbf{2m \times log(M)}} \quad (33)$$

Where, **M** is the total number of pixel of an image, **m** is the mean of the image. This function preserves the contrast, edges, background of the images. This threshold function calculated at different scale level.

B. Circular kernel:

Kernel applied to the wavelet approximation coefficient, to get denoised image with all parameters undisturbed. The kernel uses here in this technique contains some components like [0 0 1 1 1 0 0; 0 1 1 1 1 1 0; 1 1 1 1 1 1 1; 1 1 1 1 1 1 1; 1 1 1 1 1 1 1; 0 1 1 1 1 1 0; 0 0 1 1 1 0 0]. Multi resolution analysis wavelet structure has used for this kernel to get result.

C. Mean-Max threshold:

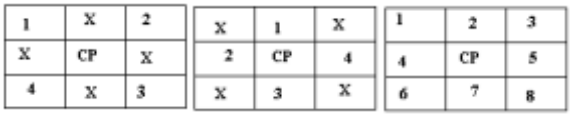
Generation of the threshold in using mean and max method after decomposition. Let xi denotes the sequence of elements; threshold can be calculated using following technique.

$$\mathbf{MAXMIN} = \{\mathbf{x_j}\} = \mathbf{MAX}\{[\mathbf{MIN}(\mathbf{x_1}, \dots \dots \mathbf{x_k}], [\mathbf{MIN}(\mathbf{x_2}, \dots \dots \mathbf{x_{k+1}})], \dots \dots + [\mathbf{MIN}(\mathbf{x_{i-k+1}}, \dots \dots \mathbf{x_i})]\} \quad (34)$$

$$\mathbf{MINMAX} = \{\mathbf{x_j}\} = \mathbf{MIN}\{[\mathbf{MAX}(\mathbf{x_1}, \dots \dots \mathbf{x_k}], [\mathbf{MAX}(\mathbf{x_2}, \dots \dots \mathbf{x_{k+1}})], \dots \dots + [\mathbf{MAX}(\mathbf{x_{i-k+1}}, \dots \dots \mathbf{x_i})]\} \quad (35)$$

D. Nearest neighbor:

This technique gives better result for different kernel structure shown in figure (5). In this kernel central pixel (CP), calculated from the neighbor value. Three different kernels have proposed for better reduction of noise using wavelet transform at different scale. Mark 'x' denotes low value at that position.



(a) **(b)** **(c)**

Figure 5: Kernel at different noise level.

V. Result

Image parameters has not disturb when denoising. In this paper calculating threshold function in spatial domain, and Lena image is used for implementation. When denoising we have to preserve contrast of the image. Image brightness in denoising kept same but preserves the background and the gray level tonalities in the image. The noise term is considered as a random phenomenon and it is uncorrelated, hence the average value of the noise results in a zero value, therefore consider proper kernel to get denoised image The low pass spatial filter reduces the noise such as bridging the gaps in the lines or curve in a given image, but not suitable for reducing the noise patterns consisting of strong spike like components [21] [22]. The high pass filters results in sharp details, it provides more visible details that obscured, hazy, and poor focus on the original image. Now wavelets preferred in denoising while preserving all the details of the image. Table 1 shows the results with existing technique and proposed denoising scheme.

TABLE 1: RESULT OF DIFFERENT TECHNIQUE WITH LENA.

Methods		Denoised image SNR			
		Gaussian	***Salt and Pepper***	***Poisson***	***Speckle***
VisuShrink		14.2110	11.8957	11.6988	12.2023
Universal		9.2979	9.6062	9.5227	9.8297
Sure shrink		16.8720	17.2951	14.3082	14.7922
Normal Shrink		15.4684	17.2498	14.5519	15.0069
Bays shrink		15.3220	16.9775	14.6518	14.9716
New threshold		15.0061	15.5274	13.9969	14.5171
Circular kernel		16.7560	18.9791	17.9646	14.2433
Mean Max approximation	Maxmin	9.5908	10.2514	10.0308	10.4287
	Meanmin	12.5421	12.9643	13.2045	13.3588
	Minmax	13.3761	9.3629	13.4904	13.8338
	Meanmax	10.7183	11.5483	11.3422	11.8212
	Sqrtth	16.9222	19.9917	16.9681	16.4639
Nearest Neighbor	Four diagonal	13.6178	13.4029	12.9661	13.6229
	Four directional	13.7226	13.1139	13.1241	13.8448
	Eight conn-ectivity	13.5136	13.0216	13.0835	13.9546

VI. CONCLUSION

This technique is computationally faster and gives better results. Some aspects that were analyzed in this paper may be useful for other denoising schemes, objective criteria for evaluating noise suppression performance of different significance measures. Our new threshold function is better as compare to other threshold function. Some function gives better edge perseverance, background information, contrast stretching, in spatial domain. In future we can use same threshold function for medical images as well as texture images to get denoised image with improved performance parameter.

REFERENCES

[1] Donoho.D.L,Johnstone.I.M, "Ideal spatial adaptation via wavelet shrinkage", Biometrika,81,pp.425-455,1994.

[2] Gao Zhing, Yu Xiaohai, "Theory and application of MATLAB Wavelet analysis tools", National defense industry publisher,Beijing,pp.108-116, 2004.

[3] Aglika Gyaourova Undecimated wavelet transforms for image de-noising, November 19, 2002.

[4] Bart Goossens, Aleksandra Pi˘zurica, and Wilfried Philips, "Image Denoising Using Mixtures of Projected Gaussian Scale Mixtures", IEEE Transactions On Image Processing, Vol. 18, No. 8, August 2009, Pp. 1689-1702

[5] Michel Misiti, Yves Misiti, Georges Oppenheim, Jean-Michel Poggi, "Wavelets and their Applications", Published by ISTE 2007 UK.

[6] C Sidney Burrus, Ramesh A Gopinath, and Haitao Guo, "Introduction to wavelet and wavelet transforms", Prentice Hall1997.S. Mallat, A Wavelet Tour of Signal Processing, Academic, New York, second edition, 1999.

[7] R. C. Gonzalez and R. Elwood's, Digital Image Processing. Reading, MA: Addison-Wesley, 1993.

[8] M. Sonka,V. Hlavac, R. Boyle Image Processing , Analysis , And Machine Vision. Pp10-210 & 646-670

[9] Raghuveer M. Rao., A.S. Bopardikar Wavelet Transforms: Introduction To Theory And Application Published By Addison-Wesley 2001 pp1-126

[10] Arthur Jr Weeks , Fundamental of Electronic Image Processing PHI 2005.

[11] Jaideva Goswami Andrew K. Chan, "Fundamentals Of Wavelets Theory, Algorithms, And Applications", John Wiley Sons

[12] Donoho, D.L. and Johnstone, I.M. (1994) Ideal spatial adaptation via wavelet shrinkage. Biometrika, 81, 425-455.

[13] D. L. Donoho and I. M. Johnstone, "Denoising by soft thresholding", IEEE Trans. on Inform. Theory, Vol. 41, pp. 613-627, 1995.

[14] Mark J. T. Smith and Steven L. Eddins, "Analysis/synthesis techniques for sub band image coding,"IEEE Trans. Acoust., Speech and Signal Process., vol. 38,no. 8, pp. 1446–1456, Aug. 1990.

[15] F. Luisier, T. Blu, and M. Unser, "A new SURE approach to image denoising: Inter-scale orthonormal wavelet thresholding," IEEE Trans. Image Process., vol. 16, no. 3, pp. 593–606, Mar. 2007.

[16] X.-P. Zhang and M. D. Desai, "Adaptive denoising based on SURE risk," IEEE Signal Process. Lett., vol. 5, no. 10, pp. 265–267, Oct. 1998.

[17] Thierry Blu, and Florian Luisier "The SURE-LET Approach to Image Denoising". IEEE Transactions On Image Processing, VOL. 16, NO. 11, pp 2778 – 2786 , NOV 2007 .

[18] H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch: 'Adaptive Bayesian wavelet shrinkage', J. Amer. Stat. Assoc., Vol. 92, No 440, Dec. 1997, pp. 1413-1421

[19] Chang, S. G., Yu, B., and Vetterli, M. (2000). Adaptive wavelet thresholding for image denoising and compression. IEEE Trans. on Image Proc., 9, 1532–1546.

[20] Andrea Polesel, Giovanni Ramponi, And V. John Mathews, "Image Enhancement Via Adaptive Unsharp Masking" IEEE Transactions On Image Processing, Vol. 9, No. 3, March 2000, Pp505-509

[21] G. Y. Chen, T. D. Bui And A. Krzyzak, Image Denoising Using Neighbouringwavelet Coefficients, Icassp ,Pp917-920

[22] Sasikala, P. (2010). Robust R Peak and QRS detection in Electrocardiogram using Wavelet Transform. International Journal of Advanced Computer Science and Applications - IJACSA, 1(6), 48-53.

[23] Kekre, H. B. (2011). Sectorization of Full Kekre ' s Wavelet Transform for Feature extraction of Color Images. International Journal of Advanced Computer Science and Applications - IJACSA, 2(2), 69-74.

AUTHORS PROFILE

Ruikar Sachin D has received the postgraduate degree in Electronics and Telecommunication Engineering from Govt Engg College, Pune University, India in 2002. He is currently pursuing the Ph.D. degree in Electronics Engineering, SGGS IET , SRTMU Nanded, India. His research interests include image denoising with wavelet transforms, image fusion and image in painting.

Dharmpal D Doye received his BE (Electronics) degree in 1988, ME (Electronics) degree in 1993 and Ph. D. in 2003 from SGGS College of Engineering and Technology, Vishnupuri, Nanded (MS) – INDIA. Presently, he is working as Professor in department of Electronics and Telecommunication Engineering, SGGS Institute of Engineering and Technology, Vishnupuri, Nanded. His research fields are speech processing, fuzzy neural networks and image processing.

Multicasting over Overlay Networks – A Critical Review

M.F.M Firdhous
Faculty of Information Technology,
University of Moratuwa,
Moratuwa,
Sri Lanka.
Mohamed.Firdhous@uom.lk

***Abstract* – Multicasting technology uses the minimum network resources to serve multiple clients by duplicating the data packets at the closest possible point to the clients. This way at most only one data packets travels down a network link at any one time irrespective of how many clients receive this packet. Traditionally multicasting has been implemented over a specialized network built using multicast routers. This kind of network has the drawback of requiring the deployment of special routers that are more expensive than ordinary routers. Recently there is new interest in delivering multicast traffic over application layer overlay networks. Application layer overlay networks though built on top of the physical network, behave like an independent virtual network made up of only logical links between the nodes. Several authors have proposed systems, mechanisms and protocols for the implementation of multicast media streaming over overlay networks. In this paper, the author takes a critical look at these systems and mechanism with special reference to their strengths and weaknesses.**

Keywords – Multicasting; overlay networks; streaming media;

I. INTRODUCTION

Media multicasting is one of the most attractive applications that can exploit the network resources least while delivering the most to the clients. Multicast is a very efficient technology that can be used to deliver the same content to multiple clients simultaneously with minimum bandwidth and server loading. The use of this technology is many and web TV, IP TV, web radio, online delivery of teaching are few of them [1-4]. In all these applications, the server continues to deliver the content while clients can join and leave the network any time to receive the content, but they would receive it only from where they joined the stream.

Traditionally multicasting was tied to the underlying network with special multicast routers making the necessary backbone. Multicast routers are special type of routers with the capability of duplicating and delivering the same data packet to many outgoing links depending on which links clients reside downstream. Traditional IP based routers are unicast that receive data packets on one link and either forward that packet only to one outgoing link or drop it depending on the routing table entries. Broadcasting is totally disabled in the internet due to the unnecessary congestion caused by broadcast traffic that may bring the entire internet down in a short time.

Using the network layer multicast routers to create the backbone of the internet is not that attractive as these routers are more expensive compared to unicast routers. Also, the absence of a multicast router at any point in the internet would defeat the objective of multicasting throughout the internet downstream from that point. Chu et al., have proposed to replace the multicast routers with peer to peer clients for duplicating and forwarding the packets to downstream clients [5]. In this arrangement, the duplicating and forwarding operation that makes multicasting attractive compared to unicast and broadcast would be carried out at the application layer. The multimedia application installed in end nodes would carry out the duplicating and forwarding operation in addition to the display of the content to the downstream nodes that request the stream from an upstream node.

A peer to peer network is formed by nodes of equal capability and act both as client and server at the same time depending on the function performed. Peer to peer networks and applications have advantages over traditional client server, such as the elimination of single point of failure, balance the network load uniformly and to provide alternate path routing easily in case of link failures [6]. A peer to peer network forms an overlay network on top of the existing network infrastructure. The formation of this overlay network by the peer to peer nodes make the resulting network resilient to changes in the underlying network such as router and link failures and congestion.

An overlay network is a computer network which is built on top of another network. Nodes in the overlay can be thought of as being connected by virtual or logical links, each of which corresponds to a path, perhaps through many physical links, in the underlying network. Usually overlay network run at the application layer of the TCP/IP stack making use of the underlying layer and independent of them [7]. Figure 1 shows the basic architecture of an overlay network. The nodes in an overlay network will form a network architecture of their own that may be totally different and independent of the underlying network.

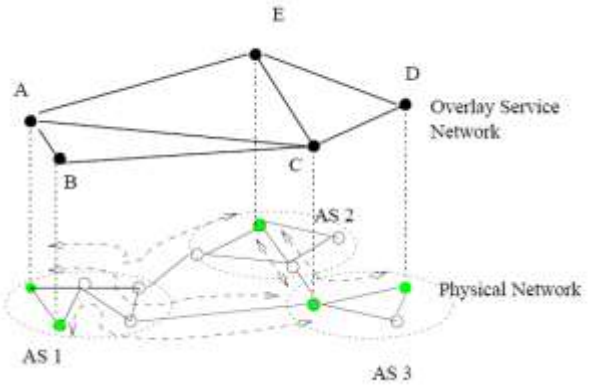


Figure 1: An Overlay Network

II. MULTICASTING OVER OVERLAY NETWORKS

Several authors have proposed protocols and services how to implement overlay based multicast over the internet. In this section, an in depth analysis would be carried out on the some of the most recent protocols and application in terms of their architecture, advantages and disadvantages.

Chen et al., have proposed ACOM Any-source Capacity-constrained Overlay Multicast system. This system is made up of three multicast algorithms namely Random Walk, Limited Flooding, and Probabilistic Flooding on top of a non-DHT (Distributed Hash table) overlay network with simple structures [8]. ACOM divides the receiving nodes into multicast groups based on the upload bandwidth of a node as the upload bandwidth is determining factor as a node may be required to transfer multiple copies of the same packet over the uplink. The number of nodes in a multicast group hence depends on the capacity of the uplink making the nodes with higher capacity to support large number of neighbors. An overlay network is established for each multicast group that transforms the multicast stream to a broadcast stream with the scope of the overlay.

The overlay network is made up of two components, namely an unrestricted ring that is fundamentally different from the location specific DHT based ring and the random graph among the nodes. The ring maintenance is carried out by requiring the nodes to the next node as a successor and a few other nodes in order to avoid the problem of ring breakage due to node leaving the network.

Packet delivery with the overlay network is carried out using a two phase process. In Phase 1, packets are forwarded to a random number of neighbors within a specific number (k – a system parameter) of hops. This follows a tree structure of delivery in Phase 1. For the purpose of delivery in Phase 2, the ring is partitioned into segments and each Phase 1 node is made responsible for a segment and required to deliver the packet to its successor which in turn forwards to its successor until the packet reaches a node that has already received the packet.

In practice what ACOM does is to forward the packet in Phase 1 using a tree structure to a number of random nodes and then in Phase 2 the packet is forwarded in a unicast fashion down the network until all the nodes within the multicast subgroup receives the packet.

Even though ACOM presents several advantages such as maintenance of virtual tree in place of multicast tree bound to the physical networks, it has certain disadvantages too. The main disadvantage is the total disregard of the physical distances when setting up of the virtual tree. ACOM also has certain other disadvantages in the formation and maintenance of the ring network.

Liu and Ma have proposed a framework called Hierarchy Overlay Multicast Network based on Transport Layer Multi-homing (HOMN-SCTP) [9]. This framework uses transport layer multi-homing techniques. HOMN-SCTP can be shared by a variety of applications, and provide scalable, efficient, and practical multicast support for a variety of group communication applications. The main component of the framework is the Service Broker (SvB) that is made up of the cores of of HOMN-SCTP and Bandwidth- Satisfied Overlay Multicast (BSOM). The BSOM searches for multicast paths to form overlay networks for upper layer QoS-sensitive applications, and balance overlay traffic load on SvBs and overlay links.

HOMN-SCTP is capable of supporting multiple applications on it and helps these applications meet the required QoS requirements. Since it is built on top of TCP, error control is totally delegated to the underlying layer. Nevertheless HOMN-SCTP is not suitable for applications such as streaming media that run on top of UDP.

Fair Load Sharing Scheme proposed by Makishi et al., concentrates on improving throughput instead of reducing the delay on multisource multimedia delivery such as video conferencing using Application Layer Multicasting (ALM) [10]. This scheme is based on tree structure and tries to find the tree with the highest possible throughput. The main objective of the protocol is to improve the receiving bit rate of the node that is having lowest bit rate out of all receivers. The proposed scheme is an autonomous distributed protocol that constructs the ALM where each node refines its own subtree. The end result of this refinement of subtrees is the automatic refinement of the entire tree structure.

This scheme assumes that there is backbone ring network that has unlimited bandwidth and all the receiving nodes can connect to this ring. The links from the backbone to the end nodes make the tree network which this protocol tries to improve for the receiver bandwidth. The end nodes initially make a basic (ad-hoc) tree network that would be improved iteratively over time until the all the nodes receive an optimum bandwidth in terms of bit rate. Once the basic tree structure has been established using unicast links, the nodes communicate with each other exchanging estimation queries with communication quality in terms of bandwidth to be allocated and the delay. When a better path than that is currently being used is found, the tree readjusts itself moving to the node with the better quality. This process is continued until the optimum link allocation is achieved.

Peng and Zhang have discussed the problem associated with the intranet based multimedia education platform [4]. They mainly analyze the principles of this education platform, but in the process they introduce Computer Assisted Instruction (CAI) multicasting network as the backbone of the

network. The CAI multicasting network is based on the IPv4 special class D multicasting (224.0.0.0 ~ 239.255.255.255) address block. The multicasting network has been implemented using the Winsock2 mechanism. Once the server node (teacher) has initiated the stream the user (students) can join the multicast network and receive the multicast stream on their computers. This system is suitable only for in class teaching or within limited distance where high networking resources are available.

Reduced delay and delay jitter are very important to viewing quality of any video from the viewers' point. In traditional IP multicasting RSVP and DiffServ algorithms were used to reserve resources prior to starting the streaming. Each router on the path from the server to the client uses a dynamic scheduling algorithm to deliver the packet based on the QoS requirements. Szymanski and Gilbert have proposed a Guaranteed Rate (GR) scheduling algorithm for computing the transmission schedules for each IQ packet-switched IP router in a multicast tree. The simulation results have shown that this algorithm manages on average two packets in a queue resulting in very low delay and jitter which can practically ignored as zero jitter [11].

Even though this algorithm virtually eliminates delay jitter, it is bound to the underlying layers as it needs to be implemented on routers.

Lua et al., have proposed a Network Aware Geometric Overlay Multicast Streaming Network. This network exploits the locality of the nodes in the underlay for the purpose of node placement, routing and multicasting. This protocol divides the nodes into two groups called SuperPeers and Peers. SuperPeers form the low latency, high bandwidth backbone and the Peer connect to the nearest SuperPeer to receive the content [12].

SuperPeers have been elected based on two criteria, namely: the SuperPeer should have sufficient resources to serve other SuperPeers and Peers and they must be reliable in terms of stability not join and leave the network very frequently. Network embedding algorithm computes node coordinates and geometric distances between nodes to estimate the performance metrics of the underlying network such as latency. Peers joining the overlay network calculate the total Round Trip Time (RTT) to the SuperPeers and join the SuperPeer that has the lowest RTT. A SuperPeer joining the network calculates the RTT to all the existing SuperPeers and joins the ones with the lowest RTT and then creates connection with other six SuperPeers around it.

When a SuperPeer leaves the network, the other SuperPeers would detect this by the loss of heartbeat signal from the node that has left and will reorganize themselves by sending discovery broadcast messages to all the SuperPeers. The Peers who are affected by the leaving of a SuperPeer need to reconnect to the overlay network by selecting the nearest SuperPeer.

The main strengths of this scheme can be summarized as it has good performance in terms of latency and efficient transmission of packets via the high speed backbone network whereas the weaknesses include the heavy dependence on the geographical locality, election of SuperPeers.

Pompili et al., have presented two algorithms called DIfferentiated service Multicast algorithm for Internet Resource Optimization (DIMRO) and DIfferentiated service Multicast algorithm for Internet Resource Optimization in Groupshared applications (DIMRO-GS) to build virtual multicast trees on an overlay network [13].

DIMRO constructs virtual source-rooted multicast trees for source-specific applications taking the virtual link available bandwidth into account. This avoids traffic congestion and fluctuation on the underlay network. Traffic congestion in the underlay would cause low performance. This keeps the average link utilization low by distributing data flows among the least loaded links. (DIMRO-GS) builds a virtual shared tree for group-shared applications by connecting each member node to all the other member nodes with a source-rooted tree computed using DIMRO.

Both these algorithms support service differentiation without the support of the underlying layers. Applications with less stringent QoS requirements reuse resources already exploited by members with more stringent requirements. Better utilization of network bandwidth and improved QoS are achieved due to this service differentiation.

System built using these algorithms would result in better performance due to differentiation of applications based on QoS requirements, but node dynamic may bring the quality of the system down.

Wang et al., have proposed an adapted routing scheme that minimizes delay and bandwidth consumption [14]. This routing algorithm creates an optimum balanced tree where the classical Dijkstra's algorithm is used to compute the shortest path between two nodes. In this scheme the Optimal Balance of Delay and Bandwidth consumption (OBDB) is formulated as:

$$OBDB = (1-\alpha)D + \alpha B$$

Where D is the minimal delay criteria and B is the minimal bandwidth consumption criteria.

From the above formula, it can be seen that one routing method may lead to another bad performance. The value of α is calculated based on the nature of the application. As the delay and bandwidth can be tuned to suit application requirements, applications performance can be controlled. But these algorithms will have performance issues in the face of node dynamics.

Kaafar et al., have proposed an overlay multicast tree construction scheme called LCC: Locate, Cluster and Conquer. The objective of this algorithm is to address scalability and efficiency issues [15]. The scheme is made up of two phases. One is a selective locating phase and the other one is the overlay construction phase. The selective locating phase algorithm locates the closest existing set of nodes (cluster) in the overlay for a newcomer. The algorithm does not need full knowledge of the network to carry out this operation, partial knowledge of location-information of participating nodes is sufficient for this operation. It then

allows avoiding initially randomly-connected structures with neither virtual coordinates system embedding nor fixed landmarks measurements. Then, on the basis of this locating process, the overlay construction phase consists in building and managing a topology-aware clustered hierarchical overlay.

This algorithm builds an efficient initial tree architecture with partial knowledge of the network but node dynamics may result in poor performance with time.

Walters et al., have studied the effect of the attack by adversaries after they become members of the overlay network [16]. Most of the overlay protocol can handle benign node failures and recover from those failures with relative ease, but they all fail when adversaries in the network start attacking the nodes. In this study, they have identified, demonstrated and mitigated insider attacks against measurement-based adaptation mechanisms in unstructured multicast overlay networks.

Attacks usually target the overlay network construction, maintenance, and availability and allow malicious nodes to control significant traffic in the network, facilitating selective forwarding, traffic analysis, and overlay partitioning. The techniques proposed in this work decrease the number of incorrect or unnecessary adaptations by using outlier detection. The proposed solution is based on the performance of spatial and temporal outlier analysis on measured and probed metrics to allow an honest node to make better use of available information before making an adaptation decision.

This algorithm creates a resilient overlay network in the both structured and unstructured overlay network in the presence of malicious attacks. But the strict nature of the algorithm may delay the adaptation of the network in the event of node dynamics disrupting the flow of information.

Alipour et al., have proposed an overlay protocol known as Multicast Tree Protocol (OMTP) [17]. This protocol can be used to build an overlay tree that reduces the latency between any two pair of nodes. The delay between the nodes has been reduced by adding a shortcut link by calculating the utility link between two groups.

The main advantage of this algorithm is the efficient data transfer but the efficiency may be affected by node dynamics in the overlay network.

Bista has proposed a protocol where the nodes informs the other nodes its leaving time when it joins the network [18]. Using this leaving time information new nodes are joined at the tree in such a manner early leaving nodes would make the leaf nodes down the line and the nodes that would stay longer would be at the higher levels. It has also been proposed a proactive recovery mechanism so that even if an upstream node leaves the tree, the downstream nodes can rejoin at predetermined nodes immediately, so that the recovery time of the disrupted nodes is the minimum.

These algorithms have several drawbacks including, the prior notice of the duration of stay in the network, arrangement of nodes based on the time of stay and central control to manage the node information.

Gao et al., have proposed a hybrid network combining the IP multicast network and the mesh overlay network [19]. The main objectives of the design were to build a network with high performance (low end-to-end transmission latency, high bandwidth mesh overlay links), low end-to-end hop count and high reliability.

This algorithm results in a good structure combining multicast and overlay, but the resulting network is still dependant on the physical network and managing the mesh network is expensive in terms of network resources.

Wang et al., have proposed hybrid overlay network combining a tree and mesh networks called mTreebone [20]. In the mTreebone, the tree forms the backbone and local nodes make a mesh to share content. The backbone tree network has been constructed by identifying the stable nodes as the churn of the backbone would be more expensive in terms of service disruption than the leaf node churn. On top of the tree based overlay network a mesh network has been created in order to handle the effect of node dynamics. Since the mesh network has been updated regularly using keep alive packets, any change in the mesh network immediately notified to all the other mesh nodes. In this design heuristics have been used to predict the stability of the nodes assuming the age of the in the network to be directly proportional to the probability of it staying longer in the network. That is longer a node stays in the network, larger the probability it would stay even further and more stable.

This algorithm results in a resilient overlay network but the network may carry duplicate packets in some part of the network. Also, the maintenance of the mesh network requires large network resources.

Guo and Jha have shown that the main problem in overlay based multicast networks is to optimize routing among CDN servers in the multicast overlay backbone in such a manner that it reduces the maximal end-to-end latency from the origin server to all end hosts is minimized [21]. They have identified this as the Host-Aware Routing Problem (HARP) in the multicast overlay backbone. The main reason for HARP is the last mile latency between end hosts and their corresponding proxy servers. The author of this paper have framed HARP as a constrained spanning tree problem and shown that it is NP-hard. As a solution they have presented a distributed algorithm for HARP. They also have provided a genetic algorithm (GA) to validate the quality of the distributed algorithm.

This structure results in a low latency routing path improving QoS but the non-consideration of node dynamics will affect the overall quality of the network.

Table I summarize the systems, algorithms and mechanism discussed above with special reference to their advantages and disadvantages. hows the comparison of these works with special reference to their advantages and disadvantages.

TABLE I: COMPARISON OF MULTICASTING OVERLAY NETWORKS

	Work	System/Protocol Proposed	Advantages	Disadvantages
1.	[4]	Computer Assisted Instruction (CAI) Multicasting Network	This system is suitable for in class teaching or teaching within a limited area using the high technology to larger classes.	This system uses the existing technologies to build a platform and hence no new technology has been introduced in terms of multicasting. This technique may work well on an intranet but cannot be ported to the internet as the internet lacks support for class D multicast addresses.
2.	[8]	Any-source Capacity-constrained Overlay Multicast System	Only maintenance of virtual tree is necessary, no strict multicast trees are maintained. Resilient to node dynamics as node maintain multiple neighbors in its neighbor table. Simple maintenance of unrestricted ring.	The random nature of the tree formation in Phase 1 totally ignores the physical distance from the source to that node. The researchers have considered this as an advantage, but this will make unnecessary delay in delivering a packet to a node that may be physically closer to the source but not selected as a Phase 1 node. Even though, the unrestricted ring is better than location specific DHT ring in terms of node creation and maintenance, the unrestricted node is inefficient in forwarding packets and the logical neighbors may not always be neighbors physically. Phase 2 forwarding is essentially unicast and cascaded in ACOM which may create a lot of delay in the ultimate delivery to the last node. The problem will be more severe in case of a high capacity Phase 1 node as it will have a large number of neighbors. Converting any portion of the multicast network to a broadcast network is inherently in efficient as broadcast is an inefficient protocol in terms of network utilization.
3.	[9]	Hierarchy Overlay Multicast Network	This framework builds an infrastructure on which multiple applications can be built. The framework helps application to meet the required QoS. The framework makes use of the facilities of the underlying layer for the error control as it is built on top of TCP.	The framework has been built on top of TCP and hence is not suitable for best effort services running on UDP especially streaming media applications. TCP is a high overhead protocol compared to UDP and hence, this protocol would inherently have high overhead. This framework will not meet the QoS requirement in terms of delay and delay jitter on high loss links as TCP would create delay and delay jitter on high loss networks and hence not suitable for real time applications like media streaming.
4.	[10]	Fair Load Sharing Scheme	This scheme works purely on the application layer without directly depending on the lower layers. The protocol results in the optimum receiving tree structure for a given situation.	This protocol is only suitable for network with stable receiving nodes. Dynamic nature of the nodes joining and leaving the network will make the tree fail as there is no mechanism in the protocol to handle node dynamics. The dynamic nature of the internet would make the tree structure oscillating as the communication quality heavily depends on the dynamics of the underlying network. Tree management is costly in terms of information exchanged between the nodes as the continuous information exchange is needed.
5.	[11]	Guaranteed Rate (GR) Scheduling Algorithm	This algorithm virtually eliminates delay jitter in received video stream resulting in high quality reception. The algorithm is simple enough to implement on any IP router.	This algorithm needs to be implemented on IP routers and hence bound to the underlay network. This does not provide a solution to the existing problem of how to use the existing network as it is to transport multimedia streaming without depending on the underlay network.
6.	[12]	Network Aware Geometric Overlay Multicast Streaming Network	This algorithm has good performance in terms of latency as the physical location of the node has been computed and used as a parameter in forming the overlay network. Breaking the network into two layers result in efficient transmission of	Since the algorithm heavily depends on the geographical location, an efficient geographical location computing algorithm is vital in addition to managing the overlay network. Since the election of SuperPeers depends on the condition that the SuperPeer should have high dependability in terms of churn frequency, past historical data may be necessary to

			packets as the SuperPeers are connected to each other via a high speed backbone network.	determine the reliability of a node accurately. SuperPeers should also have sufficient resources to support other SuperPeers and Peers in the network, if the resource availability information is obtained from the node itself, this would be an invitation to rogue nodes to highjack the entire backbone network using false information. The Peers that are affected by a SuperPeer leaving the network do not have automatic transfer to another SuperPeer, this drastically affect the reliability of the entire multicast operation.
7.	[13]	DIfferentiated service Multicast algorithm for Internet Resource Optimization (DIMRO) and DIfferentiated service Multicast algorithm for Internet Resource Optimization in Groupshared applications (DIMRO-GS)	These algorithms result in good performance for both QoS stringent applications and non QoS stringent applications due to service differentiation.	The node dynamics has not been considered when designing these protocols, hence node dynamics would drastically bring the quality of the network down.
8.	[14]	Adapted Routing Scheme	These algorithms results in better performance for any kind of application as delay and bandwidth can be tuned to meet the application requirements.	The node dynamics has not been considered in this routing scheme and hence node dynamics would drastically bring the quality of the network down.
9.	[15]	LCC : Locate, Cluster and Conquer multicast tree	This algorithm creates the initial tree architecture very efficiently with partial knowledge of the overlay network.	The node dynamics has not been considered in this routing scheme and hence node dynamics would drastically bring the quality of the network down.
10.	[16]		This algorithm creates a resilient overlay network in the face of attacks by malicious nodes. The algorithm works on unstructured overlays and hence can be easily adapted to structured overlay networks too.	The strict nature of the algorithm delays the adaptation of the network to genuine node dynamics. The delay in network adaptation may result in unnecessary disruptions to the streaming and affect the quality of service of the application.
11.	[17]	Multicast Tree Protocol	This algorithm results in efficient data transfer between the source to destination in terms of reduced delay.	The node dynamics has not been considered in constructing the tree and hence node churn would result in broken trees affecting the downstream nodes.
12.	[18]	Proactive Fault Resilient Overlay Multicast Network	A proactive mechanism has been proposed that keeps the downstream nodes informed when to expect an upstream node leaving for the purpose of reconnection.	The setup shown in this algorithm is very artificial as the nodes need to inform the other nodes their duration of stay at the beginning itself. This against the basic spirit of the overlay network where nodes can join and leave the network at their choice. Arranging the nodes in the in the reverse order of the duration of the stay is very impractical as it may result in an inefficient structure if the longest staying node is very far away from the source. Node dynamics has not been properly considered in this design as the affected nodes need to rejoin the network themselves. A central control would be needed to keep the information about the nodes joining time and duration of stay for proper operation of these protocols.
13.	[19]	Hybrid IP multicast mesh overlay network	Results in a good structure combining both IP multicast and mesh overlay.	This network is not totally independent of the underlying network as it depends on the IP multicast protocol. Establishing a full mesh network is not efficient as it would need a large memory maintain information about each and every node. This problem would become more acute as the network size grows to very large. This would result in scalability problems. Maintaining the full mesh information also has the problem of updating the mesh information table as all the nodes need to update the information table every time a node

				joins or leaves the network. This would result in instability in the network as the network grows.
14.	[20]	mTreebone	This structure show better resilience to node dynamics compared to pure tree structured overlay networks. Also this structure would have lesser load on the backbone tree as it would carry only one data packet at any one time.	This structure has the shortcoming of data duplicates of content and unnecessary congestion in the local network in managing the mesh network at the local level. The maintenance of the mesh network is also more expensive as it needs constant updates about the node structure and availability and requires large memory to maintain the mesh information on each and every mesh node.
15.	[21]	Distributed Algorithm for HARP	This structure optimizes the routing between the source node and the receiving nodes in such a manner that the total latency is reduced. This results in the better QoS in terms of reducing the maximum delay between the source and the clients.	The node dynamics has not been considered in constructing the tree and hence node churn would result in broken trees affecting the downstream nodes.

III. CONCLUSION

In this work, the author has taken a critical look at the literature on multicasting over overlay networks. Multicasting is one of the most promising applications over the Internet as it requires relatively lower overhead to serve a large number of clients compared to the traditional unicast or broadcast applications. Under the most optimum conditions, multicast systems will have only one data packet in any part of the network irrespective of how many clients are served downstream. Also, the multicast server will transmit only one data packet irrespective of how many clients receive the copies of such packets. In traditional multicast networks, multicast routers placed at strategic locations duplicate the packets depending on the requests they receive. Deploying multicast routers throughout the Internet is not practical due to the amount of investment required for such an operation. Hence researchers have recently focused their attention on using overlay networks to realize the goal of implementing multicast applications in the Internet. By implementing multicasting over application layer overlay networks, the function of duplicating packets has been moved from the network layer (Internet Protocol layer) to application layer.

One of the most challenging tasks in overlay networks in the management of node dynamics. In overlay networks, nodes can join and leave the network at their will. In traditional multicast networks deployed using multicast routers, node dynamics is not a major concern as the infrastructure is considered to be stable and only the leaf or end nodes join and leave the network. The churn of end nodes will not affect any other client node as they are not dependant on each other.

Using overlay networks for multicasting presents a new challenge as the end nodes are required to play a dual role of clients as well as forwarding agents to other client nodes downstream. Node dynamics will have different effects on the end user applications depending on the type of application. Real time applications will be more affected by node dynamics than non-real time applications due to the disruptions resulting from such dynamics.

In this work, media streaming has been selected as the application to be run over the overlay network for the purpose of testing the quality of the overlay networks. Media streaming has been selected as the application due to its popularity and the stringent QoS requirements that are required to be met for successful deployment of such applications in the Internet. Media streaming requires non disrupted flow of packets from the source to destination. Hence managing node dynamics is very important for successful implementation of these applications on the overlay networks.

In this paper, the author presents a critical review of systems, algorithms and mechanism proposed in the recent literature. Special attention has been paid to the advantages and disadvantages of these proposed systems with respect to managing node dynamics. The paper looks at each proposal critically on the mechanism proposed, their strengths and weaknesses. Finally, the results of the analysis have been presented in a table for easy reference.

REFERENCES

[1] Katsumi Tanaka, "Research on Fusion of the Web and TV Broadcasting," in 2nd International Conference on Informatics Research for Development of Knowledge Society Infrastructure (ICKS 2007), Kyoto, Japan, 2007, pp. 129-136.

[2] Luis Montalvo et al., "Implementation of a TV studio based on Ethernet and the IP protocol stack," in IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB '09), Bilbao, Spain, 2009, pp. 1-7.

[3] Steven M Cherry, "Web Radio: Time to Sign Off?," IEEE Spectrum, p. 1, 2002.

[4] Yongkang Peng and Yilai Zhang, "Research and application of the intranet-based multimedia education platform," in International Conference on Networking and Digital Society, Wenzhou, China, 2010, pp. 294 - 297.

[5] Yang-hua Chu, Sanjay G Rao, Srinivasan Seshan, and Hui Zhang, "A Case for End System Multicast," IEEE Journal on Selected Areas in Communications, vol. 20, no. 8, pp. 1456-1471, October 2002.

[6] Alf Inge Wang and Peter Nicolai Motzfeldt, "Peer2Schedule - an experimental peer-to-peer application to support present collaboration," in International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2007), New York, NY, USA, 2007, pp. 418-425.

[7] Loïc Baud and Patrick Bellot, "A purpose adjustable overlay network," in 6th EURO-NF Conference on Next Generation Internet (NGI), , Paris, France, 2010, pp. 1-8.

[8] Shiping Chen, Baile Shi, Shigang Chen, and Ye Xia, "ACOM: Any-source Capacity-constrained Overlay Multicast in Non-DHT P2P

Networks," IEEE Transactions on Parallel and Distributed Systems, vol. 18, no. 9, pp. 1188-1201, September 2007.

[9] Jiemin Liu and Yuchun Ma, "Implementing Overlay Multicast with SCTP," in 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM), Chengdu, China, 2010, pp. 1-4.

[10] Jun Makishi, Debasish Chakraborty, Toshiaki Osada, Gen Kitagata, and Atushi Takeda, "A Fair Load Sharing Scheme for Multi-source Multimedia Overlay Application Layer Multicast," in International Conference on Complex, Intelligent and Software Intensive Systems, 2010, pp. 337-344.

[11] Ted H Szymanski and Dave Gilbert, "Internet Multicasting of IPTV With Essentially-Zero Delay Jitter," IEEE Transactions on Broadcasting, vol. 55, no. 1, pp. 20-30, March 2009.

[12] Eng Keong Lua, Xiaoming Zhou, Jon Crowcroft, and Piet Van Mieghem, "Scalable multicasting with network-aware geometric overlay," Journal of Computer Communications, vol. 31, no. 3, pp. 464-488, February 2008.

[13] Dario Pompili, Caterina Scoglio, and Luca Lopez, "Multicast algorithms in service overlay networks," Journal of Computer Communications 31 (2008) 489–505, vol. 31, no. 3, pp. 489-505, February 2008.

[14] Zhaoping Wang, Xuesong Cao, and Ruimin Hu, "Adapted Routing Algorithm in the Overlay Multicast," in International Symposium on Intelligent Signal Processing and Communication Systems, Xiamen, China, 2007, pp. 634-637.

[15] Mohammed Ali Kaafar, Thierry Turletti, and Walid Dabbous, "A Locating-First Approach for Scalable Overlay Multicast," in 25th IEEE International Conference on Computer Communications (INFOCOM 2006), Barcelona, Spain, 2006, pp. 1-2.

[16] Aron Walters, David Zage, and Cristina Nita-Rotaru, "Mitigating Attacks Against Measurement-Based Adaptation Mechanisms in Unstructured Multicast Overlay Networks," in 14th IEEE International Conference on Network Protocols (ICNP '06), Santa Barbara, CA, USA, 2006, pp. 65 - 74.

[17] Yousef Alipour, Amirmasoud Bidgoli, and Mohsen Moradi, "Approaches for latency reduction in multicast overlay network," in International Conference on Computer Design and Applications (ICCDA), Qinhuangdao, China, 2010, pp. 349-354.

[18] Bhed Bahadur Bista, "A Proactive Fault Resilient Overlay Multicast for Media Streaming," in International Conference on Network-Based Information Systems (NBIS '09), Indianapolis, IN, USA, 2009, pp. 17-23.

[19] Han Gao, Michael E Papka, and Rick L Stevens, "Extending Multicast Communications by Hybrid Overlay Network," in IEEE International Conference on Communications (ICC '06), Istanbul, Turkey, 2006, pp. 820-828.

[20] Feng Wang, Yongqiang Xiong, and Jiangchuan Liu, "mTreebone: A Collaborative Tree-Mesh Overlay Network for Multicast Video Streaming," IEEE Transactions on Parallel and Distributed Systems, vol. 21, no. 3, pp. 379-392, March 2010.

[21] Jun Guo and Sanjay Jha, "Host-aware routing in multicast overlay backbone," in IEEE Network Operations and Management Symposium (NOMS 2008), Salvador, Bahia,Brazil, 2008, pp. 915-918.

AUTHOR'S PROFILE

Mohamed Fazil Mohamed Firdhous is a senior lecturer attached to the Faculty of Information Technology of the University of Moratuwa, Sri Lanka. He received his BSc Eng., MSc and MBA degrees from the University of Moratuwa, Sri Lanka, Nanyang Technological University, Singapore and University of Colombo Sri Lanka respectively. In addition to his academic qualifications, he is a Chartered Engineer and a Corporate Member of the Institution of Engineers, Sri Lanka, the Institution of Engineering and Technology, United Kingdom and the International Association of Engineers. Mohamed Firdhous has several years of industry, academic and research experience in Sri Lanka, Singapore and the United States of America.

Adaptive Equalization Algorithms:An Overview

Garima Malik
Electronics & Communication Engineering
Punjabi University Patiala
Punjab, India
garimamalik82@gmail.com

Amandeep Singh Sappal
Electronics & Communication Engineering
Punjabi University Patiala
Punjab, India
sappal73as@pbi.ac.in

***Abstract*—The recent digital transmission systems impose the application of channel equalizers with short training time and high tracking rate. Equalization techniques compensate for the time dispersion introduced by communication channels and combat the resulting inter-symbol interference (ISI) effect. Given a channel of unknown impulse response, the purpose of an adaptive equalizer is to operate on the channel output such that the cascade connection of the channel and the equalizer provides an approximation to an ideal transmission medium. Typically, adaptive equalizers used in digital communications require an initial training period, during which a known data sequence is transmitted. A replica of this sequence is made available at the receiver in proper synchronism with the transmitter, thereby making it possible for adjustments to be made to the equalizer coefficients in accordance with the adaptive filtering algorithm employed in the equalizer design. In this paper, an overview of the current state of the art in adaptive equalization techniques has been presented.**

Keywords- Channel Equalizer; Adaptive Equalize; Least Mean Square; Recursive Least Squares.

I. INTRODUCTION

One of the most important advantages of the digital transmission systems for voice, data and video communications is their higher reliability in noise environment in comparison with that of their analog counterparts. Unfortunately most often the digital transmission of information is accompanied with a phenomenon known as intersymbol interference (ISI) [1]. Briefly this means that the transmitted pulses are smeared out so that pulses that correspond to different symbols are not separable. Depending on the transmission media the main causes for ISI are:

- cable lines – the fact that they are band limited;
- cellular communications – multipath propagation

Obviously for a reliable digital transmission system it is crucial to reduce the effects of ISI and it is where the equalizers come on the scene. The need for equalizers [2] arises from the fact that the channel has amplitude and phase dispersion which results in the interference of the transmitted signals with one another. The design of the transmitters and receivers depends on the assumption of the channel transfer function is known. But, in most of the digital communications applications, the channel transfer function is not known at enough level to incorporate filters to remove the channel effect at the transmitters and receivers.For example, in circuit switching communications, the channel transfer function is usually constant, but, it changes for every different path from the transmitter to the receiver. But, there are also non stationary channels like wireless communications. These channels' transfer functions vary with time, so that it is not possible to use an optimum filter for these types of channels. So, In order to solve this problem equalizers are designed. Equalizer is meant to work in such a way that BER (Bit Error Rate) should be low and SNR (Signal-to-Noise Ratio) should be high. Equalizer gives the inverse of channel to the Received signal and combination of channel and equalizer gives a flat frequency response and linear phase [1] [4] shown in figure 1.

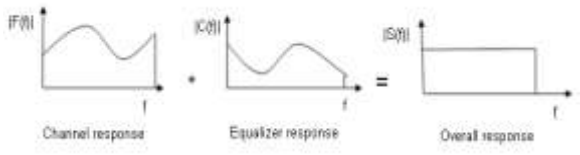


Figure 1. Concept of equalizer

The static equalizer is cheap in implementation but its noise performance is not very good [3]-[20]. As it is told before, most of the time, the channels and, consequently, the transmission system's transfer functions are not known. Also, the channel's impulse response may vary with time. The result of this is that the equalizer cannot be designed. So, mostly preferred scheme is to exploit adaptive equalizers. An adaptive equalizer is an equalization filter that automatically adapts to time-varying properties of the communication channel. It is a filter that self-adjusts its transfer function according to an optimizing algorithm.

The rest of this paper is organized as follows. In section II, discussed the basic concept of transversal equalizers are introduced followed by a simplified description of some practical adaptive equalizer. Different adaptation algorithms are discussed in section III. Many methods exist for both the identification and correction processes of adaptive equalization are presented in section IV. The application of adaptive filtering are covered briefly in section V. The conclusion along with future research directions are discussed in Section VI.

II. CHANNEL EQUALIZATION

As mentioned in the introduction the intersymbol interference imposes the main obstacles to achieving increased digital transmission rates with the required accuracy. ISI problem is resolved by channel equalization [5] in which the aim is to construct an equalizer such that the impulse response

of the channel/equalizer combination is as close to $z^{-\Delta}$ as possible, where Δ is a delay. Frequently the channel parameters are not known in advance and moreover they may vary with time, in some applications significantly. Hence, it is necessary to use the adaptive equalizers, which provide the means of tracking the channel characteristics. The following figure shows a diagram of a channel equalization system.

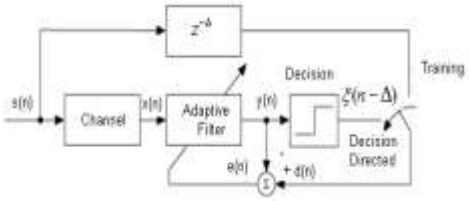


Figure2. Digital transmission system using channel equalization

In the previous figure, s(n) is the signal that you transmit through the communication channel, and x(n) is the distorted output signal. To compensate for the signal distortion, the adaptive channel equalization system completes the following two modes:

- Training mode - This mode helps you determine the appropriate coefficients of the adaptive filter. When you transmit the signal $s(n)$ to the communication channel, you also apply a delayed version of the same signal to the adaptive filter. In the previous figure, $z^{-\Delta}$ is a delay function and $d(n)$ is the delayed signal, $y(n)$ is the output signal from the adaptive filter and e(n) is the error signal between $d(n)$ and $y(n)$. The adaptive filter iteratively adjusts the coefficients to minimize $e(n)$ .After the power of e(n) converges, $y(n)$ is almost identical to $d(n)$,which means that you can use the resulting adaptive filter coefficients to compensate for the signal distortion.
- Decision-directed mode - After you determine the appropriate coefficients of the adaptive filter, you can switch the adaptive channel equalization system to decision-directed mode. In this mode, the adaptive channel equalization system decodes the signal and $y(n)$ produces a new signal, which is an estimation of the signal $s(n)$ except for a delay of Δ taps [6].

Here, Adaptive filter plays an important role. The structure of the adaptive filter [7] is showed in Fig.3

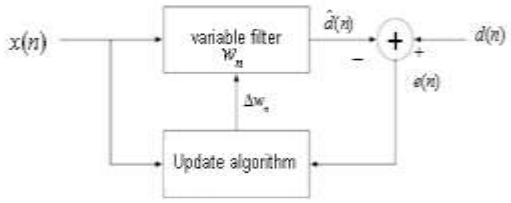


Figure3. Adaptive filter

To start the discussion of the block diagram we take the following assumptions:

The input signal is the sum of a desired signal $d\ (n)$ and interfering noise v (n)

$$x(n) = d(n) + v(n) \tag{1}$$

The variable filter has a Finite Impulse Response (FIR) structure. For such structures the impulse response is equal to the filter coefficients. The coefficients for a filter of order p are defined as

$$w_n = [w_n(0), \quad w_n(1), \quad ..., \quad w_n(p)]^T \tag{2}$$

the error signal or cost function is the difference between the desired and the estimated signal

$$e(n) = d(n) - \hat{d}(n) \tag{3}$$

The variable filter estimates the desired signal by convolving the input signal with the impulse response. In vector notation this is expressed

$$\hat{d}(n) = w_n * x(n) \tag{4}$$

Where

$$x(n) = [x(n), \quad x(n-1), \quad ..., \quad x(n-p)]^T \tag{5}$$

is an input signal vector. Moreover, the variable filter updates the filter coefficients at every time instant

$$w_{n+1} = w_n + \Delta w_n \tag{6}$$

Where Δw_n is a correction factor for the filter coefficients .The adaptive algorithm generates this correction factor based on the input and error signals.

III. Adaptation Algorithms

There are two main adaptation algorithms one is least mean square (LMS) and other is Recursive least square filter (RLS).

A. Least Mean Squares Algorithm (LMS)

Least mean squares (LMS) algorithms are a class of adaptive filter used to mimic a desired filter by finding the filter coefficients that relate to producing the least mean squares of the error signal (difference between the desired and the actual signal). It is a stochastic gradient descent method in that the filter is only adapted based on the error at the current time. LMS filter is built around a transversal (i.e. tapped delay line) structure. Two practical features, simple to design, yet highly

effective in performance have made it highly popular in various application. LMS filter employ, small step size statistical theory, which provides a fairly accurate description of the transient behavior. It also includes H∞ theory which provides the mathematical basis for the deterministic robustness of the LMS filters [1]. As mentioned before LMS algorithm is built around a transversal filter, which is responsible for performing the filtering process. A weight control mechanism responsible for performing the adaptive control process on the tape weight of the transversal filter [9] as illustrated in Figure 4.

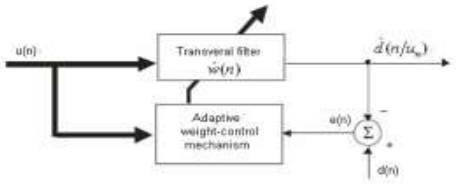


Figure4. Block diagram of adaptive transversal filter employing LMS algorithm

The LMS algorithm in general, consists of two basics procedure:

Filtering process, which involve, computing the output $(d(n-d))$ of a linear filter in response to the input signal and generating an estimation error by comparing this output with a desired response as follows:

$$e(n) = d(n) - y(n) \qquad (7)$$

$y(n)$ is filter output and is the desired response at time n

Adaptive process, which involves the automatics adjustment of the parameter of the filter in accordance with the estimation error.

$$\hat{w}(n+1) = \hat{w}(n) + \mu(u)e^{*}(n) \qquad (8)$$

Where μ is the step size, $(n+1)$ = estimate of tape weight vector at time $(n+1)$ and If prior knowledge of the tape weight vector (n) is not available, set $(n) = 0$

The combination of these two processes working together constitutes a feedback loop, as illustrated in the block diagram of Figure 4. First, we have a transversal filter, around which the LMS algorithm is built; this component is responsible for performing the filtering process. Second, we have a mechanism for performing the adaptive control process on the tap weight of the transversal filter- hence the designated "adaptive weight –control mechanism" in the figure [1].

i. LMS is the most well-known adaptive algorithms by a value that is proportional to the product of input to the equalizer and output error.
ii. LMS algorithms execute quickly but converge slowly, and its complexity grows linearly with the no of weights.
iii. Computational simplicity
iv. In which channel parameter don't vary very rapidly [10].

B. Recursive Least Square Algorithm (RLS)

The Recursive least squares (RLS)[11] adaptive filter is an algorithm which recursively finds the filter coefficients that minimize a weighted linear least squares cost function relating to the input signals. This in contrast to other algorithms such as the least mean squares (LMS) that aim to reduce the mean square error. In the derivation of the RLS, the input signals are considered deterministic, while for the LMS and similar algorithm they are considered stochastic. Compared to most of its competitors, the RLS [11] exhibits extremely fast convergence. However, this benefit comes at the cost of high computational complexity, and potentially poor tracking performance when the filter to be estimated changes.

As illustrated in Figure 5, the RLS algorithm has the same to procedures as LMS algorithm, except that it provides a tracking rate sufficient for fast fading channel, moreover RLS algorithm is known to have the stability issues due to the covariance update formula p(n) [13], which is used for automatics adjustment in accordance with the estimation error as follows: [14].

$$p(0) = \delta^{-1} I \qquad (9)$$

Where p is inverse correlation matrix and δ is regularization parameter, positive constant for high SNR and negative constant for low SNR.

For each instant of time $n = 1, 2, 3.....$

$$\pi(n) = p(n-1)u(n) \qquad (10)$$

$$k(n) = \frac{\pi(n)}{\lambda + u^{H}(n)\pi(n)} \qquad (11)$$

Time varying gain vector

$$\xi(n) = d(n) - \hat{w}^{H}(n-1)u(n) \qquad (12)$$

Then the priori estimation error

$$\hat{w}(n) = \hat{w}(n-1) + k(n)\xi^{*}(n) \qquad (13)$$

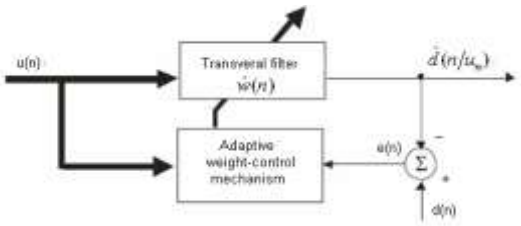


Figure5. Block diagram of adaptive transversal filter employing RLS algorithm

IV. ADAPTIVE EQUALIZATION TECHNIQUES

Different kinds of Equalizer are available in the text like Fractionally Spaced Equalizer, Blind Equalization, Decision-Feedback Equalization, Linear Phase Equalizer, T-Shaped Equalizer, Dual Mode Equalizer and Symbol spaced Equalizer. But most widely used equalizers are discussed as follow:

A. Symbol Spaced Equalizer

A symbol-spaced linear equalizer consists of a tapped delay line that stores samples from the input signal. Once per symbol period, the equalizer outputs a weighted sum of the values in the delay line and updates the weights to prepare for the next symbol period. This class of equalizer is called symbol-spaced [20] because the sample rates of the input and output are equal. In this configuration the equalizer is attempting to synthesize the inverse of the folded channel spectrum which arises due to the symbol rate sampling (aliasing) of the input. In typical application, the equalizer begins in training mode together information about the channel, and later switches to decision –directed mode. Here, The new set of weights depends on these quantities:

- The current set of weights
- The input signal
- The output signal
- For adaptive algorithms other than CMA, a reference signal, d, whose characteristics depend on the operation mode of the equalizer.

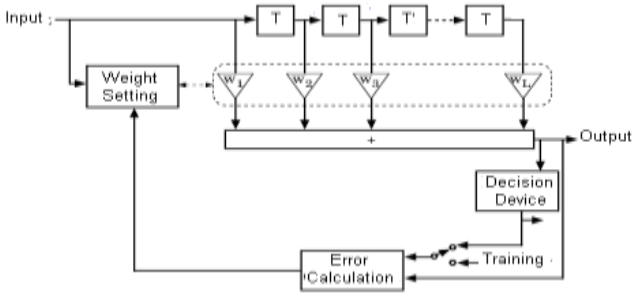


Figure6. Symbol spaced Equalizer

B. Decision-Feedback Equalization

The basic limitation of a linear equalizer, such as transversal filter, is the poor perform on the channel having spectral nulls. A decision feedback equalizer (DFE) is a nonlinear equalizer that uses previous detector decision to eliminate the ISI on pulses that are currently being demodulated. In other words, the distortion on a current pulse that was caused by previous pulses is subtracted. Figure 7 shows a simplified block diagram of a DFE where the forward filter and the feedback filter can each be a linear filter, such as transversal filter. The nonlinearity of the DFE stems from the nonlinear characteristic of the detector that provides an input to the feedback filter. The basic idea of a DFE[19] is that if the values of the symbols previously detected are known, then ISI contributed by these symbols can be cancelled out exactly at the output of the forward filter by subtracting past symbol values with appropriate weighting. The forward and feedback tap weights can be adjusted simultaneously to fulfill a criterion such as minimizing the MSE. The DFE structure is particularly useful for equalization of channels with severe amplitude distortion, and is also less sensitive to sampling phase offset. The improved performance comes about since the addition of the feedback filter allows more freedom in the selection of feed forward coefficients. The exact inverse of the channel response need not be synthesized in the feed forward filter, therefore excessive noise enhancement is avoided and sensitivity to sampler phase is decreased.

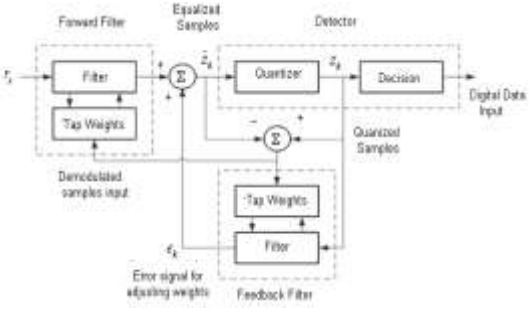


Figure7. Decision feedback equalizer

The advantage of a DFE implementation is the feedback filter, which is additionally working to remove ISI, operates on noiseless quantized levels, and thus its output is free of channel noise.

One drawback to the DFE structure surfaces when an incorrect decision is applied to the feedback filter. The DFE output reflects this error during the next few symbols as the incorrect decision propagates through the feedback filter. Under this condition, there is a greater likelihood for more incorrect decisions following the first one, producing a condition known as error propagation. On most channels of interest the error rate is low enough that the overall performance degradation is slight.

C. Blind Equalization

Blind equalization of the transmission channel has drawn achieving equalization without the need of transmitting a training signal. Blind equalization algorithms that have been proposed are the constant modulus algorithm (CMA) and multimodal's algorithm (MMA). This reduces the mean-squared error (MSE) to acceptable levels. Without the aid of training sequences, a blind equalization is used as an adaptive equalization in communication systems.

In the blind equalization algorithm, the output of the equalizer is quantized and the quantized output is used to update the coefficients of the equalizer. Then, for the complex signal case, an advanced algorithm was presented in. However, the convergence property of this algorithm is relatively poor.[15] In an effort to overcome the limitations of decision directed equalization, the desired signal is estimated at the receiver using a statistical measure based on a priori symbol properties, this technique is referred to as blind equalization [16]. In the Blind Error is selected as the basis for the filter coefficient update. In general, blind equalization directs the coefficient adaptation process towards the optimal filter parameters even when the initial error rate is large. For best results the error calculation is switched to decision directed method after an initial period of equalization, we call this the shift blind method. Referring to, the Reference Selector selects the Decision Device Output as the input to the error calculation and the Error Selector selects the Standard Error as the basis for the filter coefficient update.

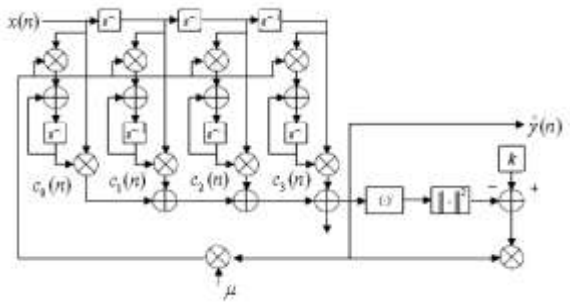


Figure8. Blind Equalization

D. Fractionally Spaced Equalizer

A fractionally spaced equalizer [17] is a linear equalizer that is similar to a symbol-spaced linear equalizer. By contrast, however, a fractionally spaced equalizer receives say K input samples before it produces one output sample and updates the weights, where K is an integer. In many applications, K is 2. The output sample rate is 1/T, while the input sample rate is K/T. The weight-updating occurs at the output rate, which is the slower rate. Sometimes the input to the equalizer is oversampled such that the sample interval is shorter than the symbol interval and the resulting equalizer is said to be fractionally spaced. Equalizer Taps are spaced closer than the reciprocal of symbol rate. Advantages of FSE are it has ability to be not affected by aliasing problem, shows fast convergence and sample rate is less the symbol rate [18].

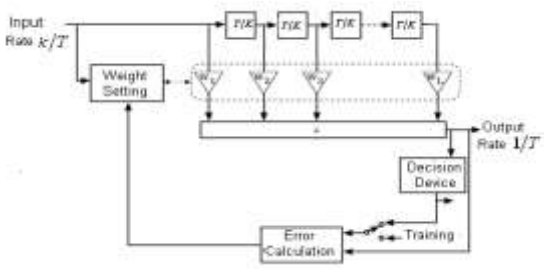


Figure9. Fractional spaced equalizer

V. Applications of Adaptive Filtering

The ability of an adaptive filter to operate satisfactorily in an unknown environment and track time variations of input statistics makes the adaptive filter a powerful device for signal processing and control applications. Indeed, adaptive filters have been successfully applied in such devices fields as communications, radar, sonar and biomedical engineering. Although these applications are quite different in nature, nevertheless, they have one basic feature in common: An input vector and a desired response are used to compute an estimation error, which is in turn used to control the values of a set of adjustable filter coefficients. The adjustable coefficients may take the form of tap weights reflection coefficients, or rotation parameters, depending on the filter structure employed. However, the essential difference between the various applications of adaptive filtering arises in the manner in which the desired response is extracted. In this context, we may distinguish four basic classes of adaptive filtering application [1].

TABLE I. FOUR BASIC CLASSES OF ADAPTIVE FILTERING APPLICATION

Class of adaptive filtering	*Application*	*Purpose*
Identification	System identification	Given an unknown dynamical system, the purpose system identification is to design an adaptive filter that provides an approximation to the system.
	Layered earth modeling	In exploration seismology, a layered modeled of the earth is developed to unravel the complexities of the earth's surface.
Inverse modeling	Equalization	Given a channel of unknown impulse response, the purpose of an adaptive equalizer is to operate on the channel output such that the cascade connection of the channel and the equalizer provides an approximation to an ideal transmission medium.
Prediction	Predictive coding	The adaptive prediction is used to develop a model of a signal of interest (e.g., a speech signal); rather than encode the signal directly, in predictive coding the prediction error is encoded for transmission or storage. Typically, the prediction error has a smaller variance than the original signal-Hence the basis for improved encoding.
	Spectrum analysis	In this application, predictive modeling is used to estimate the power spectrum of a signal of interest.
Interference cancellation	Noise cancellation	The purpose of an adaptive noise canceller is to subtract noise from a received signal in adaptively controlled manner so as to improve the signal-to-noise ratio. Echo cancellation, experienced on telephone circuits, is a special form of noise cancellation.Noise cancellation is also used in electrocardiography.
	Beamforming	A beamforming is spatial filter consist of an array of antenna elements with adjustable weights (coefficients). The twin purposes of an adaptive beamformer are to adaptively control the weights so as to cancel interfering signal

		impinging on the array from unknown direction and, at the same time, provide protection to a target signal of interest.

VI. CONCLUSION

The continuing evolution of communication standards and competitive pressure in the market-place dictate that communication system architects must start the engineering design and development cycle while standards are still in a fluid state. Third and future generation communication infrastructure must support multiple modulation formats and air interface standards. FPGAs provide the flexibility to achieve this goal, while simultaneously providing high levels of performance. The SDR implementation of traditionally analog and digital hardware functions opens-up new levels of service quality, channel access flexibility and cost efficiency. In this paper, the several aspects of adaptive equalizer for communication systems are reviewed. Bandwidth-efficient data transmission over telephone and radio channels is made possible by the use of adaptive equalization to compensate for the time dispersion introduced by the channel. Spurred by practical applications, a steady research effort over the last two decades has produced a rich body of literature in adaptive equalization and the related more general fields of reception of digital signals, adaptive filtering, and system identification. There is still more work to be done in adaptive equalization of nonlinearities with memory and in equalizer algorithms for coded modulation systems. However, the emphasis has already shifted from adaptive equalization theory toward the more general theory and applications of adaptive filters, and toward structures and implementation technologies which are uniquely suited to particular applications.

ACKNOWLEDGMENTS

The authors would thanks the reviewers for their help in improving the document.

REFERENCES

[1] Analog and Digital Communications, S.Haykins, Prentice Hall, 1996.

[2] Modern Digital and analog Communications system, B.P.Lathi

[3] S. U. H. Qureshi, "Adaptive equalization," Proceedings of the IEEE 1985, vol. 73, no. 9, pp. 1349-1387.

[4] B. Petersen, D. Falconer, "Suppression of adjacent-channel, cochannel and intersymbol interference by equalizers and linear combiners," IEEE Trans on Communication, vol. 42, pp. 3109-3118, Dec. 1994

[5] Adinoyi, S. Al-Semari, A. Zerquine, "Decision feedback equalisation of coded I-Q QPSK in mobile radio environments," Electron. Lett. vol. 35, No1, pp. 13-14, Jan. 1999.

[6] Wang Junfeng, Zhang Bo, "Design of Adaptive Equalizer Based on Variable Step LMS Algorithm," Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10) Jiaozuo, P. R. China, 14-15, pp. 256-258, August 2010.

[7] Antoinette Beasley and Arlene Cole-Rhodes, "Performance of Adaptive Equalizer for QAM signals," IEEE, Military communications Conference, 2005. MILCOM 2005, Vol.4, pp. 2373 - 2377.

[8] En-Fang Sang, Hen-Geul Yeh, "The use of transform domain LMS algorithm to adaptive equalization," International Conference on Industrial Electronics, Control, and Instrumentation, 1993. Proceedings of the IECON '93, Vol.3, pp.2061-2064, 1993.

[9] Kevin Banovic, Raymond Lee, Esam Abdel-Raheem, and Mohammed A. S. Khalid, "Computationally-Efficient Methods for Blind Adaptive Equalization," 48th Midwest Symposium on Circuits and Systems, Vol.1, pp.341-344, 2005.

[10] En-Fang Sang, Hen-Geul Yeh, "The use of transform domain LMS algorithm to adaptive equalization," International Conference on Industrial Electronics, Control, and Instrumentation, 1993. Proceedings of the IECON '93, Vol.3, pp.2061-2064, 1993.

[11] Mahmood Farhan Mosleh, Aseel Hameed AL-Nakkash, "Combination of LMS and RLS Adaptive Equalizer for Selective Fading Channel," European Journal of Scientific Research ISSN 1450-216X Vol.43, No.1, pp.127-137, 2010.

[12] C Chen, C., Chen, K. and Chiueh, T. 2003 "Algorithm and Architecture Design for a low Complexity Adaptive Equalizer," National Taiwan University, Taipei, Taiwan, 0-7803-7761- 31031S17.0, IEEE.

[13] Iliev,G and Kasabov, N. 2000 "Channel Equalization using Adaptive Filtering with Averaging," 5th Joint coference on Information Science (JCIS), Atlantic city, USA., 2000.

[14] Sharma, O., Janyani, V and Sancheti, S. 2009. "Recursive Least Squares Adaptive Filter a better ISI Compensator," International Journal of Electronics, Circuits and Systems.

[15] J Yang, J.-J. Werner, G. A. Dumont, "The multimodal's blind equalization and its generalized algorithms," IEEE Journal on selected areas in communication, vol.20, no.5, pp. 997-1015, 2002.

[16] 10 D. N. Godard, "Self-recovering equalization and Carrier Tracking in Two-dimensional Data communications Systems," IEEE Trans. Communication, vol. COM-28, November 1980

[17] J.R. Treichler, I. Fijalkow, and C.R. Johnson, JR., "Fractionally Spaced Equalizers," IEEE Signal Processing Magazine, Vol.13, Isssue: 3, pp.65-81, 1996.

[18] Kevin Banovi´c, Mohammed A. S. Khalid and Esam Abdel-Raheem, "A Configurable Fractionally-Spaced Blind Adaptive Equalizer for QAM Demodulators," Digital Signal Processing, Vol. 17, No.6, pp. 1071-1088, Nov. 2007.

[19] Vladimir D. Orlic, Miroslav Lutovac, "A solution for efficient reduction of intersymbol interference in digital microwave radio," Telsiks 2009, pp.463-464.

[20] Antoinette Beasley and Arlene Cole-Rhodes, "Performance of Adaptive Equalizer for QAM signals," IEEE, Military Communications Conference, 2005. MILCOM 2005, Vol.4, pp. 2373 - 2377.

[21] Haque, A. K. M. F. (2010). Multipath Fading Channel Optimization for Wireless Medical Applications. International Journal of Advanced Computer Science and Applications - IJACSA, 1(4), 33-36.

[22] Bappy, D. M., Dey, A. K., & Saha, S. (2010). OFDM System Analysis for reduction of Inter symbol Interference Using the AWGN Channel Platform. International Journal of Advanced Computer Science and Applications - IJACSA, 1(5), 123-125.

AUTHORS PROFILE

Garima Malik received her Bachelor of Engineering B.E and pursuing Master of Technology M.Tech from Punjabi University Patiala, India.

Amandeep Singh Sappal is currently his Ph.D in Electronic & Communication from Punjabi University Patiala and presently he is working as an Assistant Lecturer in Punjabi University Patiala, India.

Pulse Shape Filtering in Wireless Communication-A Critical Analysis

A. S Kang
Assistant Professor, Deptt of Electronics and Communication Engg
SSG Panjab University Regional Centre
Hoshiarpur, Punjab, INDIA
askang_85@yahoo.co.in

Vishal Sharma
Assistant Professor, Deptt of Electronics and Communication Engg
University Institute of Engineering and Technology
Panjab University Chandigarh - 160014, INDIA
vishaluiet@yahoo.co.in

***Abstract*—The goal for the Third Generation (3G) of mobile communications system is to seamlessly integrate a wide variety of communication services. The rapidly increasing popularity of mobile radio services has created a series of technological challenges. One of this is the need for power and spectrally efficient modulation schemes to meet the spectral requirements of mobile communications. Pulse shaping plays a crucial role in spectral shaping in the modern wireless communication to reduce the spectral bandwidth. Pulse shaping is a spectral processing technique by which fractional out of band power is reduced for low cost, reliable , power and spectrally efficient mobile radio communication systems. It is clear that the pulse shaping filter not only reduces inter-symbol interference (ISI), but it also reduces adjacent channel interference. The present paper deals with critical analysis of pulse shaping in wireless communication.**

Keywords- WCDMA, Pulse Shaping;

I. INTRODUCTION

Third generation cellular systems are being designed to support wideband services like high speed Internet access, video and high quality image transmission with the same quality as the fixed networks. Research efforts have been underway for more than a decade to introduce multimedia capabilities into mobile communications. Different standard agencies and governing bodies are trying to integrate a wide variety of proposals for third generation cellular systems. [1-7]

One of the most promising approaches to 3G is to combine a Wideband CDMA (WCDMA) air interface with the fixed network of GSM. Several proposal supporting WCDMA were submitted to the International Telecommunication Union (ITU) and its International Mobile Telecommunications for the year 2000 (IMT2000) initiative for 3G. All these schemes try to take advantage of the WCDMA radio techniques without ignoring the numerous advantages of the already existing GSM networks. The standard that has emerged is based on ETSI's Universal Mobile Telecommunication System (UMTS) and is commonly known as UMTS Terrestrial Radio Access (UTRA) [1]. The access scheme for UTRA is Direct Sequence Code Division Multiple Access (DSCDMA). The information is spread over a band of approximately 5 MHz. This wide bandwidth has given rise to the name *Wideband* CDMA or WCDMA.[8-9]

The future mobile systems should support multimedia services. WCDMA systems have higher capacity, better properties for combating multipath fading, and greater flexibility in providing multimedia services with defferent transmission rate and different QoS requirements and has been investigate worldwide [4-6]. CDMA mobile systems are interference limited and therefore reducing interference can directly increase system capacity.

II. WIDEBAND CDMA

In WCDMA, the CDMA technology (Code Division Multiple Access) air interface is implemented along with GSM networks. WCDMA is a Third Generation technology and works towards the interoperability between the various 3G technologies and networks being developed worldwide. WCDMA transmits on a 5 MHz wide radio channel and hence is called Wideband CDMA. This 5 MHz wide radio channel is in contrast to CDMA which transmits in one or more pairs of 1.25 MHz wide radio channel. WCDMA uses Direct Sequence spreading, where spreading process is done by directly combining the baseband information to high chip rate binary code. The Spreading Factor is the ratio of the chips (UMTS = 3.84Mchips/s) to baseband information rate. Spreading factors vary from 4 to 512 in FDD UMTS. Spreading process gain can in expressed in dBs (Spreading factor 128=21dBgain).

Figure 1.shows the CDMA spreading

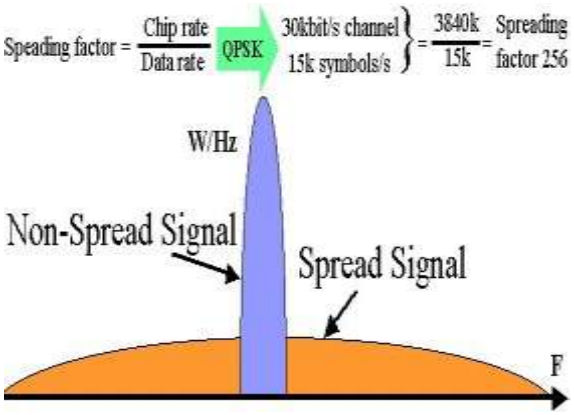


Figure 3.CDMA spreading

A. *Pulse Shape filtering in Wireless Communication*

Linear modulation methods such as QAM,QPSK, OQPSK have received much attention to their inherent high spectral efficiency However for the efficient amplification of transmitted signal, the Radio Frequency Amplifier is normally operated near the saturation region and therefore exhibit non linear behavior. As a result significant spectral spreading occurs, when a signal with large envelope variations propagates through such an amplifier and creates large envelope fluctuations.

To satisfy the ever increasing demands for higher data rates as well as to allow more users to simultaneously access the network, interest has peaked in what has come to be known as wideband code division multiple access (WCDMA).The basic characteristics of WCDMA waveforms that make them attractive for high data rate transmissions are their advantages over other wireless systems. It emphasizes that how the choice of spread bandwidth affects the bit error rate of system [10]

III. MULTIPATH EFFECTS

Multipath propagation through linear dispersive media introduces distortion in signal during the wireless transmission. Due to this, there is degradation in BER performance, unless it is compensated with some suitable techniques at the receiver. In addition to frequency selectivity, the wireless channel also experiences time variations, which arise due to relative motion between transmitter and receiver which in turn needs to acquire mobile channel states and needs to be optimized. In a digital communication system, digital information can be sent on a carrier through changes in its fundamental characteristics such as: phase, frequency, and amplitude. In a physical channel, these transitions can be smoothed, depending on the filters implemented in transmission. In fact, the use of a filter plays an important part in a communications channel because it is effective at eliminating spectral leakage, reducing channel width, and eliminating interference from adjacent symbols (Inter Symbol Interference, ISI)[1-9].Transmitting a signal at high modulation rate through a band-limited channel can create intersymbol interference. As the modulation rate increases, the signal's bandwidth increases. When the signal's bandwidth becomes larger than the channel bandwidth, the channel starts to introduce distortion to the signal. This distortion is usually seen as intersymbol interference. The signal's spectrum is determined by the pulse shaping filter used by the transmitter. Usually the transmitted symbols are represented as a time sequence of dirac delta pulses.

IV. NEED OF EFFICIENT PULSE SHAPING

In communications systems, two important requirements of a wireless communications channel demand the use of a pulse shaping filter. These requirements are:

a) Generating band limited channels, and

b) Reducing Inter Symbol Interference (ISI) arising from multi-path signal reflections.

Both requirements can be accomplished by a pulse shaping filter which is applied to each symbol. In fact, the sinc pulse, shown below, meets both of these requirements because it efficiently utilizes the frequency domain to utilize a smaller portion of the frequency domain, and because of the windowing affect that it has on each symbol period of a modulated signal. A sinc pulse is shown below in figure 2 along with an FFT spectrum of the givensignal.

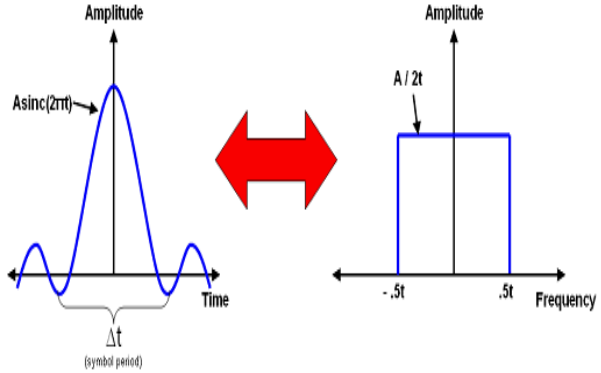


Figure 2 Time vs. Frequency Domain for a Sinc Pulse

The sinc pulse is periodic in nature and is has maximum amplitude in the middle of the symbol time. In addition, it appears as a square wave in the frequency domain and thus can effectively limit a communications channel to a specific frequency range. [10-12]

A. *Reducing Channel Bandwidth*

Fundamentally, modulation of a carrier sinusoid results in constant transitions in its phase and amplitude. Below, figure 3. shows the time domain of a carrier sinusoid with a symbol rate that is half of the carrier. It is clear that phase/amplitude transitions occur at every two periods of the carrier and sharp transitions occur, when filtering is not applied. [10-12]

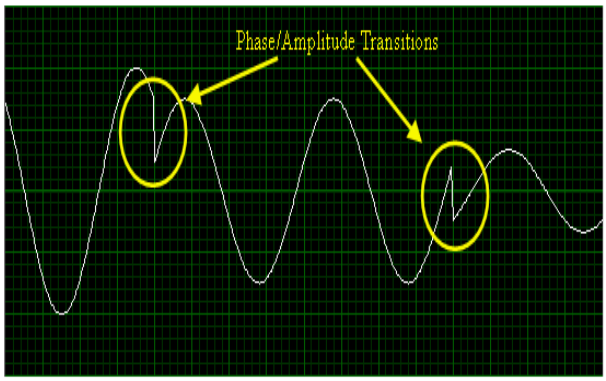


Figure 3. Phase and Amplitude Transitions in an Unfiltered Modulated Signal

The sharp transitions in any signal result in high-frequency components in the frequency domain. By applying a pulse-shaping filter to the modulated sinusoid, the sharp transitions are smoothed and the resulting signal is limited to a specific frequency band. Below, it is shown time-domain modulated sinusoid.

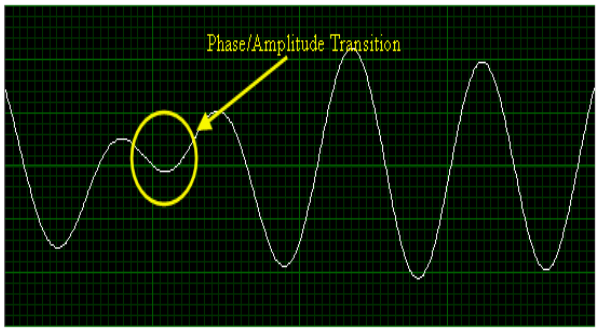


Figure 4. Smoothed Phase and Amplitude Transitions in a Filtered Modulated Signal

Fig 4 shows smoothed phase and amplitude transitions in a filtered modulated signal. It happens much more gradually when filtering is implemented. As a result, the frequency information of the sinusoid becomes more concentrated into a specified frequency band. [10]The sharp transitions do cause high frequency components in the frequency domain. Now, once a filter has been applied to the carrier signal, these high frequency components of the signal have been removed. Thus, the majority of the channel power is now limited to a specific defined bandwidth. It is clear that the required bandwidth for a channel is directly related to the symbol rate and is centered at the carrier frequency.

B. *Reducing Inter-Symbol Interference (ISI)*

In band limited channels, intersymbol interference (ISI) can be caused by multi-path fading as signals are transmitted over long distances and through various mediums. More specifically, this characteristic of the physical environment causes some symbols to be spread beyond their given time interval. As a result, they can interfere with the following or preceding transmitted symbols. One solution to this problem is the application of the pulse shaping filter. By applying this filter to each symbol that is generated, it is possible to reduce channel bandwidth while reducing ISI. In addition, it is common to apply a match filter on the receiver side to minimize these affects. Fig 5 shows the output in time domain. It is clear that the maximum amplitude of the pulse-shaping filter occurs in the middle of the symbol period. In addition, the beginning and ending portions of the symbol period are attenuated. Thus, ISI is reduced by providing a pseudo-guard interval which attenuates signals from multi-pathreflections[10].

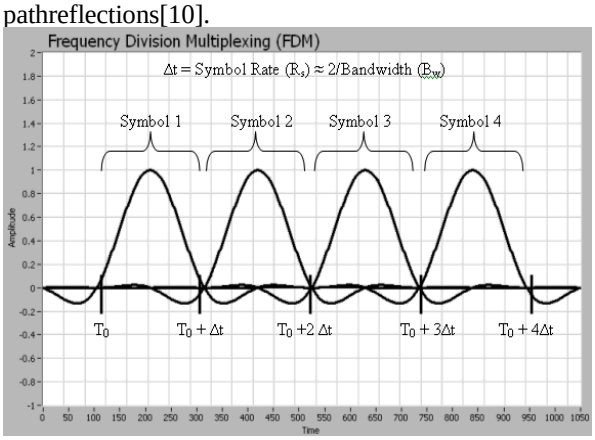


Figure 5 Filter Output in the Time Domain.

C. *Pulse Shaping and Matched Filtering*

The matched filter is perhaps equally as important as the pulse-shaping filter. While the pulse shaping filter serves the purpose of generating signals such that each symbol period does not overlap, the matched filter is important to filter out what signal reflections do occur in the transmission process. Because a direct-path signal arrives at the receiver before a reflected signal does, it is possible for the reflected signal to overlap with a subsequent symbol period. This is shown in the fig 6. It is clear, the matched filter reduces this affect by attenuating the beginning and ending of each symbol period. Thus, it is able to reduce intersymbol interference. One of the most common choices for a matched filter is the root raised cosinefilter[10].

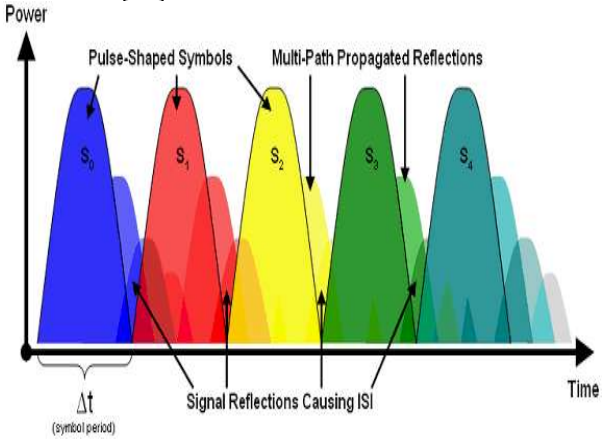


Figure 6. ISI Caused by Multi-Path Distortion

V. DIFFERENT PULSE SHAPES

1) *Rectangular Pulse*
2) *Raised Cosine Pulse*
3) *Square Root Raised Cosine Pulse*
4) *Gaussian Pulse*

5) *Flipped Exponential Pulse*
6) *Flipped Hyperbolic Secant Pulse*
7) *Flipped Inverse Hyperbolic Secant Pulse*

Data transmission systems that must operate in a bandwidth-limited environment must contend with the fact that constraining the bandwidth of the transmitted signal necessarily increases the likelihood of a decoding error at the receiver. Bandwidth limited systems often employ pulse-shaping techniques that allow for bandwidth containment while minimizing the likelihood of errors at the receiver.[12]

A. Rectangular Pulse

The most basic information unit in a digital transmission scheme is a rectangular pulse. It has a defined amplitude, A, and defined duration, T. Such a pulse is shown in Figure 7, where A = 1, T = To, with the pulse centered about the time origin at t = 0. Typically, a sequence of such pulses (each delayed by T seconds relative to the previous one) constitutes the transmission of information. The information, in this case, is encoded in the amplitude of the pulse. The simplest case is when a binary 0 is encoded as the absence of a pulse (A = 0) and a binary 1 is encoded as the presence of a pulse (A = constant). Since each pulse spans the period T, the maximum pulse rate is 1/T pulses per second, which leads to a data transmission rate of 1/T bits per second.[12]

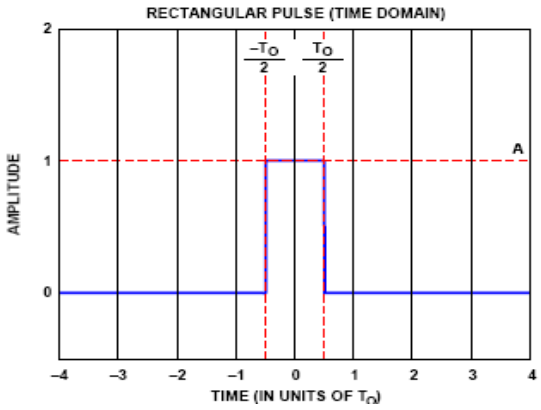


Figure 7. A Single Rectangular Pulse ($T=T_0$, A=1)

The pulses used to transmit symbols occupy a fixed time interval, T .Thus, the pulse rate is 1/T pulses per second, which leads to a symbol rate of 1/T symbols per second. The unit, symbols per second, is often referred to as baud. The data transmission rate in bits per second is the baud rate multiplied by the number of bits represented by each symbol. For example, if a symbol represents four bits, then the bit rate is four times the symbol rate. This means that a lower transmission rate can be used to transmit symbols as opposed to directly transmitting bits, which is the primary reason that the more sophisticated data transmission systems encode groups of bits into symbols. The logic 1 is represented by the presence of a pulse of unit amplitude and a logic 0 by the absence of a pulse (that is, zero amplitude).

B. Raised Cosine Pulse

As shown in Figure 3, the spectrum of a rectangular pulse spans infinite frequency. In many data transmission applications, the transmitted signal must be restricted to a certain bandwidth. This can be due to system design constraints In such instances, the infinite bandwidth associated with a rectangular pulse is not acceptable. The bandwidth of the rectangular pulse can be limited, however, by forcing it to pass through a low-pass filter. The act of filtering the pulse causes its shape to change from purely rectangular to a smooth contour without sharp edges.[12] Therefore, the act of filtering rectangular data pulses is often referred to as pulse shaping. Unfortunately, limiting the bandwidth of the rectangular pulse necessarily introduces a damped oscillation. That is, the rectangular pulse exhibits nonzero amplitude only during the pulse interval, whereas the smoothed (or filtered) pulse exhibits ripples both before and after the pulse interval. At the receiver, the ripples can lead to incorrect decoding of the data, because the ripples associated with one pulse interfere with the pulses before and after it. However, the choice of a proper filter can yield the desired bandwidth reduction while maintaining a time domain shape that does not interfere with the decoding process of the receiver.[12]

This filter is the well-known raised cosine filter and its frequency response is given by

$$H(w)=\tau \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 0\le w\le c$$
$$\tau\{\cos^2[\tau(w-c)/4\alpha]\} \quad \ldots\ldots\ldots\ldots c\le w\le d$$
$$0 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots w>d$$

where w is the radian frequency $2\pi f$,

τ is the pulse period

α is roll off factor

c is equal to $\pi(1-\alpha)/\tau$

d is equal to $\pi(1+\alpha)/\tau$

A plot of the raised cosine frequency response is shown in Figure 8 (normalized to $\tau = 1$). The raised cosine filter gets its name from the shape of its frequency response, rather than its impulse (or time domain) response.

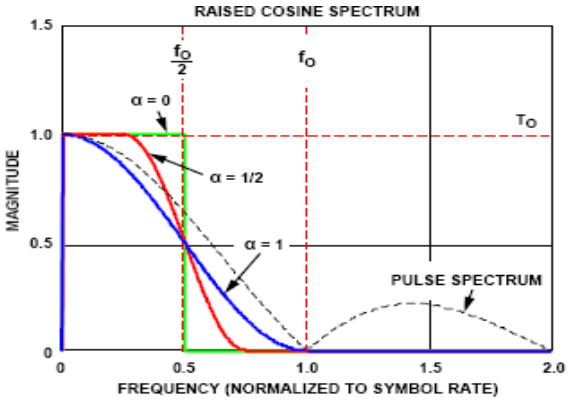


Figure 8. The Raised Cosine Frequency Response

Ideal Response of Raised Cosine Filter is shown in figure 9. below[10]

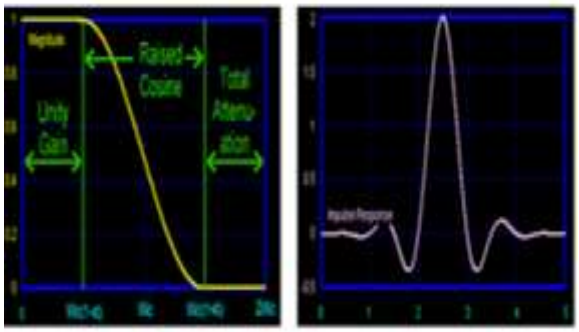


Figure 9. Ideal Response of Raised Cosine Filter

Ideal Raised Cosine filter frequency response consists of unity gain at low frequencies,raised cosine function in the middle and total attenuation at high frequencies.The root raised cosine filter is generally used in series pairs so that total filtering effect is that of raised cosine filter.Sometimes it is desirable to implement the raised cosine response as the product of two identical responses, one at the transmitter and the other at the receiver. In such cases, the response becomes a square-root raised cosine response since the product of the two responses yields the desired raised cosine response.[12]

C. Square Root Raised Cosine

The frequency Response of the Square-Root Raised Cosine is given as below.

$H(w)=\sqrt{\tau}$……0≤w≤c

$\sqrt{\tau}\{\cos[\tau(w-c)/4\alpha]\}$……c≤w≤d

0…….w>d

The variable definitions are the same as for the raised cosine response.The consequence of pulse shaping is that it distorts the shape of the original time domain rectangular pulse into a smoothly rounded pulse with damped oscillations (ripples) before and after the ±½ To points. The ripples result from the convolution of the rectangular pulse with the raised cosine impulse response (convolution is the process of filtering in the time domain).This is a source of decision-making error at the receiver known as Intersymbol Interference (ISI). Reduced bandwidth means larger ripple, which exacerbates ISI and increases the likelihood of an incorrect decision (that is, error) at the receiver.[12] Obviously, a trade off exists between bandwidth containment in the frequency domain and ripple attenuation in the time domain. It is this trade off of bandwidth containment vs. ripple amplitude that must be considered by design engineers when developing a data transmission system that employs pulse shaping.Ideal response of Square Root Raised Cosine filter is shown in figure 10 below.[10]

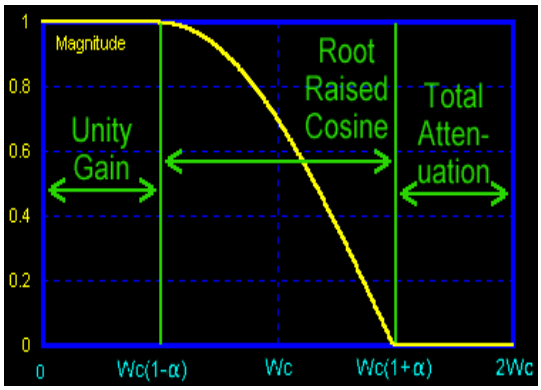


Figure 10. Ideal Square Root Raised Cosine Filter frequency Response

D. Gaussian Pulse

This gives an output pulse shaped like a Gaussian function. The Gaussian filter is a pulse shaping technique that is typically used for Frequency Shift Keying (FSK) and Minimum Shift Keying (MSK) modulation. This filter is unlike the raised cosine and root raised cosine filters because it does not implement zero crossing points. The impulse response for the Gaussian filter is defined bythefollowingequation:

$$h_G(t) = Q\left(\frac{2\pi\alpha}{\sqrt{\ln 2}}\left(n-\frac{1}{2}\right)\right) - Q\left(\frac{2\pi\alpha}{\sqrt{\ln 2}}\left(n+\frac{1}{2}\right)\right)$$

Graphical representation of the impulse response of gaussian filter is shown in figure11. As described above, it is clear that there are no zero crossings for this type of filter[10].

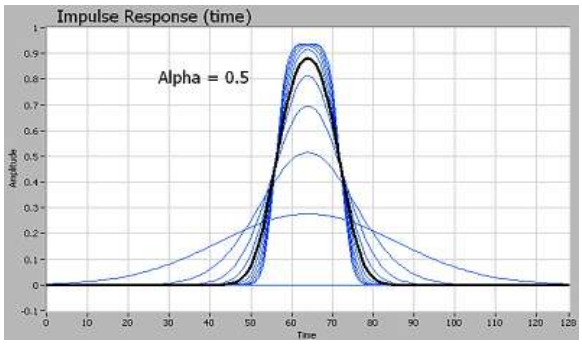


Figure 11. Impulse Response of Gaussian Filter

E. Flipped Exponential Pulse

This pulse is proposed by Beaulieu(fexp)[11]. It is known as flipped-exponential pulse.In this pulse a new parameter,β=ln 2/αβ has been introduced .The frequency and impulse responses of this family are given as below.

$S2(f)=1$……$f\leq B(1-\alpha)$;

$S2(f)=\exp(\beta(B(1-\alpha)-f))$……$B(1-\alpha)<f\leq B$

$S2(f)=1-\exp(\beta(f-B(1+\alpha)))$……...$B<f\leq B(1+\alpha)$

$S2(f)= 0$…… ……$B(1+\alpha)<f$

$S2(t)=(1/T)\text{sinc}(t/T)(4\beta\pi t\sin(\pi\alpha t/T)+2\beta^2\cos(\pi\alpha t/T)-\beta^2)/[(2\pi t)^2+\beta^2]$

Two more pulses have been derived from from flipped exponential pulse[11]

F. Flipped Secant Hyperbolic Pulse

The first one is known as flipped-hyperbolic secant (fsech) pulse, has frequency and impulse responses defined as $S3(f)=1$…….$f\leq B(1-\alpha)$;

$=\text{sech}(\gamma|f|-B(1-\alpha)))$…...$B(1-\alpha)<f\leq B$

$=1-\text{sech}(\gamma(B(1+\alpha)-f))$….$B<f<\leq B(1+\alpha)$

$= 0$……..$B(1+\alpha)<f$

$S3(t)=(1/T)\text{sinc}(t/T)\{8\pi t\sin(\pi\alpha t/T)F_1(t)+2\cos(\pi\alpha t/T)[1-2F_2(t)]+4F_3(t)-1\}$

where $\gamma=\ln(\sqrt{3}+2)/\alpha\beta$

The impulse response has been obtained through an expansion in series of exponentials.

G. Flipped Inverse Secant Hyperbolic Pulse

The second pulse is referred to as flipped-inverse hyperbolic secant (farcsech) pulse. It has frequency response defined as

$S4(f)=1$…….$f\leq B(1-\alpha)$

$=1-(1/2\alpha\beta\gamma)\text{arcsech}(1/2\alpha\beta(B(1+\alpha)-f))$……..$B(1-\alpha)<f\leq B$

$=(1/2\alpha\beta\gamma)\text{arcsech}(1/2\alpha\beta(f-B(1-\alpha)))$……...$B<f\leq B(1+\alpha)$

$= 0$…….$B(1+\alpha)<f$

Its impulse response s4(t)can be evaluated through a numerical inverse Fourier transform. The plot of frequency responses of the above pulses is shown in figure12 below. It is clear that the pulses are real and even, with optimum time sampling.

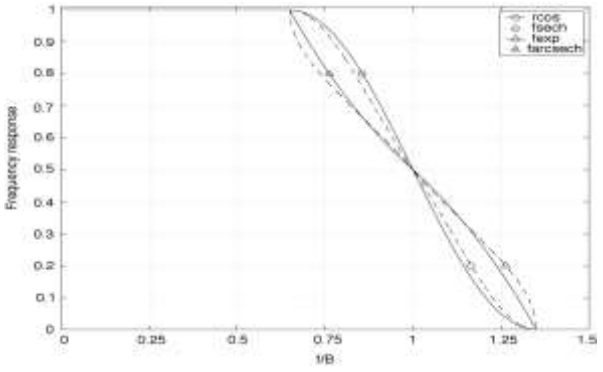


Figure 12. Frequency Responses of the different pulses with roll-off factor α= 0.35

VI. CONCLUSION

Digital Signal processing techniques are being used to improve the performance of 3G systems WCDMA (Wideband Code-Division Multiple Access), an ITU standard derived from Code-Division Multiple Access (CDMA), is officially known as IMT-2000 direct spread spectrum. W-CDMA is a third-generation (3G) mobile wireless technology that promises much higher data speeds to mobile and portable wireless devices than commonly offered in today's market. The pulse shaping filter plays a critical role in WCDMA performance enhancement. [13-20]. **T**he present paper has critically analyzed the pulse shape filtering in wireless communication. The application of signal processing techniques to wireless communications is an emerging area that has recently achieved dramatic improvement in results and holds the potential for even greater results in the future as an increasing number of researchers from the signal processing and communication areas participate in this expanding field.

VII. IMPACT OF STUDY

1. The study is useful to improve the performance of WCDMA Network.
2. In the planning of WCDMA Network.
3. To achieve the flexibility in use of data rates in different environments.
4. Design of future cellular mobile communication network.

VIII. FUTURE SCOPE OF WORK

The following points have to be concentrated to extend its application to wide range of future directions:

1. Different multipath fading channels can replace AWGN channel in the model to simulate the system under different mobile radio channels.
2. More influencing parameters can be incorporated in the simulation model using adaptive signal processing.
3. The simulation model can be developed for Tan, Beaulieu and Damen pulse shaping families by incorporating more variables.
4. DSP algorithms can be developed for performance enhancement of WCDMA based wireless system using optimized values of parameters of pulse shaping filters.
5. Simulation study can be extended to different data rates such as 144 kbps, 384 kbps & 2Mbps.

ACKNOWLEDGEMENT

The first author is thankful to Dr B S Sohi, Director UIET,Sec-25,PANJAB UNIVERSITY Chandigarh for discussion and valuable suggestions during technical discussion at WOC conference at Panjab Engg College Chandigarh during presentation of research paper,"Digital Processing and Analysis of Pulse Shaping Filter for wireless communication "The help rendered by S Harjitpal Singh,a

research scholar of NIT Jalandhar as well as Mr Hardeep Singh,Research Scholar, Communication Signal Processing research Lab, Deptt of Electronics Technology GNDU Amritsar is also acknowledged.

REFERENCES

[1] R.M Piedra and A. Frish, "Digital Signal Processing comes of age" IEEE spectrum, vol 33,No.5,pp70,May 1996.

[2] J Stevens ,"DSPs in Communication "IEEE Spectrum vol.35,No.9,pp39-46,Sept.98.

[3] C. Iniaco and D. Embres," "The DSP decision-fixed or floating?" IEEE Spectrum, Vol.3,No.9,pp 72-74,1996.

[4] P.M. Grant, "Signal Processing Hardware and Software" IEEE signal Processing Magazine,vol13,No.1,pp 86-92,Jan 1992.

[5] G. Giannkis, "Highlights of Signal Processing for communications" IEEE Signal Processing Magazine, vol 16,No.2,pp 14-51,March,1999.

[6] F.Adachi,"Wireless Past and Future-Evolving Mobile Communication Systems,"IEICE Trans Fundamentals,Vol.E84-A,No.1,pp.55-60,Jan2001.

[7] H. Holma and A. Toskala, "WCDMA for UMTS, "John Wiley &Sons.Ltd 2002.

[8] www.3GPP.org

[9] Tero Ojanpera and Ramjee Prasad, "Wideband CDMA for Third Generation Mobile Communications," Artech House, Boston, London,1998.

[10] National Instrument Signal Generator tutorial, Pulse shaping to improve spectral efficiency. www.ni.com/rf.

[11] Antonio Assalini,Andrea M.Tonello,"Improved Nyquist Pulses",IEEE Communication letters ,vol 8 No 2, pp no 87-89, Feb2004.

[12] Ken Gentile,"The care and feeding of digital pulse shaping filters"www.rfdesign.com,pp-50-61**,** 2002.

[13] ASKang,ErVishalSharma"Comparatative Performance Analysis of Pulse Shaping Techniques for Wireless Communication"proceedings of WECON -2008,International Conference on Wireless networks and Embedded Systems"(presesnted)**Chitkara Institute of Engg & Technology,Rajpura,Distt Patiala,** pp 113-115,2008, Oct2008.

[14] ASKang,ErVishalSharma**"**Digital Signal Processing For Cellular Wireless Communication Systems presented at IEEE-ISM 2008,**Indian Institute of Science,Bangalore(INDIA**),pp-323-327, 3Dec-6Dec2008.

[15] A S Kang ,Er Vishal Sharma "Digital Processing and analysis of pulse shaping Filter for wireless Communication",presented at 2nd National Conference(Co –Sponsered by IEEE, Chandigarh Sub Section) On Wireless and Optical Communication WOC-2008 at **PEC Chandigarh**(INDIA) pp110-113, 18-19 Dec,2008.

[16] A S Kang,Er Vishal Sharma "Pulse Shaping Filters in Wireless Communication-A Critical Review"Proceedings of National conference on optical and wireless communication(NCOW-2008, Co-Sponsered by Institution of Engineers(IE), INDIA) **DAVIET,Jalandhar**(INDIA) ,pp 291-297, November 27-28, 2008.

[17] A S Kang,Vishal Sharma "Spectral analysis of filters in celluilar wireless communication systems", International Conference on Systemics,Cybernetics,Informatics,ICSCI-2009,**Hyderabad**,pp-48-55, January7-10,2009.

[18] AS Kang,Vishal Sharma IEEE -0196,"Study of Spectral Analysis of Filters in Cellular Communication Systems" by A S Kang, Vishal Sharma in IEEE International advance Computing Conference (**IACC09**)held at **Thapar University ,Patiala**,(**presented),**pp 1934-1938**,** 6-7March 2009.

[19] ASKang,VishalSharmaIEEE0197,"Simulative Investigation of Pulse shaping in WCDMA Wireless Communication" by A S Kang, Vishal Sharma in IEEE International Advance Computing Conference(**IACC09**) held at **Thapar University ,Patiala(presented),**pp 1939-1943**,**6-7 March2009.

[20] ASKang,Vishal Sharma "Study of pulse shaping fiters in WCDMA under different interferences",**IEEE-SIBCON2009**(IEEE Siberian Conference on Control and Communications)March2009**Russia**,pp37-41,28 March2009.

[21] Adibi, S. (2010). Traffic Classification – Packet- , Flow- , and Application-based Approaches. International Journal of Advanced Computer Science and Applications - IJACSA, 1, 6-15.

[22] Sharma, S. (2010). Efficient Implementation of Sample Rate Converter. International Journal of Advanced Computer Science and Applications - IJACSA, 1(6), 35-41.

Arabic Cursive Characters Distributed Recognition using the DTW Algorithm on BOINC: Performance Analysis

Zied TRIFA, Mohamed LABIDI and Maher KHEMAKHEM
Department of Computer Science, University of Sfax
MIRACL Lab
Sfax, Tunsia
trifa.zied@gmail.com, mohamedlabidi@yahoo.fr
maher.khemakhem@fsegs.rnu.tn

***Abstract*— Volunteer computing or volunteer grid computing constitute a very promising infrastructure which provides enough computing and storage powers without any prior cost or investment. Indeed, such infrastructures are the result of the federation of several, geographically dispersed, computers or/and LAN computers over the Internet. Berkeley Open Infrastructure for Network Computing (BOINC) is considered the most well-known volunteer computing infrastructure. In this paper, we are interested, rather, by the distribution of the Arabic OCR (Optical Character Recognition) based on the DTW (Dynamic Time Warping) algorithm on the BOINC, in order, to prove again that volunteer computing provides very interesting and promising infrastructures to speed up, at will, several greedy algorithms or applications, especially, the Arabic OCR based on the DTW algorithm. What makes very attractive the Arabic OCR based on the DTW algorithm is the following, first, its ability to recognize, properly, words or sub words, without any prior segmentation, from within a reference library of isolated characters. Second, its good immunity against a wide range of noises. Obtained first results confirm, indeed, that the Berkeley Open Infrastructure for Network Computing constitutes an interesting and promising framework to speed up the Arabic OCR based on the DTW algorithm.**

Keywords*— *Volunteer Computing; BOINC; Arabic OCR; DTW algorithm;

I. INTRODUCTION

Arabic OCR based on the DTW algorithm provides very interesting recognition and segmentation rates. One of the advantages of the DTW algorithm is its ability to recognize, properly, words or connected characters without their prior segmentation. In our previous studies achieved on high and medium quality documents [2], [3], [5] we obtained an average of more than 98% as recognition rate and more than 99% as segmentation rate. The purpose of the DTW algorithm is to perform the optimal time alignment between a reference pattern and an unknown pattern in order to ease the evaluation of their similarity. Unfortunately, the drawback of the DTW is its complex computing [6], [7], [8]. Consequently, several solutions and approaches have been proposed to speed up the DTW algorithm, [1], [7], [8], [4], [9], [5].

In this paper we show and confirm, through an experimental study, how volunteer computing, which present the advantage to be costless, can speed up, substantially also, the execution time of the Arabic OCR using the DTW algorithm. More specifically, we show how BOINC can achieve such a mission.

The reminder of this paper is organized as follows; section (2) describes the Arabic OCR using the DTW algorithm. Section (3), gives an overview on volunteer computing especially BOINC (Berkeley Open Infrastructure for Network Computing). The proposed approach and the corresponding performance evaluation are detailed in Section (4). Conclusion remarks and some future investigation are presented in section (5).

II. MECHANISM OF THE ARABIC OCR BASED ON THE DTW ALGORITHM

Words in Arabic are inherently written in blocks of connected characters. We need a prior segmentation of these blocks into separated characters. Indeed many researchers have considered the segmentation of Arabic words into isolated characters before performing the recognition phase. The viability of the use of DTW technique, however, is its ability and efficiency to perform the recognition without prior segmentation [4, 2].

We consider in this paper a reference library of R trained characters forming the Arabic alphabet in some given fonts, and denoted by C_r : r=1, 2,…, R. The technique consists to use the DTW pattern method to match an input character against the reference library. The input character is thus recognized as the reference character that provides the best time alignment, namely character A is recognized to be C_k if the summation distance S_k corresponding to the matching of A to reference character C_k satisfies the following equation [4, 5].

$$S_k = min_{1 \le r \le R}\{S_r\} \quad (1)$$

Let T constitutes a given connected sequence of Arabic characters to be recognized. T is then composed of a sequence of N feature vectors T_i that are actually representing the concatenation of some sub sequences of feature vectors representing each an unknown character to be recognized. As portrayed in Fig.1 text T lies on the time axis (the X-axis) in such a manner that feature vector T_i is at time i on this axis.

The reference library is portrayed on the Y-axis, where reference character C_r is of length l_r, $1 \leq r \leq R$. Let S (i, j, r) represent the cumulative distance at point (i, j) relative to reference character C_r . The objective here is to detect simultaneously and dynamically the number of characters composing T and recognizing these characters. There surely exists a number k and indices (m_1, m_2, ..., m_k) such that $Cm_1 \oplus Cm_2 \oplus \ldots \oplus Cm_k$ represent the optimal alignment to text T where $\oplus$ denotes the concatenation operation. The path warping from point (1, 1, m_1) to point (N, l_{mk},k) and representing the optimal alignment is therefore of minimum cumulative distance that is:

$$S(N, l_{mk}, k) = min_{1 \leq r \leq R}\{S(N, l_r, r\} \quad (2)$$

This path, however, is not continuous since it spans many different characters in the distance matrix. We therefore must allow at any time the transition from the end of one reference character to the beginning of another reference character. The end of reference character C_r is first reached whenever the warping function reaches point (i, l_r, r), i $= \left\lceil \frac{l_{r+1}}{2} \right\rceil$,...,N. As we can see from Fig.1, the end of reference characters C_1, C_2, C_3 are first reached at time 3, 4, 3 respectively. The end points of reference characters are shown in Fig.1 inside diamonds and points at which transitions occur are within a circle. The warping function always reaches the ends of the reference characters. At each time i, we allow the start of the warping function at the beginning of each reference character along with addition of the smallest cumulative distance of the end points found at time (i - 1) [4,5]. The resulting functional equations are:

$$S(i,j,r) = D(i,j,r) + min_{\substack{1 \leq i \leq N \\ 1 \leq j \leq l_r \\ 1 \leq r \leq R}} \begin{Bmatrix} S(i-1,j,r), \\ S(i-1,j-1,r), \\ S(i-1,j-2,r) \end{Bmatrix} \quad (3)$$

With the boundary conditions:

$$S(i,1,r) = D(i,1,r) + min_{\substack{1+\lceil \frac{lr+1}{2} \rceil \leq i \leq N \\ l \leq j \leq l_r \\ l \leq r \leq R}} S(i-1, l_K, k) \quad (4)$$

To trace back the warping function and the optimal alignment path, we have to memorize the transition times among reference characters. This can easily be accomplished by the following procedure:

$$b(i,j,r) = trace\ min_{\substack{1 \leq i \leq N \\ 1 \leq j \leq l_r \\ 1 \leq r \leq R}} \begin{Bmatrix} b(i-1,j,r), \\ b(i-1,j-1,r), \\ b(i-1,j-2,r) \end{Bmatrix} \quad (5)$$

Where trace min is a function that returns the element corresponding to the term that minimizes the functional equations. The functioning of this algorithm is portrayed on Fig.1 by means of the two vectors VecA and VecB, where VecB(i) represents the reference character giving the least cumulative distance at time i, and VecA(i) provides the link to the start of this reference character in the text T . The heavy marked path through the distance matrix represents the optimal alignment of text T to the reference library. We observe that the text is recognized as C1 $\oplus$ C3 [5].

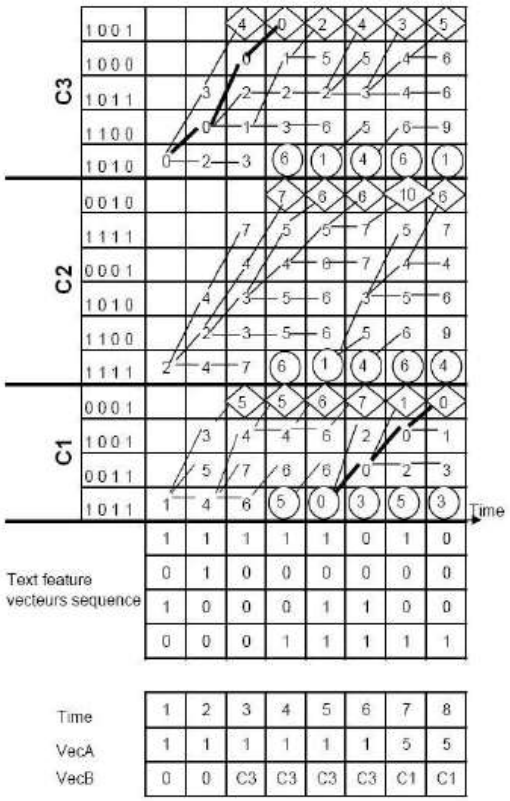


Figure 1. DTW Wrapping mechanism.

III. VOLUNTEER COMPUTING AND BOINC

Volunteer computing is a form of distributed computing in which the general public volunteer make available and storage resources to scientific research projects. Early volunteer computing projects include the Great Internet Mersenne Prime Search [12], SETI@home [10], Distributed.net [11] and Folding@home [13]. Today the approach is being used in many areas, including high energy physics, molecular biology, medicine, astrophysics, and climate dynamics. This type of computing can provide great power (SETI@home, for example, has accumulated 2.5 million years of CPU time in 7 years of operation). However, it requires attracting and retaining volunteers, which places many demands both on projects and on the underlying technology.

The Berkeley Open Infrastructure for Network Computing (BOINC) is a framework to deploy distributed computing platforms based on volunteer computing. It is developed at U.C. Berkeley Spaces Sciences Laboratory and it is released under an open source license.

BOINC is the evolution of the original SETI@home project, which started in 1999 and attracted millions of participants worldwide [16].

This middleware provides a generic framework for implementing distributed computation applications within a heterogeneous environment. The system is designed as a

software platform utilizing computing resources from volunteer computers [14].

BOINC software is divided in two main components: the server and the client side of the software. BOINC allows sharing computing resources among different autonomous projects. A BOINC project is identified by a unique URL which is used by BOINC clients to register with it. Every BOINC project must run a host with the server side of the BOINC software.

BOINC software on the server side comprises several components: one or more scheduling servers that communicate with BOINC clients, one or more data servers that distribute input files and collect output files, a web interface for participants and project administrators and a relational database that stores information about work, results, and participants.

BOINC software provides powerful tools to manage the applications run by a project. For instance it allows to easily define different application versions for different target architectures. A "workunit" describes a computation to be performed, associating a (unique) name with an application and the corresponding input files.

Not all kind of applications are suitable to be deployed on BOINC. Ideally, candidate applications must present "independent parallelism" (divisible into parallel parts with few or no data dependencies) and a low data/compute ratio since output files will be sent through a typically slow commercial Internet connection [16].

In this paper, we show how volunteer computing such as BOINC can speed up the execution time of Arabic OCR based on the DTW algorithm.

IV. THE DTW DATA DISTRIBUTION OVER BOINC

The Arabic OCR based on the DTW procedure described in the preceding section presents many ways on which one could base its parallelization or distribution. The idea of the proposed approach is how to take advantages of the enough power provided by BOINC to speed up the execution time of the DTW algorithm?

A BOINC project uses a set of servers to create, distribute, record, and aggregate the results of a set of tasks that the project needs to perform to accomplish its goal. The tasks are evaluating data sets, called workunits. The servers distribute the tasks and corresponding workunits to clients (software that runs on computers that people permit to participate in the project). When a computer running a client would otherwise be idle (in the context of volunteer computing, a computer is deemed to be idle if the computer's screensaver is running), it spends the time working on the tasks that a server assigns to the client. When the client has finished a task, it returns the result obtained by completing the task to the server. If the user of a computer that is running a client begins to use the computer again, the client is interrupted and the task it is processing is paused while the computer executes programs for the user. When the computer becomes idle again, the client continues processing the task it was working on when the client was interrupted.

To be added into a BOINC project, applications must incorporate some interaction with the BOINC client: they must notify the client about start and finish, and they must allow for renaming of any associated data files, so that the client can relocate them in the appropriate part of the guest operating system and avoid conflicts with workunits from other projects [16].

We propose to split optimally the binary image of a given Arabic text to be recognized into a set of binary sub images and then assign them among some volunteer computers which are already subscribed to our project.

BOINC uses a simple but a rich set of abstraction files, applications, and data. A project defines application versions for various platforms (Windows, Linux/x86, Mac OS/X, etc.).

An application can consist of an arbitrary set of files. A workunit represents the inputs to a computation: the application (but not a particular version) presents a set of references input files, and sets of command line arguments and environment variables. Each workunit has parameters such as computing, memory and storage requirements and a soft deadline for completion. A result represents the result of a computation, it consists of a reference to a workunit and a list of references to output files.

Files can be replicated, the description of a file includes a list of URLs from which it may be downloaded or uploaded. When the BOINC client communicates with a scheduling server it reports completed work, and receives an XML document describing a collection of the above entities. The client then downloads and uploads files and runs applications; it maximizes concurrency, using multiple CPUs when possible and overlapping communication and computation.

A. Experimental Study

Significant reduction in the elapsed time, defined as the time elapsing from the start until the completion of the text recognition, can be realized by using a distributed architecture. This effect is known as the speedup factor. This factor is properly defined as the ratio of the elapsed time using sequential mode with just one processor to the elapsed time using the distributed architecture.

Next we consider only the case where the volunteer computers participating in the work are homogeneous. It means that all the corresponding interconnected computers are homogeneous in terms of computing power, hardware configuration and operating system. We ran several experiments on several specific printed Arabic texts.

Our experiments aim at proving that volunteer computing present, indeed, interesting infrastructures to speed up the execution process of the Arabic OCR based on the DTW algorithm.

During these experiments, we have considered the following conditions:

- The number of pages is 100,
- The number of lines per page is 7,
- The average number of characters per line is 55,

- The number of characters per page is 369,
- The reference library contains 103 characters,
- We have used 16 dedicated homogeneous workers having the exact configuration: 3GHZ CPU frequency, 512 Mega Octets RAM and running Windows XP-professional.

To reach our expectation, we have studied the effect of the distribution of approximately a hundred (100) of similar printed Arabic text pages over a variable number of volunteer computers.

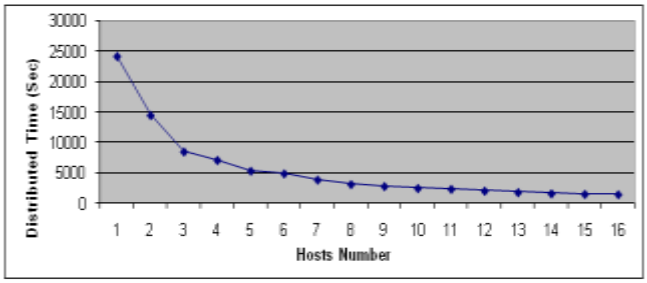


Figure 2. Distributed execution time.

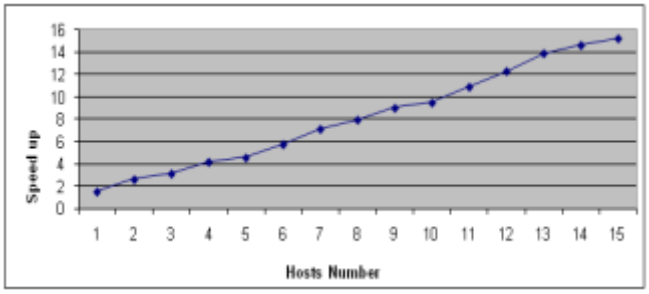


Figure 3. Speedup of the DTW algorithm.

Fig. 2 and 3 illustrate the obtained results of our experiment. These figures show in particular that:

- The execution time of the DTW algorithm decreases with the number of computers used. Each time you add a computer the execution time of recognition decrease. The average test time for 1 computer was approximately 6.20 hours and the average test time for 16 computers was 0.4 hours. It clearly shows an exponential decrease in the amount of time required to complete the tests.
- However, the speedup factor increases with the number of computers used.
- If we use 16 computers then the execution time reaches the value 1450 seconds and the speedup factor reaches the value 15. This result is very interesting, because in this case our proposed OCR system is able to recognize more than 830 characters per second, compared with the existing commercial Arabic OCR [2].

Consequently, volunteer computing constitute interesting infrastructures to speedup, drastically, the execution time of the Arabic OCR based on the DTW algorithm. Moreover, and thanks to the enough computing power provided by such infrastructures, we can think, now, about the improvement of the recognition rate of our system by adding to it some complementary approaches or techniques.

V. Conclusion and Perspective

This paper has shown how volunteer computing present interesting infrastructures to speed up, substantially, the execution time of the Arabic OCR based on the DTW algorithm. Indeed, conducted experiments confirm that such infrastructures can help a lot in building a powerful Arabic OCR based on the combination (integration) of some strong complementary approaches or techniques which require enough computing power.

Several investigations are under studies especially the way to exploit in a large scale the BOINC and the way to improve the recognition rate of the Arabic OCR based on the DTW algorithm in order to build a powerful Arabic OCR system.

References

[1] M. Khemakhem, A. Belghith, M. Labidi « The DTW data distribution over a grid computing architecture », International Journal of Computer Sciences and Engineering Systems (IJCSES), Vol.1, N°.4, p. 241-247, December 2007

[2] M. Khemakhem, A. Belghith: « Towards a distributed Arabic OCR based on the DTW algorithm », the International Arab Journal of Information Technology (IAJIT), Vol. 6, N°. 2, p. 153-161, April 2009.

[3] M. Khemakhem et al., Arabic Type Writen Character Recognition Using Dynamic Comparison, Proc. 1st Computer conference Kuwait, March. 1989.

[4] M. Khemakhem, A. Belghith, M. Ben Ahmed, Modélisation architecturale de la Comparaison Dynamique distribuée Proc, Second International Congress On Arabic and Advanced Computer Technology, Casablanca, Morocco, December 1993.

[5] M. Khemakhem and A. Belghith., A Multipurpose Multi-Agent System based on a loosely coupled Architecture to speed up the DTW algorithm for Arabic printed cursive OCR. Proc. IEEE AICCSA-2005, Cairo, Egypt, January 2005.

[6] A.Cheung et al. An Arabic optical character recognition system using recognition based segmentation, Pattern Recongnition. Vol. 34, 2001.

[7] G. R. Quénot et al., A Dynamic Programming Processor for Speech Recognition, IEEE JSSC, Vol. 24, N(F 9)q.20, April 1989.

[8] Philip G. Bradford, Efficient Parallel Dynamic Programming, Proc. the 30 Annual Allerton Conference on Communication, Control an Computing, University of Illinois, 185-194, 1992.

[9] C. E. R. Alves et al, Parallel Dynamic Programming for solving the String Editing Problem on CGM/BSP, Proc. SPAA'02, August 10-13, 2002 Winnipeg, Manitoba, Canada.

[10] D.P. Anderson, J. Cobb, E. Korpela, M. Lebofsky, D. Werthimer. "SETI@home: An Experiment in Public-Resource Computing". Communications of the ACM, November 2002

[11] Distributed.net, http://distributed.net

[12] GIMPS, http://www.mersenne.org/prime.htm

[13] S.M. Larson, C.D. Snow, M. Shirts and V.S. Pande. "Folding@Home and Genome@Home: Using distributed computing to tackle previously intractible problems in computational biology". Computational Genomics, Horizon Press, 2002.

[14] D.P. Anderson. "BOINC: A System for Public-Resource Computing and Storage". 5th IEEE/ACM International Workshop on Grid Computing, Nov. 8 2004.

[15] Einstein@Home, http://einstein.phys.uwm.edu/

[16] B.Antoli, F. Castejón, A.Giner, G.Losilla, J.M Renolds, A.Rivero, S.sangiaos, F.Serrano, A. Tarancón, R. Vallés and J.L. Velasco "ZIVIS: A City Computing Platform Based on Volunteer Computing"

[17] D.P. Anderson, E. Korpela, and R. Walton. "High-Performance Task Distribution for Volunteer Computing". First IEEE International Conference

[18] Gupta, S., Doshi, V., Jain, A., & Iyer, S. (2010). Iris Recognition System using Biometric Template Matching Technology. International Journal of Advanced Computer Science and Applications - IJACSA, 1(2), 24-28. doi: 10.5120/61-161.

[19] Padmavathi, G. (2010). A suitable segmentation methodology based on pixel similarities for landmine detection in IR images. International Journal of Advanced Computer Science and Applications - IJACSA, 1(5), 88-92.

[20] Miriam, D. D. H. (2011). An Efficient Resource Discovery Methodology for HPGRID Systems. International Journal of Advanced Computer Science and Applications - IJACSA, 2(1).

[21] Hamdy, S., El-messiry, H., Roushdy, M., & Kahlifa, E. (2010). Quantization Table Estimation in JPEG Images. *International Journal of Advanced Computer Science and Applications - IJACSA, 1*(6), 17-23.

AUTHORS PROFILE

Maher Khemakhem received his master of science and his PhD degrees from the University of Paris 11, France in 1984 and 1987, respectively. He is currently assistant professor in computer science at the Higher institute of Management at the University of Sousse Tunisia. His research interests include distributed systems, performance evaluation, and pattern recognition.

Zied Trifa received his master degree of computer science from the University of Economics and management of Sfax, Tunisia in 2010. He is currently PhD student in computer science at the same University. His research interests include Grid Computing, distributed systems, and performance evaluation.

Mohamed Laabidi received his master degree of computer science from the University of Economics and management of Sfax, Tunisia in 2007. He is currently PhD student in computer science at the same University. His research interests include Cloud and Grid Computing, distributed systems, and performance evaluation.

An Empirical Study of the Applications of Data Mining Techniques in Higher Education

Dr. Varun Kumar
Department of Computer Science and Engineering
ITM University
Gurgaon, India
kumarvarun333@gmail.com

Anupama Chadha
Department of Computer Science and Engineering
ITM University
Gurgaon, India
anupamaluthra@gmail.com

***Abstract*— Few years ago, the information flow in education field was relatively simple and the application of technology was limited. However, as we progress into a more integrated world where technology has become an integral part of the business processes, the process of transfer of information has become more complicated. Today, one of the biggest challenges that educational institutions face is the explosive growth of educational data and to use this data to improve the quality of managerial decisions. Data mining techniques are analytical tools that can be used to extract meaningful knowledge from large data sets. This paper addresses the applications of data mining in educational institution to extract useful information from the huge data sets and providing analytical tool to view and use this information for decision making processes by taking real life examples.**

Keywords- Higher education,; Data mining; Knowledge discovery; Classification; Association rules; Prediction; Outlier analysis;

I. INTRODUCTION

In modern world a huge amount of data is available which can be used effectively to produce vital information. The information achieved can be used in the field of Medical science, Education, Business, Agriculture and so on. As huge amount of data is being collected and stored in the databases, traditional statistical techniques and database management tools are no longer adequate for analyzing this huge amount of data.

Data Mining (sometimes called data or knowledge discovery) has become the area of growing significance because it helps in analyzing data from different perspectives and summarizing it into useful information. [1]

There are increasing research interests in using data mining in education. This new emerging field, called Educational Data Mining, concerns with developing methods that discover knowledge from data originating from educational environments [1].

The data can be collected from various educational institutes that reside in their databases. The data can be personal or academic which can be used to understand students' behavior, to assist instructors, to improve teaching, to evaluate and improve e-learning systems , to improve curriculums and many other benefits.[1][2]

Educational data mining uses many techniques such as decision trees, neural networks, k-nearest neighbor, naive bayes, support vector machines and many others.[3]

Using these techniques many kinds of knowledge can be discovered such as association rules, classifications and clustering. The discovered knowledge can be used for organization of syllabus, prediction regarding enrolment of students in a particular programme, alienation of traditional classroom teaching model, detection of unfair means used in online examination, detection of abnormal values in the result sheets of the students and so on.

This paper is organized as follows: Section II describes the related work. Section III describes the research question. Section IV describes data mining techniques adopted. Section V discusses the application areas of these techniques in an educational institute. Section VI concludes the paper.

II. RELATED WORK

Data mining in higher education is a recent research field and this area of research is gaining popularity because of its potentials to educational institutes.

[1] gave case study of using educational data mining in Moodle course management system. They have described how different data mining techniques can be used in order to improve the course and the students' learning. All these techniques can be applied separately in a same system or together in a hybrid system.

[2] have a survey on educational data mining between1995 and 2005. They have compared the Traditional Classroom teaching with the Web based Educational System. Also they have discussed the use of Web Mining techniques in Education systems.

[3] have a described the use of k-means clustering algorithm to predict student's learning activities. The information generated after the implementation of data mining technique may be helpful for instructor as well as for students.

[4] discuss how data mining can help to improve an education system by enabling better understanding of the students. The extra information can help the teachers to manage their classes better and to provide proactive feedback to the students.

[6] have described the use of data mining techniques to predict the strongly related subject in a course curricula. This information can further be used to improve the syllabi of any course in any educational institute.

[8] describes how data mining techniques can be used to determine The student learning result evaluation system is an essential tool and approach for monitoring and controlling the learning quality. From the perspective of data analysis, this paper conducts a research on student learning result based on data mining.

III. RESEARCH OBJECT

The object of the present study is to identify the potential areas in which data mining techniques can be applied in the field of Higher education and to identify which data mining technique is suited for what kind of application.

IV. DATA MINING DEFINITION AND TECHNIQUES

Simply stated, data mining refers to extracting or "mining" knowledge from large amounts of data. [5] Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships helpful in decision making. The sequences of steps identified in extracting knowledge from data are: shown in Figure 1.

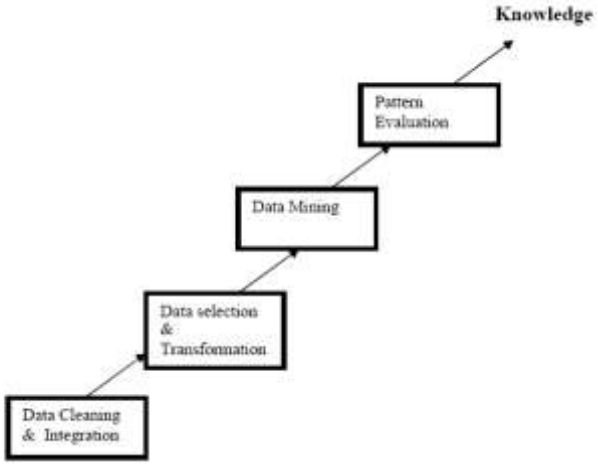


Figure 1. The steps of extracting knowledge from data

The various techniques used in Data Mining are:

A. Association analysis

Association analysis is the discovery of association rules showing attribute-value conditions that occur frequently together in a given set of data. Association analysis is widely used for market basket or transaction data analysis.

More formally, association rules are of the form $X \Rightarrow Y, i.e., "A_1 \wedge ---- \wedge A_m \rightarrow B_1 \wedge ---- \wedge B_n"$, where Ai (for i to m) and Bj (j to n) are attribute-value pairs.

(I)

The association rule X=>Y is interpreted as database tuples that satisfy the conditions in X are also likely to satisfy the conditions in Y ".

B. Classification and Prediction

Classification is the processing of finding a set of models (or functions) which describe and distinguish data classes or concepts, for the purposes of being able to use the model to predict the class of objects whose class label is unknown. The derived model may be represented in various forms, such as classification (IF-THEN) rules, decision trees, mathematical formulae, or neural networks. Classification can be used for predicting the class label of data objects. However, in many applications, one may like to predict some missing or unavailable data values rather than class labels. This is usually the case when the predicted values are numerical data, and is often specifically referred to as prediction.

IF-THEN rules are specified as **IF condition THEN conclusion**

e.g. IF age=youth and student=yes then buys_computer=yes

C. Clustering Analysis

Unlike classification and predication, which analyze class-labeled data objects, clustering analyzes data objects without consulting a known class label. In general, the class labels are not present in the training data simply because they are not known to begin with. Clustering can be used to generate such labels. The objects are clustered or grouped based on the principle of maximizing the intraclass similarity and minimizing the interclass similarity.

That is, clusters of objects are formed so that objects within a cluster have high similarity in comparison to one another, but are very dissimilar to objects in other clusters. Each cluster that is formed can be viewed as a class of objects, from which rules can be derived. [5]

Application of clustering in education can help institutes group individual student into classes of similar behavior. Partition the students into clusters, so that students within a cluster (e.g. Average) are similar to each other while dissimilar to students in other clusters (e.g. Intelligent, Weak).

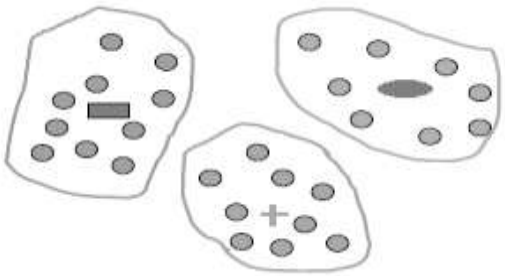


Figure 2. Picture showing the partition of students in clusters

D. Outlier Analysis

A database may contain data objects that do not comply with the general behavior of the data and are called outliers. The analysis of these outliers may help in fraud detection and predicting abnormal values.

V. POTENTIAL APPLICATIONS

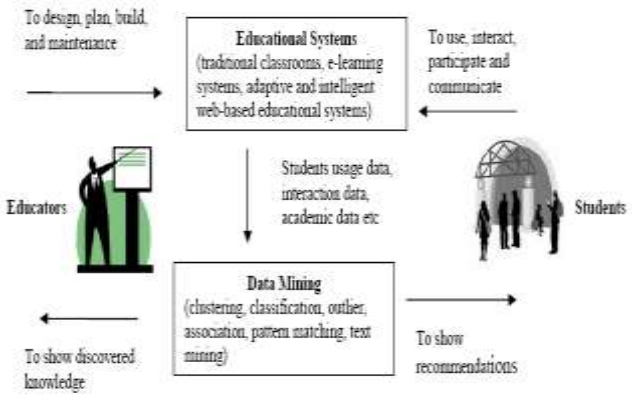


Figure 3. The cycle of applying data mining in education system [4]

The above figure illustrates how the data from the traditional classrooms and web based educational systems can be used to extract knowledge by applying data mining techniques which further helps the educators and students to make decisions.

A. Organization of Syllabus

It is important for educational institutes to maintain a high quality educational programme which will improve the student's learning process and will help the institute to optimize the use of resources. A typical student at the university level completes a number of courses (i.e. "course" and "subject" are used synonymously) prior to graduation.

Presently, organization of syllabi is influenced by many factors such as affiliated, competing or collaborating programmes of universities, availability of lecturers, expert judgments and experience. This method of organization may not necessarily facilitate students' learning capacity optimally. Exploration of subjects and their relationships can directly assist in better organization of syllabi and provide insights to existing curricula of educational programmes.

One of the application of data mining is to identify related subjects in syllabi of educational programmes in a large educational institute.[6]

A case study has been performed where the student data collected over a period of time at the Sri Lanka Institute of Information Technology (SLIIT) [7]. The main aim of the study was to find the strongly related subjects in a course offered by the institute. For this purpose following methodology was followed to:

- Identify the possible related subjects.
- Determine the strength of their relationships and determine strongly related subjects.

METHODOLOGY

In the first step, association rule mining is used to identify possibly related two subject combinations in the syllabi which also reduce our search space. In the second step(see[6]), Pearson Correlation Coefficient was applied to determine the strength of the relationships of subject combinations identified in the first step.

TABLE I. THE SUBJECTS CHOSEN BY STUDENTS

Student id	Subject 1	Subject 2	Subject 3
1	Databases	Advanced Databases	Data mining
2	Databases	Advanced Databases	Data mining
3	Databases	Advanced Databases	Data mining
4	Databases	Advanced Databases	Visual Basic
5	Databases	Advanced Databases	Web Designing

Association Rules that can be derived from Table 1 are of the form:

$$(X, subject1) \Rightarrow (X, subject2) \quad (2)$$

$$(X, subject1) \wedge (X, subject2) \Rightarrow (X, subject3) \quad (3)$$

$$(X, "Databases") \Rightarrow (X, "AdvancedDatabases")$$

[support=2% and confidence=60%] (4)

$$(X, "Databases") \wedge (X, "AdvancedDatabases" \Rightarrow (X, "DataMining")$$

[support=1% and confidence=50%] (5)

Where support factor of the association rule shows that 1% of the students have taken both the subjects "Databases" and "Advanced Databases" and confidence factor shows that there is a chance that 50% of the students who have taken "Databases" will also take "Advanced Databases"

This way we can find the strongly related subjects and can optimize the syllabi of an educational programme.

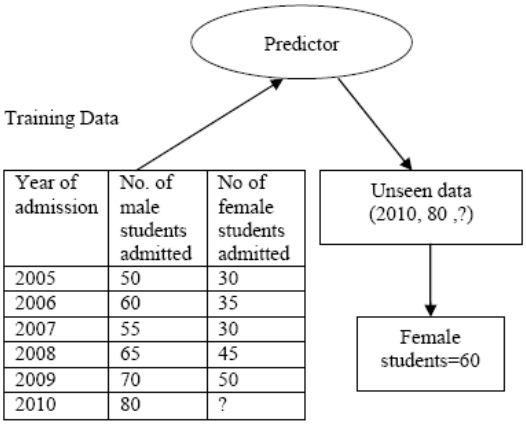
B. Predicting The Registration Of Students in an Educational Programme

Now a days educational organization are getting strong competition from other Academic competitors. To have an edge over other organizations, needs deep and enough knowledge for a better assessment, evaluation, planning, and decision making.

Data Mining helps organizations to identify the hidden patterns in databases; the extracted patterns are then used to build data mining models, and hence can be used to predict performance and behavior with high accuracy. As a result of this, universities are able to allocate resources more effectively.

METHODOLOGY

One of the application of data mining can be for example, to efficiently assign resources with an accurate estimate of how many male or female will register in a particular program by using the Prediction techniques.



Year of admission	No. of male students admitted	No of female students admitted
2005	50	30
2006	60	35
2007	55	30
2008	65	45
2009	70	50
2010	80	?

Figure 4. Prediction of female students in the coming year

In real scenario couple of other associated attributes like type of course, transport facility, hostel facility etc can be used to predict the registration of students.

C. Predicting Student Performance

One of the question whose answer almost every stakeholder of an educational system would like to know "Can we predict student performance?"

Over the years, many researchers applied various data mining techniques to answer this question.

In modern times, learning is taking on a more important role in the development of our civilization. Learning is an individual behavior as well as a social phenomenon.

It is a difficult task to deeply investigate and successfully develop models for evaluating learning efforts with the combination of theory and practice. University goals and outcomes clearly relate to "promoting learning through effective undergraduate and graduate teaching, scholarship, and research in service to University."

Student learning is addressed in some goals and outcomes related to the development of overall student knowledge, skill, and dispositions. Collections of randomly selected student work are examined and assessed by small groups of faculty teaching courses within some general education categories. [8]

With the help of data mining techniques a result evaluation system can be developed which can help teachers and students to know the weak points of the traditional classroom teaching model. Also it will help them to face the rapidly developing real-life environment and adapt the current teaching realities.

METHODOLOGY

We can use student participation data as part of the class grading policy. An instructor can assess the quality of student by conducting an online discussion among a group of students and use the possible indicators such as the time difference between posts, frequency distribution of the postings, duration between postings and replies etc.

Given this data, we can apply classification algorithms to classify the students into possible levels of quality.

Table II. Student participation data and their grades

Time difference between posts (in min)	Duration between postings and replies (in min)	Grade of the student
3	1	A
3	2	B
4	1	A
5	2	B
6	1	C
6	2	C

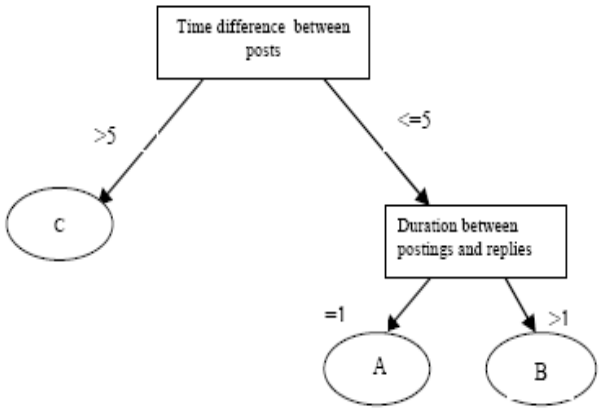


Figure 5. The Decision Tree built from the data in Table II

D. Detecting Cheating in Online Examination

We can say that online assessments are useful to evaluate students' knowledge; they are used around the world in schools -since elementary to higher education institutions- and in recognized training centers like the Cisco Academy [9].

Now a days exams are conducted online remotely through the Internet and if a fraud occurs then one of the basic problems to solve is to know: who is there? Cheating is not only done by students but the recent scandals in business and journalism show that it has become a common practice.

Data Mining techniques can propose models which can help organizations to detect and to prevent cheats in online assessments. The models generated use data comprising of different student's personalities, stress situations generated by online assessments, and common practices used by students to cheat to obtain a better grade on these exams.

E. Identifying Abnormal/ Erroneous Values

The data stored in a database may reflect outliers-|noise, exceptional cases, or incomplete data objects. These objects may confuse the analysis process, causing over fitting of the data to the knowledge model constructed. As a result, the accuracy of the discovered patterns can be poor. [5]

One of the applications of Outlier Analysis can be to detect the abnormal values in the result sheet of the students. This may be due many factors like a software fault, data entry operator negligence or an extraordinary performance of the student in a particular subject.

Table III. The result of students in four subjects

Student roll no	Marks in subject1	Marks in subject2	Marks in subject3	Marks in subject4
101	30	35	45	30
102	67	75	78	67
103	89	90	78	77
104	30	35	45	**99**

In the table shown above the result of the student in subject4 with roll no 104 will be detected as an exceptional case and can be further analyzed for the cause.

VI. CONCLUSION

In the present study, we have discussed the various data mining techniques which can support education system via generating strategic information.. Since the application of data mining brings a lot of advantages in higher learning institution, it is recommended to apply these techniques in the areas like optimization of resources, prediction of retainment of faculties in the university, to find the gap between the number of candidates applied for the post, number of applicants responded, number of applicants appeared, selected and finally joined. Hopefully these areas of application will be discussed in our next paper.

REFERENCES

[1] C. Romero, S. Ventura, E. Garcia, "Datamining in course management systems: Moodle case study and tutorial", Computers & Education, Vol. 51, No. 1, pp. 368-384, 2008

[2] C. Romero, S. Ventura "Educational data Mining: A Survey from 1995 to 2005", Expert Systems with Applications (33), pp. 135-146, 2007

[3] Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza Sattar, M. Inayat Khan, **"**Data Mining Model for Higher Education System", Europen Journal of Scientific Research, Vol.43, No.1, pp.24-29, 2010

[4] K. H. Rashan, Anushka Peiris, "Data Mining Applications in the Education Sector", MSIT, Carnegie Mellon University, retrieved on 28/01/2011

[5] Han Jiawei, Micheline Kamber, *Data Mining: Concepts and Technique.* Morgan Kaufmann Publishers,2000

[6] W.M.R. Tissera, R.I. Athauda, H. C. Fernando "Discovery of Strongly Related Subjects in the Undergraduate Syllabi using Data Mining", IEEE International Conference on Information Acquisition, 2006

[7] Sri Lanka Institute of Information Technology, *http://www.sliit.lk/,* retrieved on 28/02/2011

[8] Sun Hongjie, "Research on Student Learning Result System based on Data Mining", IJCSNS International Journal of Computer Science and Network Security, Vol.10, No. 4, April 2010

[9] Academy Connection – Training Resources In html, http*://www.cisco.com/web/learning/netacad/index,* December 28th, 2005.

[10] Wayne Smith, "Applying Data Mining to Scheduling Courses at a University", Communications of the Association for Information Systems, Vol. 16, Article 23, 2005

[11] Firdhous, M. F. M. (2010). Automating Legal Research through Data Mining. International Journal of Advanced Computer Science and Applications - IJACSA, 1(6), 9-16.

[12] The Result Oriented Process for Students Based On Distributed Data Mining. (2010). International Journal of Advanced Computer Science and Applications - IJACSA, 1(5), 22-25.

[13] Jadhav, R. J. (2011). Churn Prediction in Telecommunication Using Data Mining Technology. International Journal of Advanced Computer Science and Applications - IJACSA, 2(2), 17-19.

AUTHORS PROFILE

Presently a Research student at ITM Gurgaon , she is armed with a degree in Electronics followed by a Master of Technology and has had an accomplished academic record all through her education and career. After having a stint in software development she has taught for over 11 years to PG and UG students in Computer Applications and Information Technology. She also has to her credit a number of systems improvement projects for her previous employers besides having administrative experience relating to functioning of placement cell and coordination of various academic programmes. A distinguished faculty, her field of specialization is Software Engineering, Data Mining and Object Oriented Analysis and Design concepts.

Integrated Routing Protocol for Opportunistic Networks

Anshul Verma
Computer Science and Engineering Dept.
ABV-Indian Institute of Information Technology and Management, Gwalior, India
E-mail: anshulverma87@gmail.com

Dr. Anurag Srivastava
Applied Science Dept.
ABV-Indian Institute of Information Technology and Management, Gwalior, India
E-mail: anurags@iiitm.ac.in

***Abstract*—In opportunistic networks the existence of a simultaneous path is not assumed to transmit a message between a sender and a receiver. Information about the context in which the users communicate is a key piece of knowledge to design efficient routing protocols in opportunistic networks. But this kind of information is not always available. When users are very isolated, context information cannot be distributed, and cannot be used for taking efficient routing decisions. In such cases, context oblivious based schemes are only way to enable communication between users. As soon as users become more social, context data spreads in the network, and context based routing becomes an efficient solution. In this paper we design an integrated routing protocol that is able to use context data as soon as it becomes available and falls back to dissemination-based routing when context information is not available. Then, we provide a comparison between Epidemic and PROPHET, these are representative of context oblivious and context aware routing protocols. Our results show that integrated routing protocol is able to provide better result in term of message delivery probability and message delay in both cases when context information about users is available or not.**

Keywords-context aware routing; context information; context oblivious routing; MANET; opportunistic network.

I. INTRODUCTION

The opportunistic network is an extension of Mobile Ad hoc Network (MANET). Wireless networks' properties, such as disconnection of nodes, network partitions, mobility of users and links' instability, are seen as exceptions in traditional network. This makes the design of MANET significantly more difficult [1].

Opportunistic networks [2] are created out of mobile devices carried by people, without relying on any preexisting network topology. Opportunistic networks consider disconnections, mobility, partitions, etc. as norms instead of the exceptions. In opportunistic network mobility is used as a technique to provide communication between disconnected 'groups' of nodes, rather than a drawback to be solved.

In opportunistic networking a complete path between two nodes wishing to communicate is unavailable [3]. Opportunistic networking tries to solve this problem by removing the assumption of physical end-to-end connectivity and allows such nodes to exchange messages. By using the store-carry-and-forward paradigm [4] intermediate nodes store messages when there is no forwarding opportunity towards the destination, and exploit any future contact opportunity with other mobile devices to bring the messages closer and closer to the destination.

Therefore routing is one of the most compelling challenges. The design of efficient routing protocols for opportunistic networks is generally a difficult task due to the absence of knowledge about the network topology. Routing performance depends on knowledge about the expected topology of the network [5]. Unfortunately, this kind of information is not always available. Context information is a key piece of knowledge to design efficient routing protocols. Context information represents users' working address and institution, the probability of meeting with other users or visiting particular places. It represents the current working environment and behavior of users. It is very help full to identify suitable forwarders based on context information about the destination. We can classify the main routing approaches proposed in the literature based on the amount of context information of users they exploit. Specifically, we identify two classes, corresponding to context-oblivious and context-aware protocols.

Protocols in Context-oblivious routing class as Epidemic Routing Protocol [6] are only solution when context information about users is not available. But they generate high overhead, network congestion and may suffer high contention. Context-based routing provides an effective congestion control mechanism and with respect to context-oblivious routing, provides acceptable QoS with lower overhead. Indeed, PRoPHET [7] is able to automatically learn the past communication opportunities determined by user's movement patterns and exploit them efficiently in future. This autonomic, self-learning feature is completely absent in Context-oblivious routing schemes. But context based routing protocols provide high overhead, message delay and less success full message in absence of context information about users. We have proved this by implementing epidemic and PROPHET routing protocols in presence and absence of context information. We found epidemic is better in absence of context information while PROPHET gives better result in presence of context information. Therefore I decided to combine feature of these both protocols into a single integrated routing protocol, which will perform better in both cases when context information about user is available or not.

This paper represents our integrated routing protocol, and evaluates it through simulations. The rest of the paper is organized as follows. Section 2 describes main routing protocols of context oblivious and context based routing class which are epidemic and PROPHET, and describe some related work. In section 3 our proposed scheme is presented. In section 4 the simulation setup is given for routing protocols. Comparison and Result of epidemic, PROPHET and integrated routing protocols can be found in section 5. Finally section 6 discusses conclusion and looks into future work.

II. RELATED WORK

A. Epidemic Routing

Epidemic Routing provides the final delivery of messages to random destinations with minimal assumptions of topology and connectivity of the network. Final message delivery depends on only periodic pair-wise connectivity between mobile devices. The Epidemic Routing protocol works on the theory of epidemic algorithms [6]. Each host maintains two buffers, one for storing messages that it has originated and second for messages that it is buffering on behalf of other hosts. Each mobile device stores a summary vector that contains a compact representation of messages currently stored in buffer.

When two hosts come into communication range of one another, they exchange their summary vectors. Each host also maintains a buffer to keep list of recently seen hosts to avoid redundant connections. Summary vector is not exchanged with mobile devices that have been seen within a predefine time period. After exchanging their summary vectors mobile devices compare summary vectors to determine which messages is missing. Then each mobile device requests copies of messages that it does not contain.

Each message has a unique identification number and a hop count. The message identifier is a unique 32-bit number. The hop count field determines the maximum number of intermediate nodes that a particular message can travel. When hop count is one, messages can only be delivered to their final destinations. Larger values for hop count will increase message delivery probability and reduce average delivery time, but will also increase total resource consumption in message delivery [6].

B. PROPHET routing

PROPHET [7], a Probabilistic ROuting Protocol using History of Encounters and Transitivity makes use of observations that real users mostly move in a predictable fashion. If a user has visited a location several times before, there is more probability to visit that location again. PROPHET uses this information to improve routing performance.

To accomplish this, PROPHET maintains delivery predictability metric at every node. This metric represents message delivery probability of a host to a destination. PROPHET is similar to Epidemic Routing but it introduces a new concept of delivery predictability. Delivery predictably is the probability for a node to encounter a certain destination. When two nodes meet, they also exchange delivery predictability information with summary vectors. This information is used to update the delivery predictability information of metric.

When a message comes at a node, node checks that destination is available or not. If destination is not available, node stores the message and upon each encounters with another device, it takes decision whether or not to transfer a message. Message is transferred to the other node if the other node has higher message delivery probability to the destination [7].

C. Other work

Since routing is one of the most challenging issues in opportunistic networks, many researchers are working in this area. In this Section we are only mentioning some specific routings, which are representative of both context oblivious and context based routing protocols in opportunistic networks. The reader can also find a brief discussion on routing protocols for opportunistic networks in [8].

Context oblivious based algorithms also include network-coding-based routing [9]. In general, network coding-based routing reduces flooding, as it is able to deliver the same amount of information with fewer messages injected into the network [10].

Spray and Wait [11] routing provides a drastic way to reducing the overhead of Epidemic. Message is delivered in two steps: the spray phase and the wait phase. During the spray phase, source node and first receivers of the message spread multiple copies of the same message over the network. Then, in the wait phase each relay node stores its copy and eventually delivers it to the destination when it comes within reach.

Frequency of meetings between nodes and frequency of visits to specific physical places is used by MV [12] and MaxProp [13] as context information.

MobySpace routing [14] uses the mobility pattern of nodes as context information. The protocol uses a multi dimensional Euclidean space, named MobySpace, where possible contact between couples of nodes are represented by each axis and the probability of that contacts to occur are measured by the distance along axis. Two nodes that are close in the MobySpace, have similar sets of contacts. The best forwarding node for a message is the node that is as close as possible to the destination node in this space.

In Bubble Rap [15], social community users belong to is used as context information. Basically, Bubble Rap prefers nodes belonging to the same community of the destination as a good forwarder to this destination. If such nodes are not found, it forwards the message to the nodes, which have more chances of contact with the community of the destination. In Bubble Rap, communities are automatically detected via the patterns of contacts between nodes.

Other opportunistic routing protocols use the time lag from the last meeting with a destination as context information. Last Encounter routing [16] and Spray and Focus [17] routings are example of protocols exploiting such type of information.

Context-aware routing [18] uses an existing MANET routing protocol to connect nodes of the same MANET cloud. To transmit messages outside the cloud, a sender gives

message to the node in its current cloud that has highest message delivery probability to the destination. This node waits to get in touch with destination or enters in destination's cloud with other nodes that has higher probability of meeting the destination. In context-aware routing context information is used to calculate probabilities only for those destinations each node is aware of.

With respect to context-aware routing, HiBOp is more general, HiBOp is a fully context-aware routing protocol completely described in [19]. HiBOp exploits every type of context information for taking routing decisions and also describes mechanism to handle this information. In HiBOp, devices share their own data when they come into contact with other devices, and thus learn the context they are immersed in. Nodes seem as good forwarders, which share more and more context data with the message destination.

III. INTEGRATED ROUTING

Real users are likely to move around randomly or in predictable fashion, such that if a node has visited a location several times before, it is likely that it can visit that location again or can choose a new location that has never visited before. In this way users' movement can be predictable or unpredictable. We would like to make use of these observations to improve routing performance by combining probabilistic routing with flooding based routing and thus, we propose integrated routing protocol for opportunistic network.

To accomplish this, each node needs to know the contact probabilities to all other nodes currently available in the network. Every node maintains a probability matrix same as described in [7]. Each cell represents contact probability between to nodes x and y. Each node computes its contact probabilities with other nodes whenever the node comes in to contact with other nodes. Each node maintains a time attribute to other available nodes, the time attribute of a node is only updated when it meets with other nodes.

Two nodes exchange their contact probability matrices, when they meet. Nodes compare their own contact matrixes with other nodes. A node updates its matrix with another nodes' matrix if another node has more recent updated time attribute. In this way, two nodes will have identical contact probability matrices after communication.

A. Probability calculation

The calculations of the delivery predictabilities have described in [7]. The first thing to do is to update the metric whenever a node meets with other nodes, so that nodes that are often met have a high message delivery probability. When node x meets node y, the delivery probability of node x for y is updated by (1).

$$P'_{xy} = P_{xy} + (1 - P_{xy}) P_0 \quad (1)$$

Where P_0 is an initial probability, we used $P_0 = 0.75$. When node x does not meet with node y for some predefine time, the delivery probability decreases by (2).

$$\boldsymbol{P'_{xy} = \alpha^k P_{xy}} \quad (2)$$

Where α is the aging factor (a < 1), and k is the number of time units since the last update. When node x receives node y's delivery probabilities, node x may compute the transitive delivery probability through y to z by (3).

$$P'_{xz} = P_{xz} + (1 - P_{xz}) P_{xy} P_{yz} \beta \quad (3)$$

Where β is a design parameter for the impact of transitivity, we used β = 0.25.

B. Routing strategies

when a message arrives at a node, there might not be a path to the destination available so the node have to buffer the message and upon each encounters with another node, the decision must be made on whether or not to transfer a particular message. Furthermore, it may also be sensible to forward a message to multiple nodes to increase the probability that a message is really delivered to its destination.

Whenever a node meets with other nodes, they all exchange their messages (or as above, probability matrix). If the destination of a message is the receiver itself, the message is delivered. Otherwise, if the probability of delivering the message to its destination through this receiver node is greater than or equal to a certain threshold, the message is stored in the receiver's storage to forward to the destination. If the probability is less than the threshold, the receiver discards the message. If all neighbors of sender node have no knowledge about destination of message and sender has waited more than a configured time, sender will broadcast it to all its current neighbors. This process will be repeated at each node until it reaches to destination.

In this paper, we have developed a simple routing protocol –a message is transferred to the other node when two nodes meet, if the delivery probability to the destination of the message is higher than other node. But, taking these decisions is not an easy task. In some cases it might be sensible to select a fixed value and only give a message to nodes that have delivery probability greater than that fixed value for the destination of the message. On the other hand, when encountering a node with low delivery predictability, it is not certain that a node with a higher metric will be encountered within reasonable time. It may be possible destination is new and context information about destination is not spread in network. To solve these problems we introduce a new concept, our integrated routing distributes copies of message to all its neighbors same as flooding based techniques, after a configurable time, when node has not have any context information about destination of message.

Furthermore, we can also set the maximum number of copies of a message; a node can spread, to solve the problem of deciding how many nodes to give a certain message to. Distributing a message to a large number of nodes increases message delivery probability and decreases message delay, on the other hand, also increases resource consumption.

IV. SIMULATION SETUP

We have currently implemented four different routing protocols epidemic, PROPHET, PROPHET (with no POIs) and integrated routing protocols in ONE (Opportunistic Network

Environment) Simulator, all of which we consider in our evaluation. We are taking simulation scenario from [20], therefore we are not describing all things here. We are just showing here some important and new parameters.

We have used part of the Helsinki downtown area (4500×3400 m) as depicted in [20].For our simulations, we assume communication between modern mobile phones or similar devices. Devices has up to 20 MB of free RAM for buffering messages. Users travel on foot, in cars or trams. In addition, we have added to the map data some paths to parks, shopping malls and tram routes. We run our simulations with 100 nodes. Mobile nodes have different speed and pause time. Pedestrians move at random speeds of 0.5–1.5 m/s with pause times of 0–120 s. Cars are optional and, if present, make up 20% of the node count; they move at speeds of 10–50 km/h, pausing for 0–120 s. 0, 2, 4, or 6 trams run as speeds of 7–10 m/s and pause at each configured stop for 10–30 s. We assume Bluetooth (10 m range, 2 Mbit/s) and a low power use of 802.11b WLAN (30 m range, 4.5 Mbit/s). Mobile users (not the trams or throw-boxes) generate messages on average once per hour per node. We use message lifetimes of 3, 6, and 12 hours. We use message sizes uniformly distributed between 100 KB (text message) and 2 MB (digital photo).

Additionally, we define two scenarios POIs1 and POIs2 using different POIs each contains five groups and creates four POI groups (west containing 3, central 4, shops 22, and parks 11 POIs) [20]:

- **POIs1:** One node group runs MBM (map-based model), three choose their next destination with a probability p = 0.1 for each of the four POI groups, the last remaining one only chooses from the POI groups that are accessible by car otherwise a random target is selected.
- **POIs2:** We consider a preferred POI group for four of the node groups. A node chooses a POI with p = 0.4 from its preferred POI group, with p = 0.1 from each other POI group, and otherwise a random target.

V. SIMULATION RESULTS

Now we compare the performance of epidemic, PROPHET and integrated routing protocols in both scenarios when context information is present or not. Here PROPHET (no POIs) stands for PROPHET routing protocol without context information about users, same meaning is here of integrated (no POIs) routing protocol. No POIs means, nodes have no information about destinations behavior and moving pattern, we do this by assigning 0.0 probabilities to each POI (point of interest). While, PROPHET stands for standard probabilistic routing protocol and integrated stands for our new routing protocol with context information about users.

Here figure 1 and 2 show message delay and message delivery probability of epidemic routing. Figure 3, 4, 5 and 6 represent message delay and message delivery probabilities of PROPHET and PROPHET with no POIs routing. Message delay and message delivery probabilities of integrated and integrated with no POIs are represented in figure 7, 8, 9 and 10.

Figure 11 show comparison of message delay between all routing protocols. It is cleared by this figure, when context information is available about users PROPHET gives minimum message delay probability 0.2370. But in absence of context information it gives maximum message delay probability 0.2824, that we represent by PROPHET (no POIs). Epidemic is totally flooding based routing protocol and does not require context information for message forwarding therefore it is not affected by unavailability of context information, and gives same message delay probability 0.2738 in both cases. Our own integrated routing gives 0.2480 and 0.2603 message delay probability in presence and absence of context information.

Comparison of message delivery probability between all routing protocols is shown in figure 12. Same as in case of message delay, PROPHET gives better message delivery probability 0.2981, but on unavailability of context information it gives worst message delivery probability 0.1978. Epidemic does not use context information, therefore gives same delivery probability 0.2334 in both cases. Our integrated routing gives 0.2822 delivery probability and 0.2506 delivery probability is given by integrated (no POIs) routing.

Table 1 shows a summary of message stats report of five routing protocols, which we have implemented. Here variable "sim_time" stands for total simulation time, "created, started, relayed, aborted and dropped" represent number of messages created by simulator, started for transmission, relayed by nodes and dropped by network. Whereas "delivery_prob" stands for total probability of messages delivery, "delay_prob" stands for total probability of messages delay, "hopcount_avg" represents average of intermediate nodes travelled by messages and "buffertime_avg" stands for Average of time Messages were buffered at nodes.

Our simulation results show that PROPHET gives batter result in presence of context information. When users are very isolated, context information cannot be distributed, and cannot be used for taking effective routing decisions. In this case PHROPHET gives worst result. Epidemic gives common result in both cases we have described above. And our integrated routing gives better result in both scenarios context information is available on not. Therefore integrated routing protocol is better when users are social or isolated.

VI. CONCLUSIONS

In this work, we have proposed an integrated routing for opportunistic networks and evaluated its performance across a range of parameters' values, in comparison with Epidemic, PROPHET and PROPHET (with no POIs) routings. We have observed that our proposed integrated routing is able to meet out the challenges of other routing schemes for the

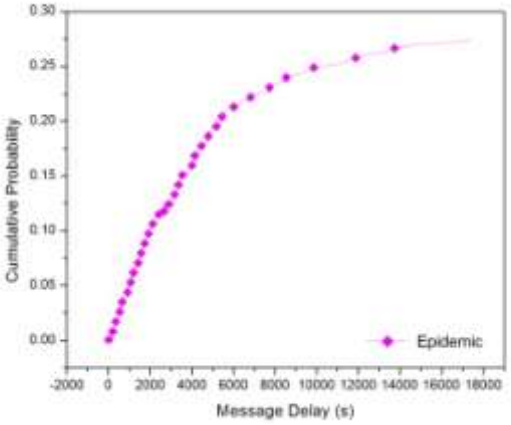


Figure 1. Message delay of Epidemic routing.

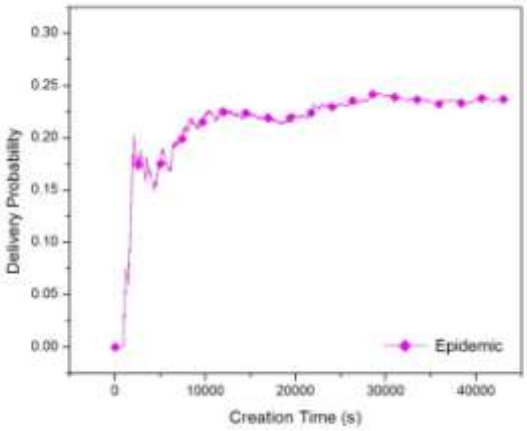


Figure 2. Delivery probability of Epidemic routing.

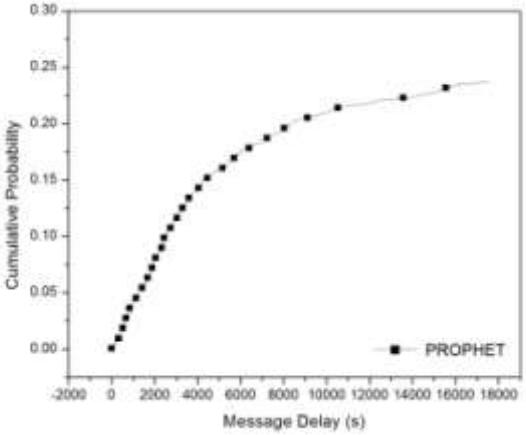


Figure 3. Message delay of PROPHET routing.

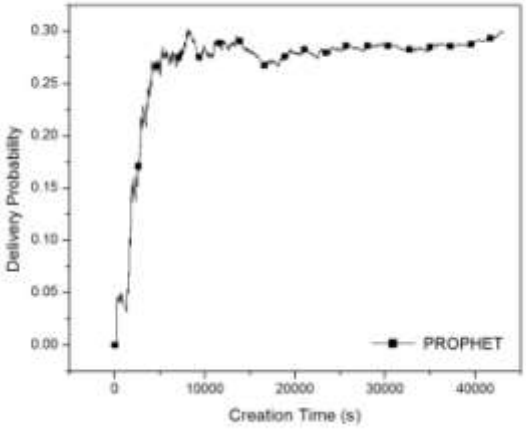


Figure 4. Delivery probability of PROPHET routing.

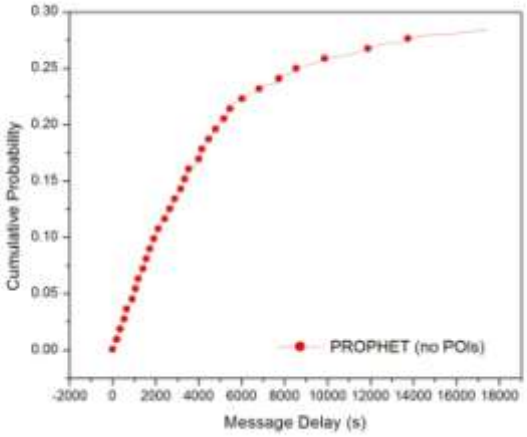


Figure 5. Message delay of PROPHET (no POIs) routing.

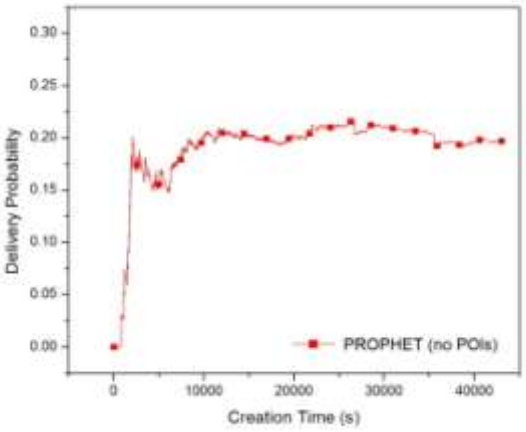


Figure 6. Delivery probability of PROPHET (no POIs) routing.

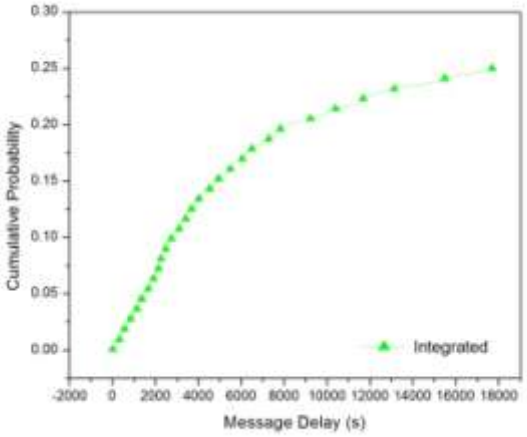


Figure 7. Message delay of Integrated routing.

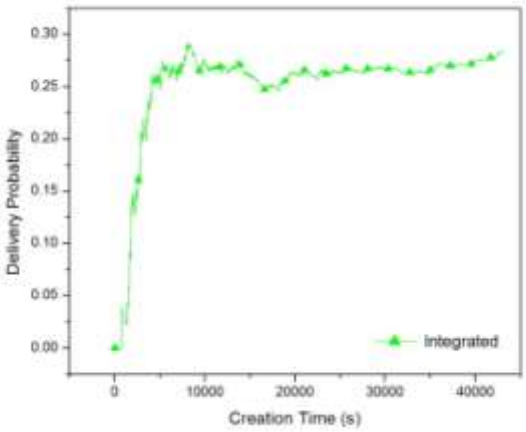


Figure 8. Delivery probability of Integrated routing.

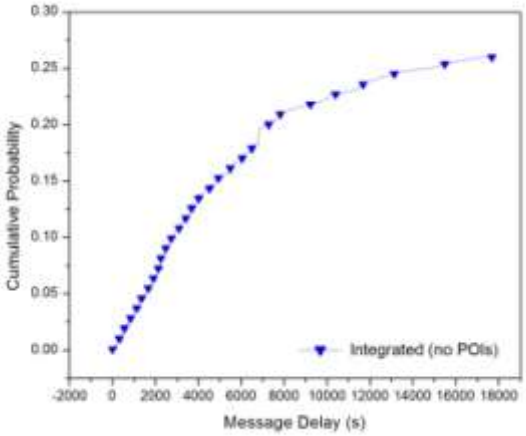


Figure 9. Message delay of Integrated (no POIs) routing.

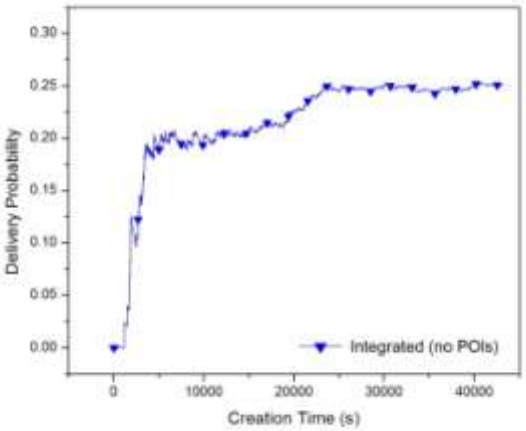


Figure 10. Delivery probability of Integrated (no POIs) routing.

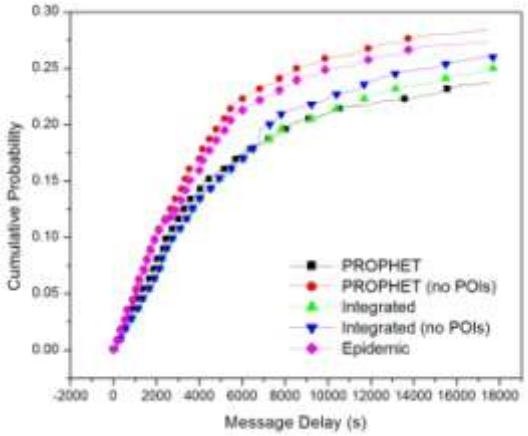


Figure 11. Comparison of Message delay of routing protocols.

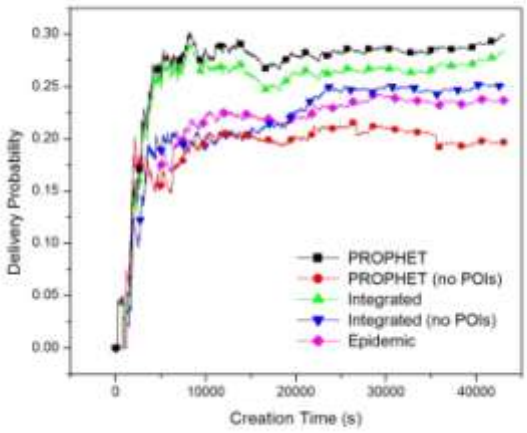


Figure 12. Comparison of Delivery probability of routing protocols.

TABLE I. MESSAGE STATS REPORT

MESSAGE STATS REPORT					
	Epidemic	PROPHET	PROPHET (no POI)	Integrated	Integrated (no POIs)
sim_time:	43200.1000	43200.1000	43200.1000	43200.1000	43200.1000
created:	1461	1461	1461	1461	1461
started:	51790	67827	58011	63532	60312
relayed:	26212	39023	30856	36451	33768
aborted:	25577	28800	27154	27079	26541
dropped:	26262	38990	30821	37520	32479
delivery_prob:	0.2334	0.2981	0.1978	0.2822	0.2506
delay_prob:	0.2738	0.2370	0.2824	0.2480	0.2603
hopcount_avg:	3.8689	3.5145	4.3112	3.6429	3.7951
buffertime_avg:	1430.1069	1054.1910	1531.6268	1132.2647	1298.4176

opportunistic networks, particularly the message delay and delivery probability, when context information about user is available or not. The present findings clearly indicates that the context-based forwarding is a very interesting approach of communication in opportunistic networks, however, in comparison to flooding-based protocols it is not suitable. The present routing is able to give better result in presence as well as absence of context information, specifically in term of message delay and delivery probability.

Despite this, a number of directions exist in integrated routing which can be further investigated. For example we can improve performance of integrated routing in terms of message delay, message delivery, network congestion and resource consumption etc. Developing a network theory to model users' social relationships and exploit these models for designing routing protocols, this is a very interesting research direction in opportunistic network.

ACKNOWLEDGMENT

The authors are thankful to ABV-IIITM, Gwalior for the financial support, they have provided for carrying out present work. We are grateful to Prof. A. Trivedi, ABV-IIITM, Gwalior for his valuable interactions.

REFERENCES

[1] Borgia, E., Conti, M., Delmastro, F. and Pelusi, L. (2005) 'Lessons from an Ad-Hoc network test-bed: middleware and routing issues', Ad Hoc and Sensor Wireless Networks, an International Journal, Vol. 1.

[2] Pelusi, L., Passarella, A. and Conti, M. (2006) 'Opportunistic networking: data forwarding in disconnected mobile ad hoc networks', IEEE Communications Magazine, Vol. 44, pp.134–141.

[3] Marco Conti and Mohan Kumar, "Opportunities in opportunistic computing" Published by the IEEE Computer Society, JANUARY 2010, pp. 42-50.

[4] Fall, K. (2003) 'A delay-tolerant network architecture for challenged internets', Paper presented in the Proceedings of the 2003 ACM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM 2003), pp.27–34.

[5] Jain, S., Fall, K. and Patra, R. (2004) 'Routing in a delay tolerant network', Paper presented in the Proceedings of the 2004 ACM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM 2004), pp.145–158, New York, NY: ACM.

[6] Vahdat, A. and Becker, D. (2000) Epidemic Routing for Partially Connected Ad Hoc Networks. Technical Report CS-2000-06, Computer Science Department. Duke University.

[7] Lindgren, A., Doria, A. and Schelen, O. (2003) 'Probabilistic routing in intermittently connected networks', ACM Mobile Computing and Communications Review, Vol. 7, pp.19–20.

[8] L.Pelusi, A.Passarella, and M.Conti, "Opportunistic Networking: dataforwarding in disconnected mobile adhoc networks", IEEE Communications Magazine, 44(11), Nov.2006.

[9] Widmer, J. and Le Boudec, J. (2005) 'Network coding for efficient communication in extreme networks', Paper presented in the Proceedings of the ACM SIGCOMM 2005 Workshop on Delay Tolerant Networking (WDTN 2005), pp.284–291.

[10] Pelusi, L., Passarella, A. and Conti, M. (2007) Handbook of Wireless Ad Hoc and Sensor Networks, Chapter Encoding for Efficient Data Distribution in Multi-Hop Ad Hoc Networks. New york, NY: Wiley and Sons Publisher.

[11] Spyropoulos, T., Psounis, K. and Ragavendra, C. (2007) 'Efficient routing in intermittently connected mobile networks: the multiple-copy case', ACM/IEEE Transactions on Networking, Vol. 16, pp.77–90.

[12] Burns, B., Brock, O. and Levine, B. (2005) 'MV routing and capacity building in disruption tolerant networks', Paper presented in the Proceedings of the 24th IEEE Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2005).

[13] Burgess, J., Gallagher, B., Jensen, D. and Levine, B. (2006) 'MaxProp: routing for vehicle-based disruption-tolerant networks', Paper presented in the Proceedings of the 25th IEEE Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2006).

[14] Leguay, J. and Friedman, T.C.V. (2006) 'Evaluating mobility pattern space routing for dtns', Paper presented in the Proceedings of the 25th IEEE Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2006), pp.1–10.

[15] Hui, P. and Crowcroft, J. (2007) 'Bubble rap: forwarding in small world dtns in every decreasing circles', Technical report, Technical Report UCAM-CL-TR684. Cambridge, UK: University of Cambridge.

[16] Grossglauser, M. and Vetterli, M. (2003) 'Locating nodes with EASE: last encounter routing in ad hoc networks through mobility diffusion', Paper presented in the Proceedings of the 22nd IEEE Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM 2003).

[17] Spyropoulos, T., Psounis, K. and Ragavendra, C. (2007) 'Efficient routing in intermittently connected mobile networks: the multiple-copy case', ACM/IEEE Transactions on Networking, Vol. 16, pp.77–90.

[18] Musolesi, M., Hailes, S. and Mascolo, C. (2005) 'Adaptive routing for intermittently connected mobile ad hoc networks', Paper presented in the Proceedings of the IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM 2005), pp.183–189.

[19] Boldrini, C., Conti, M. and Passarella, A. (2007) 'Impact of social mobility on routing protocols for opportunistic networks', Paper presented in the Proceedings of the First IEEE WoWMoM Workshop on Autonomic and Opportunistic Networking (AOC 2007).

[20] KERÄNEN, A., AND OTT, J. Increasing Reality for DTN Protocol Simulations. Tech. rep., Helsinki University of Technology, Networking Laboratory, July 2007.

[21] Ramasamy, U. (2010). Classification of Self-Organizing Hierarchical Mobile Adhoc Network Routing Protocols - A Summary. International

Journal of Advanced Computer Science and Applications - IJACSA, 1(4).

[22] Bhakthavathsalam, R., Shashikumar, R., Kiran, V., & Manjunath, Y. R. (2010). Analysis and Enhancement of BWR Mechanism in MAC 802 . 16 for WiMAX Networks. International Journal of Advanced Computer Science and Applications - IJACSA, 1(5), 35-42.

[23] Jawandhiya, P. M. (2010). Reliable Multicast Transport Protocol : RMTP. International Journal of Advanced Computer Science and Applications - IJACSA, 1(5).

AUTHORS PROFILE

ANSHUL VERMA is a M.Tech (Digital Communication) student in the Computer Science and Engineering department at ABV-Indian Institute of Information Technology and Management, Gwalior, M.P., India. He has research interest in database management systems, Opportunistic Network, wireless communications and networking.

Dr. ANURAG SRIVASTAVA, working as an Associate Professor in Applied Science Departement at ABV-Indian Institute of Information Technology and Management, Gwalior, M.P., India. His Research activities are based on Model Calculation, Computational Physics/ Nano-Biophysics, Nano Electronics, IT Localization, E-Governance, IT Applications in different domains.

An Electronic Intelligent Hotel Management System For International Marketplace

Md. Noor-A-Rahim[1], Md. Kamal Hosain[2], Md. Saiful Islam[3], Md. Nashid Anjum[4] and Md. Masud Rana[5]
[1,3,4,5] Electronics & Communications Engineering, Khulna University of Engineering and Technology (KUET), Khulna, Bangladesh.
[2]School of Engineering, Deakin University, Geelong, Victoria 3217, Australia.
E-mail: mash_0409@yahoo.com[1], mhosain@deakin.edu.au[2], sirajonece@yahoo.com[3], nashidzone@yahoo.com[4], and mamaraece28@yahoo.com[5]

***Abstract*—To compete with the international market place, it is crucial for hotel industry to be able to continually improve its services for tourism. In order to construct an electronic marketplace (e-market), it is an inherent requirement to build a correct architecture with a proper approach of an intelligent systems embedded on it. This paper introduces a web based intelligent that helps in maintaining a hotel by reducing the immediate involvement of manpower. The hotel reception policy, room facilities and intelligent personalization promotion are the main focuses of this paper. An intelligent search for existing boarders as well as room availability is incorporated in the system. For each of the facilities, a flow chart has been developed which confirms the techniques and relevant devices used in the system. By studying several scenarios, the paper outlines a number of techniques for realization of the intelligent hotel management system. Special attention is paid to the security and also prevention of power and water wastages. In this power saving scenery, an image processing approach is taken to detect the presence of any people and the darkness in a particular room. Moreover, this proposed automated computerized scheme also takes an account of the cost advantage. Considering the scarcity of manpower in several countries, the objective of this paper is to initiate the discussion and research for making the proposed systems more commercialized.**

Keywords- E-marketplac; hotel managemen; intelligent search; intelligent system; image processing algorithms; web-based application

I. INTRODUCTION

In order to triumph in the fierce competition of hotel service industry, the main goal and orientation of hotel management are how to provide customers high quality and humanized services. To compete with the international e-marketplace, a great deal of attention should pay towards the optimization of user requirements to generate recommended hotel alternatives [1]. In general sense, hotel management is the way of maintaining different activities of a hotel where a number of staffs are engaged to perform a number of these activities. At first let us take a glance to an ordinary hotel. For hiring a room in this type of hotel, the client needs to meet with the receptionist to collect the information of hotel facilities[2]. After that he is to fill up the pro forma provided by the hotel authority, then he has to pay the defined amount of money and is offered room key for his/her rented room. He/she is then finishes the formalities at reception zone through these undergoing customs. But client always wants greater privacy and reliable security. Koolmanojwong et al. [3] developed an intelligent e-marketplace for the tourism based on fuzzy to serve the customers who wants to travel but has no idea about the accommodation[4]. This system is global in the sense that anyone can use this to find the appropriate hotel according to his/her affordable means [5]. The details of the hotel management systems including the franchising, casinos, health Spas, payroll, credit, accounting control etc. are well described in [6]. However, we have designed an IHM system which is specific to a particular hotel. It helps the owner to serve the intended customers without directly involving with them. This system has included the electronic circuitry embedded with several sensors in integrated with the java programming.

The proposed intelligent hotel management (IHM) system is free from a significant number of hotel staffs that provides those facilities and fewer formalities. In mal-populated countries dearth of manpower is increasing gradually. Therefore, they have to import manpower from other countries. In this condition the IHM can be a permanent solution. Moreover, it possesses adequate security [7]. This system provides hi-tech room facilities including auto controlled door, automatic light controlling, voice active devices etc. Apart from these, it prevents the waste of electric power as well as excessive water that are the main ideas used in this paper. A short version of this approach is in [8]. Additionally, we have integrated a new image processing approach which accurately ensures the presence and darkness of the room to be occupied.

II. PROPOSED INTELLIGENT HOTEL MANAGEMENT SYSTEM

Generally, the information provided in conducting an international electronic marketplace (E-marketplace) for hotel management is not enough. The intelligent functions as a helping hand to the new visitors in such a place a great integration and initiative is offered[9]. In an ordinary hotel management system there must be a receptionist to interact with customers. Since IHM involves a fewer number of hotel staff, there should be a way in reception zone to distinguish between a client (who wants to rent a room for accommodation) and a guest (who wants to visit a border in a hotel). Some indispensable activities such as pro forma filling,

bill payment, getting the room key should be accomplished in reception zone. To provide the substantial information about the competences of the hotel, a virtual hotel guide is an urgent need. The general system function of an intelligent hotel management is shown in Fig. 1 which includes a reception zone and intelligent architecture. We have proposed an intelligent management of a modern hotel where all the facilities and services to the customers have been maintained in an efficient way.

The following Fig. 2 furnishes an overview about this proposed IHM architecture. It depicts that the reception zone is operated by the central controlling unit; room devices and other activities of rooms are operated by local controlling unit. Each local controlling unit is connected with central controlling unit. For overall security there is a security zone. It monitors the whole system including reception zone, hotel corridor and other pertinent areas from any unpleasant incidents. The following section provides a possible way to establish such type of system. This proposed system includes a reception procedure, room facilities, and border searching. In addition, we have implemented an image processing algorithm which ensures the prevention power wastage. This whole system was implemented in our lab (Fig. 3) which worked perfectly.

A. Intelligent reception procedure for hotel accommodation

When a client attends to the reception booth the infra sensor gets activated. A display is then appeared on the monitor/electronic screen containing three options: i) hotel information, ii) new entry, and iii) border searching. If the new entry option is selected a pro forma is appeared on the screen. It contains information fields about the client e.g., client's name, address, and other information. It also contains the room category and rent duration fields [10], [11]. The first selection shows the available room categories and the later one ensures the number of days for which the room is rented. After submitting the pro forma accurately, the display screen will show the amount of payment of the particular room. After completing these schedules, when the client leaves the booth the infra sensor becomes deactivated automatically and the screen resumes to the initial state[12]. This step by step information shows in the following Fig. 4 in details.

B. Intelligent management of room facilities

When the border comes in front of his rented room then another infra sensor is activated and the border is instructed automatically to type his/her room password. The door of the room will be opened automatically if the typed password matches with the password which is created and confirmed at the reception booth. If the room is empty then an intelligent image processing method runs as soon as the door is opened. This method measures the darkness of this room [13], [14]. This process takes an image of the room as soon as the door opens. The image is then processed by which the darkness of the room is compared. If the room is not enough lighted then the electric appliances are turned on automatically. The entire process is followed by the following flow chart (Fig. 5).

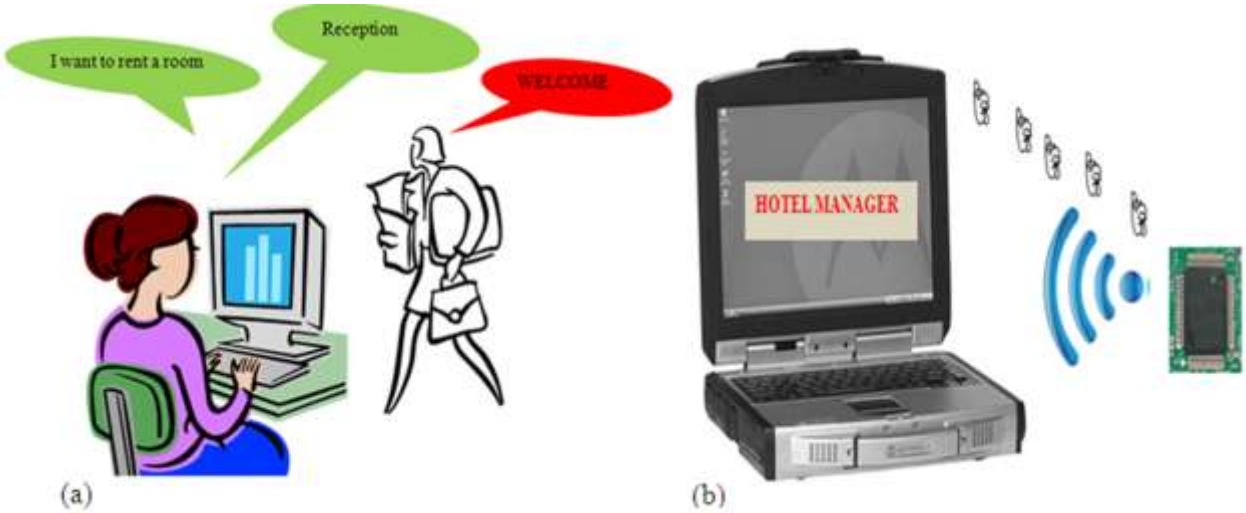


Figure 1. System function of an automated hotel, (a) Register in reception, (b) Information sending.

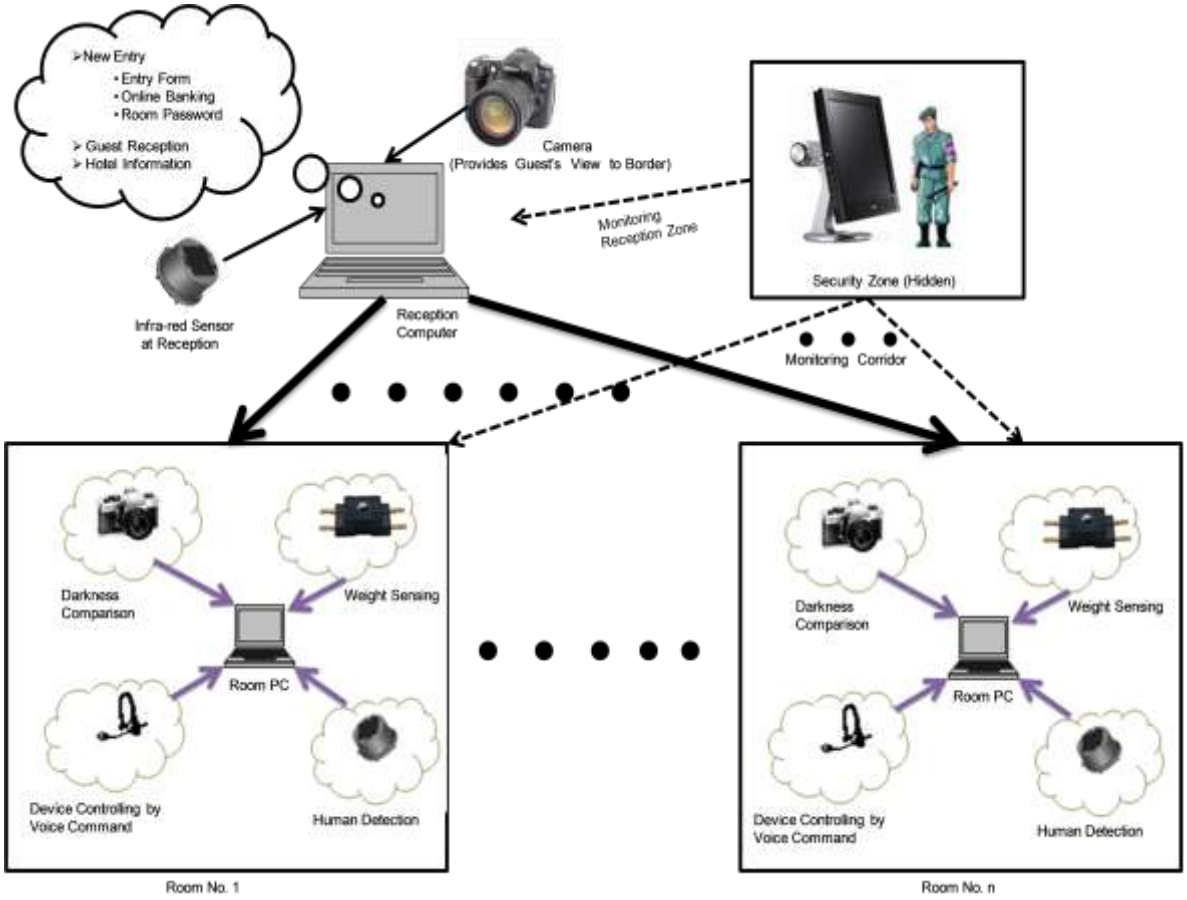


Figure 2. Block diagram illustrating the proposed intelligent hotel management system.

1) Image processing scheme: In this hi-tech room facilities, an intelligent image processing has been integrated which ensures the presence and darkness of the intended room. A web camera is set in a suitable location inside the room which can capture snap of the significant portion of the room including the area where usually the person stays. We have used the built-in function of java script which extracts the RGB value from the captured image[15], [16], [17]. This program is written and integrated with the system specifically for this design. The process of monitoring the room's light is explained in Fig. 6.

2) Controlling devices through voice command scheme: In an ordinary hotel, clients used to operate the electronic devices manually. But automatic control is more suitable and easier than manual control. This proposed IHM system provides the benefits to control the electronic equipments such as: light, fan or air-condition, TV etc. using voice command. For example, to turn on the light simply 'light on' command is enough. This is done by following process as shown in Fig. 7. Besides these facilities, this proposed method includes manual control system as well[18], [19].

3) Prevention of wastage: IHM is well concerned about wastage prevention. To prevent wastage of water it uses infra sensor which is used to auto turn on and turn off the faucet. It is done by following way as shown in Fig. 8. On the other hand IHM also prevents the electric power. When the room becomes empty the lights, fans and other electronics devices are automatically turned off.

Figure 3. Basic arrangement and test of the proposed IHM system.

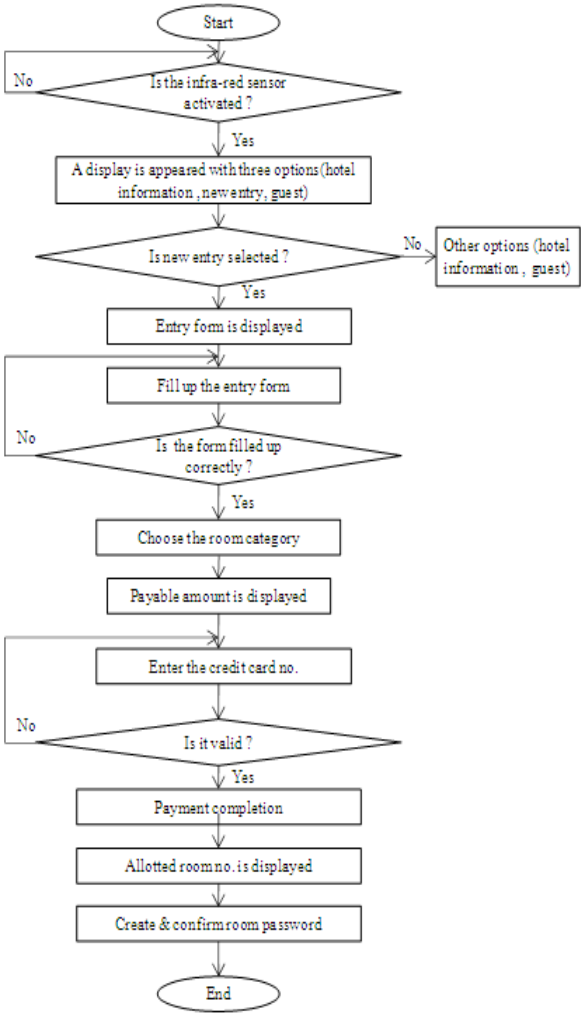


Figure 4. Flow chart of entry of a customer in a proposed hotel system.

III. Intelligent Searching of a Border

While staying in a hotel, it is important to maintain the privacy and of course with high security. Before, allowing the guest to meet with the border the border must have a chance to see him so he/ she might have the right whether allowed or not the guest. Figure 9 illustrates this interaction between the border and the guest who comes to the hotel to meet a desired border.

When the guest comes to the reception booth, again the infra sensor will be activated and the electronic screen will display three options as described in previously. Now the guest will select the border searching option (also termed as Guest option). Then there will appear a form, the purpose of which is to take the guest's name and address. It also contains two preferences to take desired border's name and room number. If the desired border is not available then a message will be displayed on the screen to inform that news [13]. But if the desired border is available then a confirmation message will be sent to the border to inform him that he has a guest. And a display with three options will be appeared on the screen which is placed in the border's room. Using first option the border will be able to verify the identity of the guest by watching him live on the screen of his own room. Using the second and third option he will be able to allow or deny the guest. If the border allows the guest then the system will generate a random password which can be [20] used for only one time and within a fixed period of time. If the border denies the guest then the guest will receive a message which informs that the border is not available.

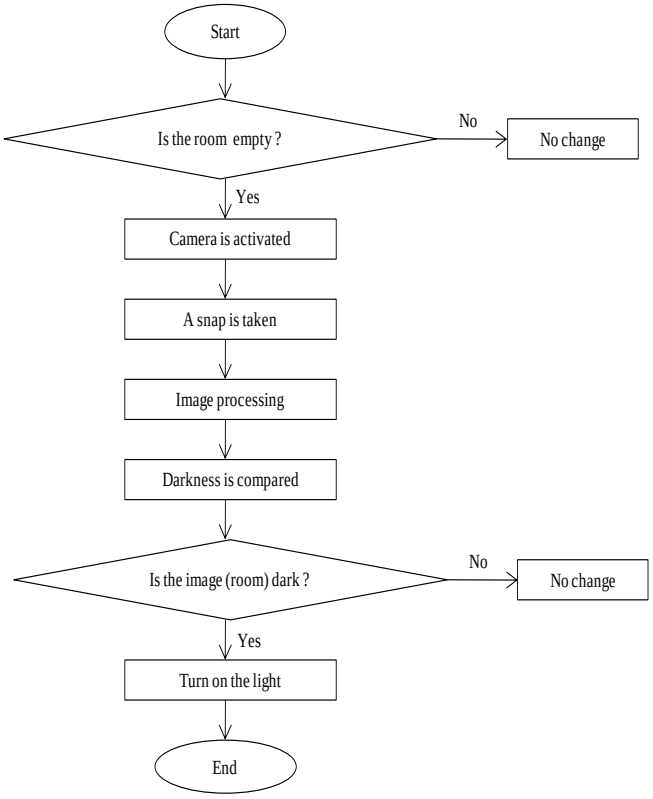


Figure 5. Flow chart of the to maintain the automatic turn on the light (when the border enters in an empty room.

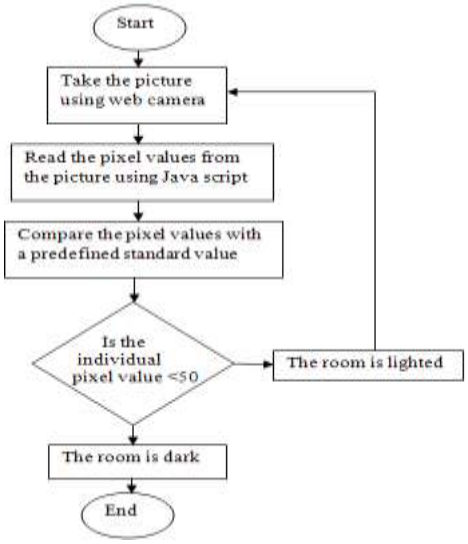


Figure 6. Flow chart of to show process of monitoring the room's condition.

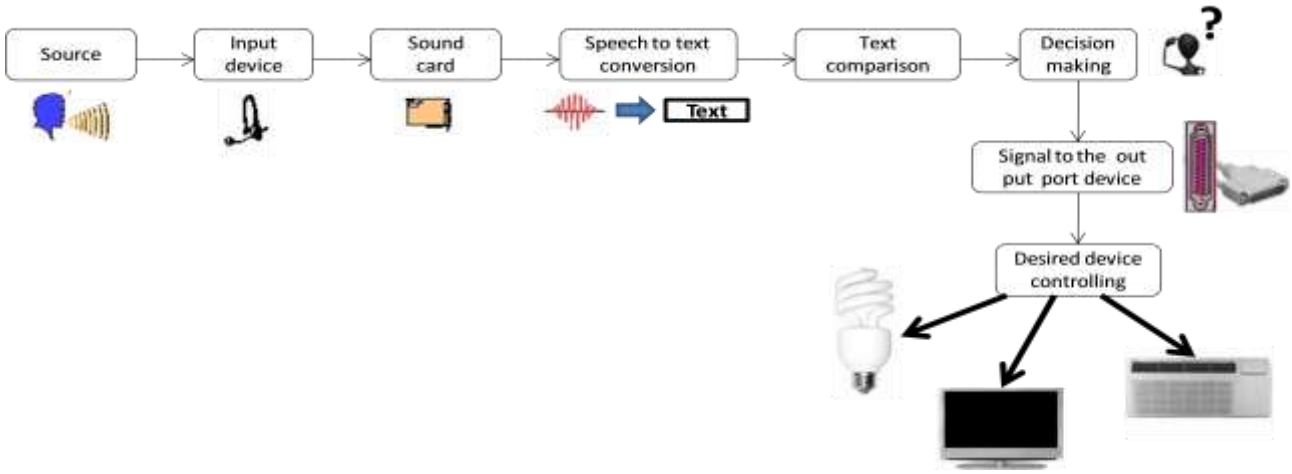


Figure 7. Block diagram of controlling devices by voice command.

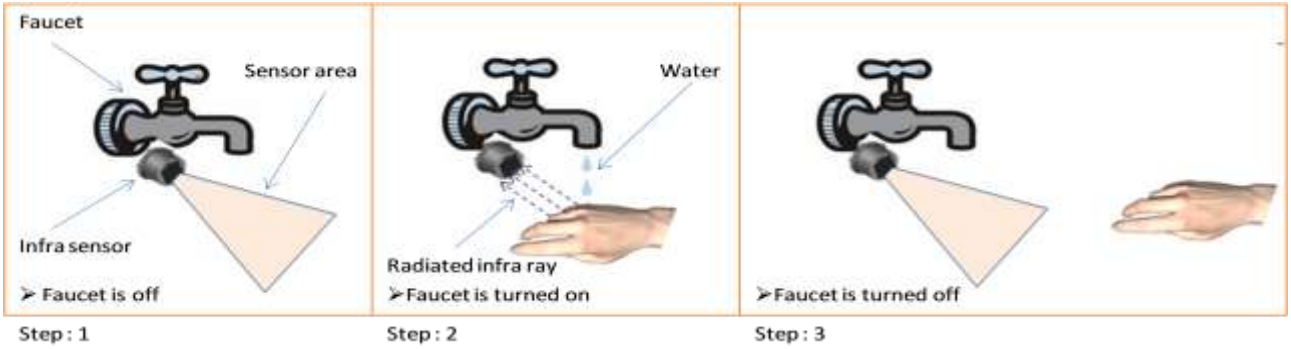


Figure 8. An approach of infra sensor for the controlling of water to be utilized.

IV. Future Works and Motivations

In order to meet the full customer requirements and to compete with the international marketplace the proposed IHM should incorporate the following facilities: intelligent room condition indicator, intelligent id recognizer, room condition indicator, system self-checking, the announcement of information, air-conditioner controller[21]. We are now considering the above approaches to be incorporated with the IHM system.

V. Conclusion

In this era of high technology, everything is attaining more and more automation dependent. Hotel management system also should be involved in the realm of automation. This proposed intelligent management system provides high level privacy than the existing conventional manual system with greater reliability. To satisfy the customer's need, this project work provides a seamless and enjoyable experience for customers.

Introducing this automotive management system in any kind of accommodation systems greatly reduced manpower and maintenance cost. In addition, the incorporation of infrared detection systems and the image processing scheme in the respective rooms helps to the prevention of power and water wastages. Moreover, the web based system increases the security and privacy by employing a web camera which ensures the live pictures of any occurrence happened in the hotel. This is our own concept and has been already successfully implemented as a project work in miniature version. In fact, this system is fast, comprehensive and flexible, but doesn't necessarily require ones to have that much skill in computer science.

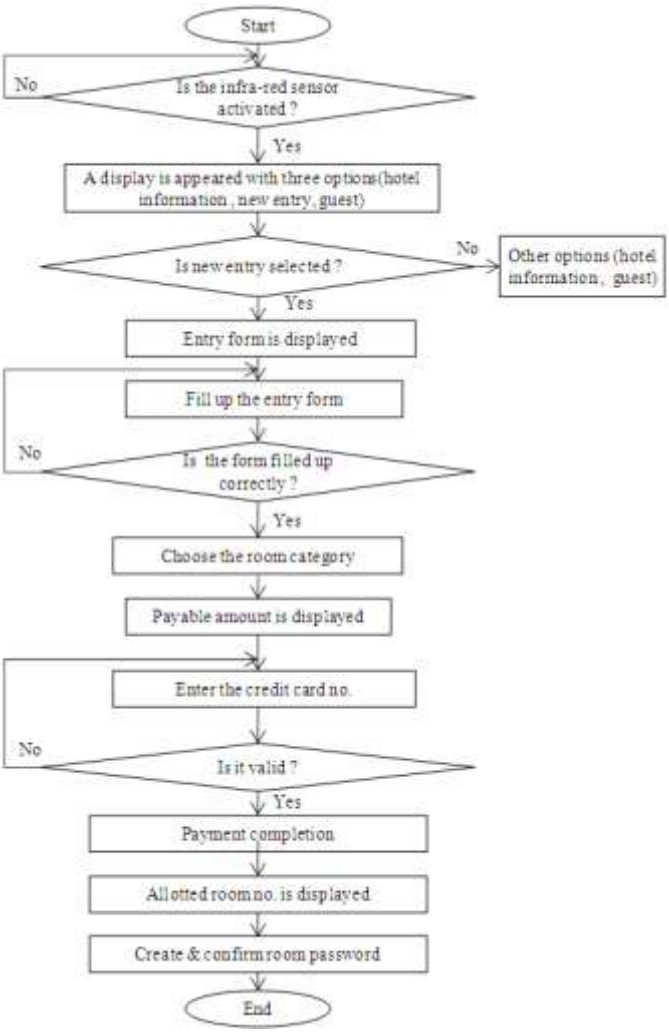


Figure 9. Flow chart to show the entry procedure of a customer in the proposed IHM.

REFERENCES

[1] S. Koolmanojwong, "Analysis and Design of B to C E-Marketplace for Tourism with UML " M.S. Thesis, Faculty of Science and Technology, Assumption University, Bangkok, Thailand 2000.

[2] M. J. O'Fallon and D. G. Rutherford. (2011). *Hotel Management and Operations | CA College of Ayurveda*. Available: http://www.ayurvedacollege.com/amazon_store/item/0470177144

[3] S. Koolmanojwong and P. Santiprabhob, "Intelligent Electronic Marketplace for Tourism."

[4] E. W. T. Ngai and F. K. T. Wat, "Design and development of a fuzzy expert system for hotel selection," *Omega*, vol. 31, pp. 275-286, 2003.

[5] G. Jingzhi, *et al.*, "Alibaba International: Building a Global Electronic Marketplace," in *e-Business Engineering, 2006. ICEBE '06. IEEE International Conference on*, 2006, pp. 545-548.

[6] W. S. Gray and S. C. Liguori, *Hotel and Motel Management and Operations*, Fourth Edition ed.: Prentice Hall, 2002.

[7] W. J. Relihan Iii, "The yield-management approach to hotel-room pricing," *The Cornell Hotel and Restaurant Administration Quarterly*, vol. 30, pp. 40-45, 1989.

[8] M. S. Islam, *et al.*, "An Automated Intelligent Hotel Management System," in *2009 Interdisciplinary Conference in Chemical, Mechanical and Materials Engineering (2009 ICCMME)*, Melbourne, Australia, 2009.

[9] J. Guo, *et al.*, "CONFENIS Special Session on the International Symposium on Electronic Marketplace Integration & Interoperability (EM2I'07)," in *Research and Practical Issues of Enterprise Information Systems II*. vol. 255, L. Xu, *et al.*, Eds., ed: Springer Boston, 2008, pp. 823-823.

[10] K. Riemer and C. Lehrke, "Biased Listing in Electronic Marketplaces: Exploring Its Implications in On-Line Hotel Distribution," *Int. J. Electron. Commerce*, vol. 14, pp. 55-78, 2009.

[11] Y.-n. Xiong and L.-x. Geng, "Personalized Intelligent Hotel Recommendation System for Online Reservation--A Perspective of Product and User Characteristics," in *Management and Service Science (MASS), 2010 International Conference on*, 2010, pp. 1-5.

[12] J. A. Bardi. (2011). *Wiley::Hotel Front Office Management 5th Edition*. Available: http://eu.wiley.com/WileyCDA/WileyTitle/productCd-EHEP001784.html?filter=TEXTBOOK

[13] Y. Ming-Ju, *et al.*, "Multicast Services of Digital Television in a New Management Framework," in *Intelligent Information Technology Application Workshops, 2008. IITAW '08. International Symposium on*, 2008, pp. 296-299.

[14] M. Ben Ghalia and P. P. Wang, "Intelligent system to support judgmental business forecasting: the case of estimating hotel room demand," *Fuzzy Systems, IEEE Transactions on*, vol. 8, pp. 380-397, 2000.

[15] R. Jiamthapthaksin, "Fuzzy information retrieval with respect to graded features," Masters Thesis, Assumption University, 2000.

[16] M. Wojcikowski, *et al.*, "An intelligent image processing sensor - the algorithm and the hardware implementation," in *Information Technology, 2008. IT 2008. 1st International Conference on*, 2008, pp. 1-4.

[17] "Session 11: intelligent image processing," in *Computational Intelligence for Measurement Systems and Applications, 2005. CIMSA. 2005 IEEE International Conference on*, 2005, pp. 186-186.

[18] D. B. Roe and Y. Wang, "A voice-controlled network for universal control of devices in the OR," *Minimally Invasive Therapy & Allied Technologies: MITAT: Official Journal Of The Society For Minimally Invasive Therapy*, vol. 9, pp. 185-191, 2000.

[19] H. G. Nik, *et al.*, "Voice Recognition Algorithm for Portable Assistive Devices," in *Sensors, 2007 IEEE*, 2007, pp. 997-1000.

[20] "Hotel Yield Management Practices Across Multiple Electronic Distribution Channels," *Information Technology & Tourism*, vol. 10, pp. 161-172, 2008.

[21] Available: http://www.smarthomeuae.com/productfile/0000000000427a.pdf

The Impact of E-Media on Customer Purchase Intention

Mehmood Rehmani, Muhammad Ishfaq Khan
Department of Management Sciences
Mohammad Ali Jinnah University
Islamabad Pakistan

Abstract--In this research paper, authors investigated the social media (e-discussion, websites, online chat, email etc) parameters that have effect over the customers buying decisions. The research focused on the development of research model to test the impact of social media on the customer purchase intention. The literature review done to explore the work done on social media. The authors identify the problem and defined the objectives of the studies. In order to achieve them, a research model is proposed that followed by the development of research hypotheses to testify the model.

Key words: - e-discussion; e-mail; website; online chat;

I. INTRODUCTION

The media that describes a variety of new sources of online information that are created, initiated, circulated and used by consumers intent on educating each other about products, brands, services, personalities, and issues is call social media [4]. Toivonen [46] termed it as an interaction of people to create, share, exchange and commenting contents in networks and virtual communities. By 1979, Tom Truscott and Jim Ellis of Duke University, created the Usenet, a discussion system that permitted Internet users to post messages. According to Kaplan & Haenlein [27], the era of Social Media probably started about 20 years earlier, when Bruce and Susan Abelson developed ‘‘Open Diary,’’ an early social networking website that brought the online diary writers into one community. Terminology of ‘‘weblog’’ was coined at the same time, and truncated as ‘‘blog’’ a year later when a blogger transformed the noun ‘‘weblog’’ into the sentence ‘‘we blog.’’ The growing access and availability of Internet further added to the popularity of this concept, which lead to the creation of social networking websites such as MySpace (in 2003) and Facebook (in 2004). This, in turn, coined the term ‘‘Social Media,’’ and contributed to the eminence it has today [27].

II. LITERATURE REVIEW

Social media encompasses a wide range of online, word-of-mouth forums including blogs [24], company sponsored discussion boards and chat rooms, Consumer-to-consumer e-mail, consumer product or service ratings websites and forums, Internet discussion boards and forums, moblogs (sites containing digital audio, images, movies, or photographs), and social networking websites, to name a few [27,34,38].

Given the fast changes in the communication system brought about by participative social media and internet [44], the better understanding of these technologies is important [33]. The increasing number of personal wireless devices such as internet connected cellphones, communication scientists anticipate the popularity of social networking websites to grow worldwide [14, 22, 27, 47]. Nielsen Net ratings (2006) reported that the number of users of top 10 social networking sites in U.S. grew from 46.8 million to 68.8 million users during one year.

The use of social media has better reach and impact [33, 44] on younger generation. Nancy [33] found relatively low penetration in the population aged 55 and older which suggests that it is not yet an appropriate time to utilize social media for this age group. Spending time on social networking sites appears to be part of young adults' daily activities [17, 20] and an average 30 minutes Facebook usage has been reported. One of the study found that about half of twelve to seventeen year olds log on daily at social networking site: 22% logged on several times per day, 26% once a day,17% three to five days perweek,15% one or two days per week, and only twenty percent every few weeks or less [30]. However, Nancy (2009) predicted a continuing increase in utilization of social media across all groups and generations in the coming years.

According to an article published in Pakistan Today, based on marketing managers’ opinions, the trend of marketing is changing in Pakistan and typical advertisements are not yielding the desired results. Companies including Wi-tribe Pakistan, Pakistan International Airlines (PIA), Sooper biscuits and Haier Pakistan have so far effectively used the technique of digital marketing, whereas various fans of these products have also started using them. Wi-tribe Marketing Manager told Pakistan Today that these techniques have proved to be quite cost-effective whereas their targeted market base was also in urban areas therefore Internet was the best way to spread their messages. Wi-tribe has gained a lot from digital marketing and fans of Wi-tribe’s Facebook page were now around 19,000 [18].

Social Networking Websites including Facebook and twitter are being used by various multinational companies in Pakistan in order to convey their messages using word of mouth technique.

Businesses are witnessing an explosion of Internet-based messages and information [29] transmitted through social media [18]. They have become a major factor in influencing various aspects of consumer behavior [41]. Unfortunately, the popular business press and academic literature offers marketing managers very little guidance for incorporating social media into their IMC strategies [31]. Therefore, many

managers lack a full appreciation for social media's role in the company's promotional efforts. Even though social media is magnifying the impact consumer-to-consumer conversations have in the marketplace, importance of these conversations has not yet been recognized [31].

Since its appearance in marketing research, purchase intention has been the subject of great attention in the academic environments. Customer behavioral intentions are considered as signals of actual purchasing choice, thus are desirable to be monitored [51]. A study on sentiment analysis of online forums and product reviews exhibited that they influence individual's purchase decisions [48].

Intention to buy is the buyer's forecast of which brand he will choose to buy. It has been used extensively in predicting the purchases of durable goods. Intention to buy may be characterized as response short of actual purchase behavior.

A survey of more than 600 youth respondents, 51 % of whom had done online purchases in the past year) established that nearly 40 % learned about the product online, but bought at a physical place or a store, whereas, only 9.3 % began and ended their search online. When asked, where they would prefer to shop, nearly 75% chose a store over online. Across the field, consumers are combining different channels and approaches, searching online to buy everything in between.

The rapid explosion of social media has melted away the barriers to the flow of information and yet the consequences have barely been felt. Marketing to the Facebook generation will not be confined to harnessing the digital channels, it will change the every way the firm communicates [9] to achieve its objectives. Internet has made it easier than before for marketers to communicate directly with consumers and target audiences, it's of much importance that marketers dramatically alter the PR and marketing strategy to maximize the effectiveness of the direct consumer communication means [42].

Jerry Wind was an early champion of digital marketing, highlighting the revolutionary changes of the Internet on consumer behavior, marketing and business strategy. He urged executives to consider the potential of this new technology to transform their businesses.

Mobile Social Media applications are expected to be the main driver of this evolution, soon accounting for over 50% of the market (Kaplan & Heinlein, 2010). Kaplan & Heinlein further states that in India, mobile phones outnumbered PCs by 10 to 1. Whereas, in Thailand, only 13 percent of the population owns a computer, while 82% have access to a mobile phone.

Pew Research Center, a Washington-based think tank, estimates that by 2020, a mobile device will be the primary Internet connection tool for majority of the people round the globe. Social Media applications mobile is likely to attract a currently unexploited base of new users. Even if per capita spending in developing countries may be low, vast population numbers make them relevant for virtually any business entity [27]. Telecom sector is seeing exorbitant growth in Pakistan. The sector is said to be growing at a fast pace yearly (Pakistan Telecommunication Authority). Mobile subscribers are more than 100 million as of Oct 2010. In fact, Pakistan has the highest mobile penetration rate in the South Asian region [50].

This study intends to explore the impact of social media on purchase intention of mobile phone customers in Pakistan. In view of the growing number of cell phone users, the factors that influence the purchase intention of customers need to be explored. Social Media is being considered playing an important role in customer buying decisions, however little studies have explored its impact over the customer purchase intention. Our study aimed at measuring the impact of social media on customers in Pakistan.

III. Statement of the Problem

In view of the growing number of internet users, the factors of social media that influence the purchase intention of customers need to be explored. Social Media is being considered playing an important role in customer buying decisions, however little studies have explored its impact over the customer purchase intention. This study intends to explore the impact of social media on purchase intention of mobile phone customers in Pakistan.

IV. Research Objectives

The objectives of this research are twofold:

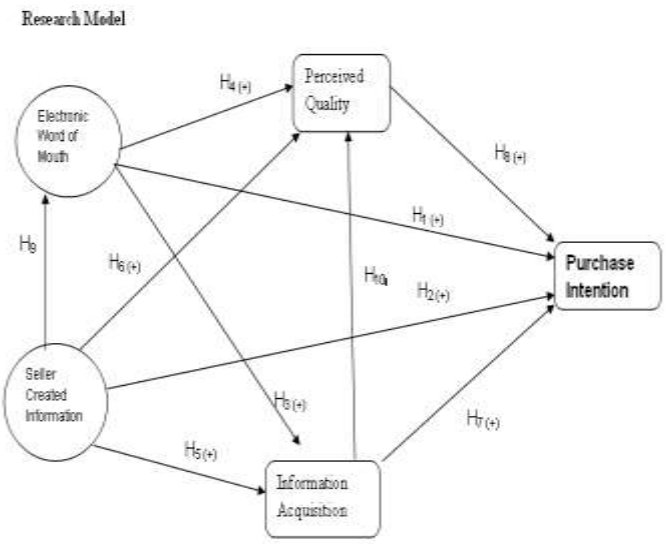
1- to identify the potential of social media for consideration as a hybrid component of the promotional mix and therefore incorporation as an integral part of IMC strategy.
2- to sensitize the marketing managers on shaping the consumer-to-consumer conversations at various social media tools, which are now driving the marketplace to a greater extent (Lefebvre, 2007; Ellison, 2007; Nancy, 2009 and Tiffany, 2009) than ever before.

V. Theoretical Framework

The literature shows that social media construct include eWOM and Seller created information which cause impact on customer's purchase intention. These two variables have been found associated with information acquisition and perceived quality which lead to change in purchase intention of a customer. Information acquisition also proved associated with perceived quality which is a mixture of service and product quality. In view of the above, the following variables have been identified from literature :

1- Electronic Word of Mouth
2- Seller created online information
3- Information Acquisition
4- Perceived Quality (Combination of Product and Service Quality)
5- Purchase Intention

The proposed research model is as under:



VI. RESEARCH HYPOTHESES

The following hypotheses have been developed to testify the model:

- H_1 Electronic word of mouth has positive impact on purchase intention
- H_2 Seller created information has positive impact on purchase intention
- H_3 Electronic word of mouth has positive impact on perceived quality.
- H_4 Electronic word of mouth has positive impact on information acquisition.
- H_5 Seller created information has positive impact on information acquisition.
- H_6 Seller created information has positive impact on perceived quality.
- H_7 Information acquisition positively impacts purchase intention of customers.
- H_8 Perceived quality positively impacts purchase intention of customers.
- H_9 Perceived quality has association with information acquisition.
- H_{10} Seller created information has association with Electronic word of mouth.

ACKNOWLEDGMENT

We are thankful to Prof. Arif Vaseer, Mohammad Ali Jinnah University and Mr. Khurram Shahzad, Assistant Professor, Riphah International University for their guidance and support in the completion of research work.

REFERENCES

[1] Ahn , Y.-Y., Han, S., Kwak, H., Moon, S. and Jeong, H. (2007), "Analysis of topological characteristics of huge online social networking services", Proceedings of the 16th International Conference on World Wide Web, Banff,ACMPress, New York, NY, pp. 835-44.

[2] Amato, Paul R. and Ruth Bradshaw (1985), "An Exploratory Study of People's Reasons for Delaying or Avoiding Helpseeking," Australian Psychologis, 20 (March), 21-31.

[3] Bansal, H. S., Irving, G. P., & Taylor, S. F. (2004). A three-component model of customer commitment to service providers. Journal of the Academy of Marketing Science, 32(3), 234−250.

[4] Blackshaw, P. and Nazzaro, M. (2004). "Consumer-Generated Media (CGM) 101: Word-of-mouth in the age of the Web fortified consumer", at http://www.nielsenbuzzmetrics.com/whitepapers (accessed: December 11, 2008

[5] Boone, L. E., & Kurtz, D. L. (2007). Contemporary marketing (13th ed.). Mason, OH: Thomson/South-Western.

[6] C. Anderson. The Long Tail: Why the Future of Business Is Selling Less of More. Hyperion, July 2006.

[7] C. Sang-Hun. To outdo Google, Naver taps into Korea's collective wisdom. International Herald Tribune, July 4 2007.

[8] Catherine Dwyer, Starr Hiltz & Katia Passerini (2007), "Trust and Privacy Concern Within Social Networking Sites: A Comparison of Facebook and MySpace", Americas Conference on Information Systems (AMCIS).

[9] Danny Meadows-klue (2008), "Falling in Love 2.0: Relationship Marketing for the Facebook Generation", Journal of Direct, Data and Digital Marketing Practice, 245-250.

[10] Della Lindsay J, Eroglu Dogan, Bernhardt Jay M, Edgerton Erin, Nall Janice. Looking to the future of new media in health marketing: deriving propositions based on traditional theories. Health Mark Q.2008;25(1-2):147–74. doi: 10.1080/07359680802126210. [Cross Ref]

[11] Elif Akagun Ergin & Handan Ozdemir Akbay (2010), "Consumers' Purchase Intentions for Foreign Products: An Empirical Research Study in Istanbul, Turkey", 2010 EABR & ETLC Conference Proceedings Dublin, Ireland.

[12] Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. Journal of Computer-Mediated Communication, 12(4), 1143–1168.

[13] Eugene, A., Carlos, C., Debora, D., Aristides, G., & Gilad, M. (2008), " Finding High-Quality Content in Social Media", Palo Alto, California, USA.

[14] Eysenbach Gunther. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res. 2008;10(3):e22. doi: 10.2196/jmir.1030.http://www.jmir.org/2008/3/e22/v10i3e22 [PMC free article] [PubMed] [Cross Ref]

[15] General Electric. (2008). Imagination at work: Our culture.Retrieved August 25, 2008, from http://www.ge.com/company/culture/index.html

[16] Gladwell, M. (2000). The Tipping Point: How Little Things Can Make a Big Difference, first published by Little Brown. ISBN 0-316-31696-2.

[17] Goh Tiong-Thye and Yen-Pei Huang. (2009). Monitoring youth depression risk in Web 2.0. The journal of information and knowledge management systems Vol. 39 No. 3, 2009 pp. 192-202.

[18] Hassan Siddique, (2011, Jan 08) Social networking sites become advertising hub. Retrieved January 13, 2011, from http://pakistantoday.com.pk/pakistan-news/Lahore/08-Jan-2011/Social-networking-sites-become-advertising-hub

[19] Hean Tat Keh &Yi Xie, (2009), "Corporate reputation and customer behavioral intentions: The roles of trust, identification and commitment", Industrial Marketing Management , 732–742.

[20] Hitwise (2007), "Social networking visits increase 11.5 percent from January to February", available at: http://hitwise.com/press-center/hitwiseHS2004/socialnetworkingmarch07.php (accessed 19 July 2007).

[21] Holak, Susan, Donald R. Lehmann, and Fareena Sultan (1987), "The Role of Expectations in the Adoption of Innovative Durables: Some Preliminary Results,"Journal of Retailing, 63 (Fall), 243-259.

[22] Horrigan J. Wireless Internet use. Washington, DC: Pew Internet & American Life Project; 2009. Jul 22, [2009-07-31]. webcite http://pewinternet.org/Reports/2009/12-Wireless-Internet-Use.aspx.

[23] Hrsky, Dan (1990), "A Diffusion Model Incorporating Product Benefits, Price, Income, and Information," Marketing Science, 9 (Fall), 342-365.

[24] Huffaker, D. (2006), Teen Blogs Exposed: The Private Lives of Teens Made Public, American Association for the Advancement of Science, St. Louis, MO.

[25] Jerry Wind & Vijay Mahajan (2001), "Convergence Marketing: Strategies for Reaching the New Hybrid Consumer", ISBN:0130650757.

[26] Jerry Wind and Vijay Mahajan (2001). Convergence Marketing: Strategies for Reaching the New Hybrid Consumer. ©2001 ISBN:0130650757

[27] Kaplan A. M., Haenlein M., (2010), "Users Of The World, Unite! The Challenges And Opportunities Of Social Media", Business Horizons, Vol.53, Issue 1.

[28] Kirmani, A. and Baumgartner, H. (2000), "Reference Points Used in Quality and Value Judgments", Marketing Letters, Vol. 11, No. 4, pp. 299-310.

[29] Lefebvre RC. The new technology: the consumer as participant rather than target audience.Social Marketing Quarterly. 2007;13(3):31–42. doi: 10.1080/15245000701544325. [Cross Ref]

[30] Lenhart, A., & Madden, M. (2007). Teens, privacy & online social networks: How teens manage their online identities and personal information in the age of MySpace. Washington, DC: Pew Internet & American Life Project.

[31] Mangold W. Glynn a, Faulds David J. (2009). Social media: The new hybrid element of the promotion mix. Business Horizons (2009) 52, 357—365

[32] Mittal, V., Kumar, P., & Tsiros, M. (1999). Attribute-level performance, satisfaction, and behavioral intentions over time. Journal of Marketing, 63(2), 88−101.

[33] Nancy Atkinson, Wen-ying Sylvia Chou, Yvonne M Hunt, Ellen Burke Beckjord, Richard P Moser, and Bradford W Hesse. (2009). "Social Media Use in the United States: Implications for Health Communication". J Med Internet Res. 2009 Oct–Dec; 11(4): e48.

[34] Nardi, B.A., Schiano, D.J. and Gumbrecht, M. (2004), "Blogging as social activity, or, would you let 900 million people read your diary?", Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, Chicago, IL, ACM Press, New York, NY, pp. 222-31.

[35] Nielsen//NetRatings (2006, May 11). Social networking sites grow 47 percent, year over year, reaching 45 percent of web users, according to Nielsen//NetRatings. Retrieved August 30, 2007, from http://www.nielsen-netratings.com/pr/pr_060511.pdf

[36] Pakistan Telecommunication Authority http://www.pta.gov.pk/index.php?cur_t=vtext&option=com_content&task=view&id=583&Itemid=1&catid=95 , Retrieved on January 20, 2011.

[37] Pascu, C., Osimo, D., Turlea, G., Ulbrich, M., Punie, Y. and Burgelman, J-C. (2008), ''Social computing: implications for the EU innovation landscape'', Foresight, Vol. 10 No. 1, pp. 37-52/

[38] Pempek Tiffany A., Yevdokiya A. Yermolayeva and Sandra L. Calvert. (2009). College students' social networking experiences on Facebook. Journal of Applied Developmental Psychology 30 (2009) 227–238.

[39] Procter and Gamble. (2008). Purpose, values, and principles. Retrieved August 25, 2008, from http://www.pg.com/company/who_we_are/ppv.shtml

[40] Rao, A. R. and Monroe, K. B. (1989), "The Effect of Price, Brand Name, and Store Name on Buyers' Perceptions of Product Quality: An Integrative Review", Journal of Marketing Research, Vol. 26, pp. 351-357.

[41] Robert, A. P., Sridhar, B., & Bronnenberg, B. J. (1997). Exploring the implications of the internet for consumer marketing. Journal of the Academy of Marketing Science .

[42] SCOTT David (2007), "The new rules of marketing and pr : how to use news releases, blogs, podcasting, viral marketing and online media to reach buyers directly".

[43] Suziki, L., & Calzo, J. (2004). The search for peer advice in cyberspace: An examination of online teen bulletin boards about health and sexuality. Journal of Applied Developmental Psychology, 25, 685–698.

[44] Tiffany A. Pempek, Yevdokiya A. Yermolayeva & Sandra L. Calvert (2009), "College students' social networking experiences on Facebook", Journal of Applied Developmental Psychology, 227–238.

[45] Tiong-Thye Goh & Yen-Pei Huang (2009), "Monitoring youth depression risk in Web 2.0", The journal of information and knowledge management systems Emerald Group Publishing Limited, 192-202.

[46] Toivonen, S. (2007), Web on the Move. Landscapes of Mobile Social Media, VTT Research Notes 2403, Espoo.

[47] Toni Ahlqvist, Asta Ba¨ck, Sirkka Heinonen & Minna Halonen (2009), "Road-mapping the societal transformation potential of social media", Emerald Group Publishing, 3-26.

[48] Turney, P.D. and Littman, M.L. (2003), "Measuring praise and criticism: inference of semantic orientation from association", ACM Trans. Inf. Syst., Vol. 21 No. 4, pp. 315-46.

[49] W. Glynn Mangold & David J. Faulds (2009), "Social media: The new hybrid element of the promotion mix", Business Horizons, Science Direct, 357-365.

[50] Wikipedia, http://en.wikipedia.org/wiki/Telecommunications_in_Pakistan, Retrieved on January 20, 201.

[51] Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The behavioral consequences of service quality. Journal of Marketing, 60(2), 31−46.

[52] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[53] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[54] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[55] K. Elissa, "Title of paper if known," unpublished.

[56] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[57] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[58] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

AUTHORS PROFILE

Mahmood Rehmani is a student of Master of Science with major in Marketing at Department of Management Sciences, Mohammad Ali Jinnah University, Islamabad, Pakistan. He is also working as Public Relation Officer with Pakistan Engineering Council. Major duties include handling newsletters, E-newsletters, online groups, facebook fan page, twitter and production of audio video material for better corporate image of Pakistan Engineering Council. Mahmood is President of Islamabad Chapter of Pakistan Council for Public Relations and also voluntarily providing Media Consultancy services to Knowledge & Information Management Society (KIMS), Human Advancement Network for the Deprived (HAND) and Women In Energy-Pakistan.

Muhammad Ishfaq Khan is supervisor and faculty member at Department of Management Sciences, Mohammad Ali Jinnah University, Islamabad. He has been teaching since 2003. During his teaching tenure, Mr. Khan delivered lectures in different degree awarding institutions like COMSATS Institute of Information Technology, Islamabad Campus, Quaid-e-Azam University, Islamabad, Mohammad Ali Jinnah University, Islamabad Campus, Riphah International University, Islamabad. Mr. Khan has submitted his research thesis at COMSATS Institute of Information Technology, Islamabad. Before this, He has completed his Master of Science in Management Sciences from COMSATS Institute of Information Technology, Islamabad. In addition, He has done his Master of Business Administration from Mohammad Ali Jinnah University, Islamabad Campus. Furthermore, Mr. Khan has received a Post Graduate Diploma in Computer Science from International Islamic University, Islamabad. Mr. Khan has more than 20 research publications in peer reviewed journals, national and international conferences. (Boston, Florida, Singapore, Japan, Hong Kong, Malaysia, Spain, New York, Australia and Pakistan.

Feed Forward Neural Network Based Eye Localization and Recognition Using Hough Transform

Shylaja S S*, K N Balasubramanya Murthy, S Natarajan
Nischith, Muthuraj R, Ajay S
Department of Information Science and Engineering,
P E S Institute of Technology,
Bangalore-560085, Karnataka, INDIA
*shylaja.sharath@pes.edu

***Abstract*— Eye detection is a pre-requisite stage for many applications such as face recognition, iris recognition, eye tracking, fatigue detection based on eye-blink count and eye-directed instruction control. As the location of the eyes is a dominant feature of the face it can be used as an input to the face recognition engine. In this direction, the paper proposed here localizes eye positions using Hough Transformed (HT) coefficients, which are found to be good at extracting geometrical components from any given object. The method proposed here uses circular and elliptical features of eyes in localizing them from a given face. Such geometrical features can be very efficiently extracted using the HT technique. The HT is based on a evidence gathering approach where the evidence is the ones cast in an accumulator array. The purpose of the technique is to find imperfect instances of objects within a certain class of shapes by a voting procedure. Feed forward neural network has been used for classification of eyes and non-eyes as the dimension of the data is large in nature. Experiments have been carried out on standard databases as well as on local DB consisting of gray scale images. The outcome of this technique has yielded very satisfactory results with an accuracy of 98.68%**

Keywords - ***Hough Transform; Eye Detection; Accumulator Bin; Neural Network;***

I. INTRODUCTION

Face recognition has become a popular area of research in computer science and one of the most successful applications of image analysis and understanding. There are a large number of commercial, securities, and forensic applications requiring the use of face recognition technologies. These applications include face reconstruction, content-based image database management and multimedia communication. Early face recognition algorithms used simple geometric models, but the recognition process has now matured into a science of sophisticated mathematical representations and matching processes.

Major advancements and initiatives in the past ten to fifteen years have propelled face recognition technology into the spotlight. Face recognition can be used for both verification and identification. Eye detection and localization have played an important role in face recognition over the years. Evidence gathering approach of Hough Transform has a major role to play in this field. Eye detection is the process of tracking the location of human eye in a face image. Previous approaches use complex techniques like neural network, Radial Basis Function networks, Multi-Layer Perceptron's etc. In the current paper human eye is modeled as a circle (iris), the black circular region of eye enclosed inside an ellipse (eye-lashes).Due to the sudden intensity variations in the iris with respect the inner region of eye-lashes the probability of false acceptance is very less. Since the image taken is a face image the probability of false acceptance further reduces. Since Hough transform has been effective at detecting shapes, and due to the circular shape of the iris, we have employed it in the eye detection and localization. The points obtained after applying Hough transform are fed to Back Propagation Neural Network (BPNN).

BPNN is a supervised method where desired output is calculated given the input and the propagation involves training input patterns through the neural network to generate the output activations. This is followed by Back propagation of the output activations through the Neural Network using the training patterns in order to generate the output activations and the hidden neurons. The back propagation used here is feed forward neural network as they can handle the high dimensional data very well.

II. HOUGH TRANSFORM

The Hough Transform, HT, has long been recognized as a technique of almost unique promise for shape and motion analysis in images containing noisy, missing and extraneous data. A good shape recognition system must be able to handle complex scenes which contain several objects that may partially overlap one another. In the case of computer vision systems the recognition algorithm must also be able to cope with difficulties caused by poor characteristics of image sensors and the limited ability of current segmentation algorithms. The Hough Transform is an algorithm which has the potential to address these difficult problems. It is a likelihood based parameter extraction technique in which pieces of local evidence independently vote for many possible instances of a sought after model. In shape detection the model captures the relative position and orientation of the constituent points of the shape and distinguishes between particular

instances of the shape using a set of parameters. The HT identifies specific values for these parameters and thereby allows image points of the shape to be found by comparison with predictions of the model instantiated at the identified parameter values

Hough transform is a general technique for identifying the locations and orientations of certain types of features in a digital image and used to isolate features of a particular shape within an image.

Lam and Yuen [18] noted that the Hough transform is robust to noise, and can resist to a certain degree if occlusion and boundary effects. Akihiko Torii and Atsushi Imiya [19] proposed a randomized Hough transform based method for the detection of great circles on a sphere. Cheng Z. and Lin Y [20] proposed a new efficient method to detect ellipses in gray-scale images, called Restricted Randomized Hough transform. The key of this method is restricting the scope of selected points when detecting ellipses by prior image processing from which the information of curves can be obtained. Yip et al. [21] presented a technique aimed at improving the efficiency and reducing the memory size of the accumulator array of circle detection using Hough transform.

P.Wang [12] compares fully automated eye localization and face recognition to the manually marked tests. The recognition results of those two tests are very close, e.g. 83.30% vs. 81.76% for experiment 1 and 97.04% vs. 96.38% for experiment 2. The performance of the automatic eye detection is validated using FRGC 1.0 database. The validation has an overall 94.5% eye detection rate, with the detected eyes very close to the manually provided eye positions.

Zhiheng Niu et al. [13] introduced a framework of 2D cascaded AdaBoost for eye localization. This framework can efficiently deal with tough training set with vast and incompact positive and negative samples by two-direction bootstrap strategy. And the 2D cascaded classifier with cascading all the sub classifiers in localization period can also speed up the procedure and achieve high accuracy. The method is said to be very accurate, efficient and robust under usual condition but not under extreme conditions.

Mark Everingham et al. [14] investigated three approaches to the eye localization task, addressing the task directly by regression, or indirectly as a classification problem. It has yielded accuracy up to 90%.

Hough transform is used for circle and ellipse detection. Hough transform was the obvious choice because of its resistance towards the noise present in the image. Compared to the aforementioned models the proposed model is simple and efficient. The proposed model can further be improved by including various features like orientation angle of eye-lashes (which is assumed constant in the proposed model), and by making the parameters adaptive.

Hough Transform implementation defines a mapping from image points into a accumulator space. The mapping is achieved in a computationally efficient manner based on the function that describes the target shape. This mapping requires much fewer computational resources than template matching technique. Hough Transform also has the ability to tolerate occlusion and noise.

The set of all straight lines in the picture plane constitutes two-parameter family. If we fix a parameterization for the family, then an arbitrary straight line can be represented by a single point in the parameter space. For reasons that become obvious, we prefer the so-called normal parameterization. As illustrated in, this parameterization specifies a straight line by the angle α of its normal and its algebraic distance p from the origin. The equation of a line corresponding to this geometry is

$$x\cos\alpha + y\cos\alpha = \rho \qquad (1)$$

If we restrict α to the interval [0, π), then the normal parameters for a line are unique. With this restriction, every line in the x-y plane corresponds to a unique point in the α- p plane.

Suppose, now, that we have some set { (x_1,y_1),..., (x_n,y_n) } of figure points and we want to find a set of straight lines that fit them. We transform the points (x_i, y_i) into the sinusoidal curves in the e-plane defined by

$$\rho = \chi_i \cos\alpha + y_i \sin\alpha \qquad (2)$$

It is easy to show that the curves corresponding to collinear figure points have a common point of intersection. This point in the α -p plane, say (α_0,p_0), defines the line passing through the collinear points. Thus, the problem of detecting collinear points can be converted to the problem of finding concurrent curves.

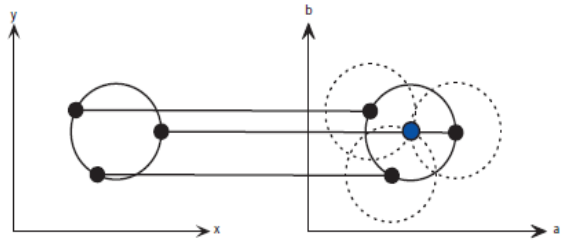


Figure 1. Circular Hough transform for x-y space to the parameter space for constant radius.

Rest of the paper is organized as follows: section III describes the proposed methodology used in this paper, experimental results are illustrated in section IV and section V & VI focus on conclusions & suggestions for further work.

III. PROPOSED METHOD

The method has 3 stages: Preprocessing, feature extraction by Hough transform followed by classification by Feed Forward Neural Network.

A. Preprocessing

The database images are initially cropped to a size of 100 x 100 resolutions. This will help us to fix the eye coordinates in a convenient and simple manner.

B. Feature Extraction by Hough Tramsfrom.

HT being the most efficient techniques is used to identify positions of arbitrary shapes most commonly circles and ellipses. The purpose of this technique is to find imperfect instances of objects in a parameter space within a certain class of shapes by a voting procedure. This voting procedure is carried out in a parameter space, from which object candidates are obtained as local maxima in a so-called accumulator space that is explicitly constructed by the algorithm for computing the Hough transform. The local maxima obtained from the accumulator array are used in training of back propagation neural network.

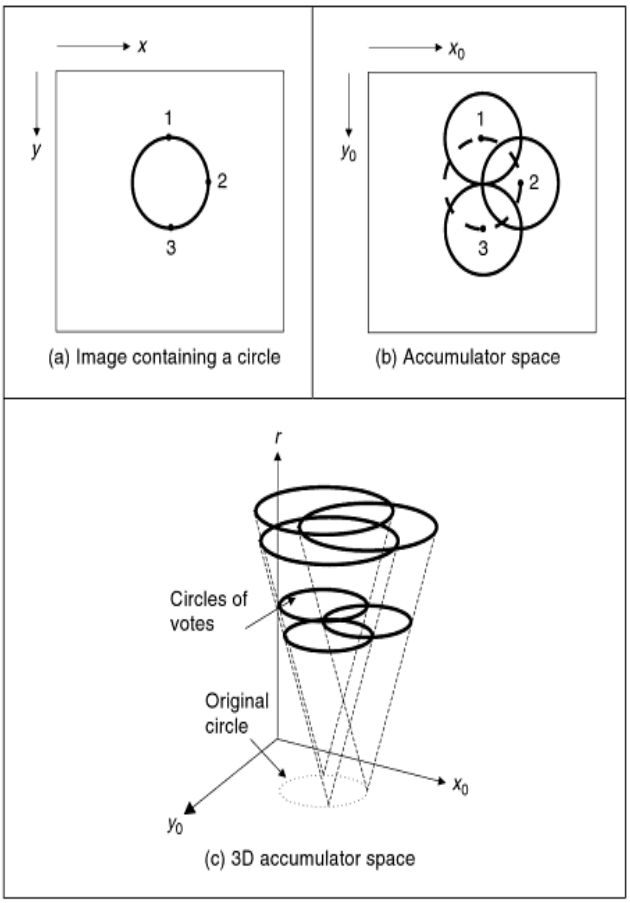


Figure 2. (a) x-y space (b) Accumulator space (c) 3D accumulator space.

For each input image in the x-y plane, a spherical filter is applied and the outcome of this is a set of points in the parameter space. These points represent the centers of probabilistic circles of pre-defined radii in the x-y plane. Mapping is done for each point in the original space to a corresponding point in the original space to a corresponding point in accumulator space and increment the value of the accumulator bin. Based on a pre-defined threshold, local maxima is identified which when back mapped on to a original space corresponds to a centre of probabilistic circle. The circle pattern is described by this equation

$$(x_p - x_0)^2 + (y_p - y_0)^2 = r^2 \qquad (3)$$

C. Feed Forward Neural Network (FFNN)

The Back propagation neural network used here is Feed Forward Neural Network (FFNN). FFNN was the first and arguably simplest type of artificial neural network devised. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes.

The number of Input Layers, hidden layers and output layers are adjusted to fit the data points to the curve. During the training phase the training data in the accumulator array is fed into the input layer. The data is propagated to the hidden layer and then to the output layer. This is the forward pass of the back propagation algorithm. Each node in the hidden layer gets input from all the nodes from input layer which are multiplexed with appropriate weights and summed. The output of the hidden node is the nonlinear transformation of the resulting sum. Similar procedure is carried out in the output layer. The output values are compared with target values and the error between two is propagated back towards the hidden layer. This is the backward pass of the back propagation algorithm. The procedure is repeated to get the desired accuracy. During the testing phase the test vector is fed into the input layer. The output of FFNN is compared with training phase output to match the correct one. This can serve as a need to do recognition of facial objects.

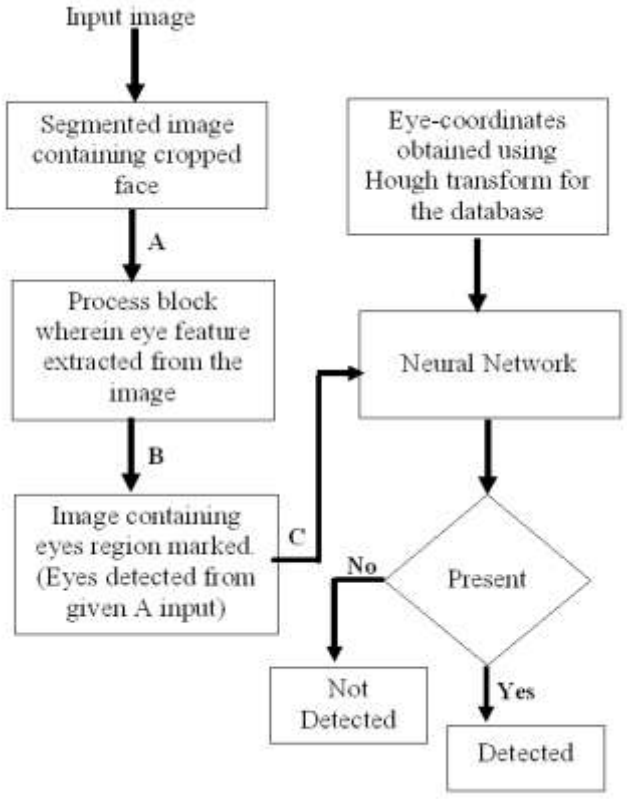


Figure 3. Block Diagram of the entire process

Note: A indicates segmentation of feature extracted face. B indicates pixel location of eye coordinates and C indicates vector (X1, Y1, X2, Y2) where (X1, Y1) and (X2, Y2) are the eye coordinates.

IV. EXPERIMENTAL RESULTS

The algorithm has been extensively tested on two standard database namely Yale, BioID and a local database. The accumulator array from the Circular Hough transform has the same dimension as the input image which is of dimension 100x100. A filtering technique is used in the search of local maxima in the accumulation array and it uses a threshold value of 0.5 set using brute force technique. The local maxima obtained are represented by peaks in the 3D view of figure 5. The technique has yielded satisfactory results for some extreme conditions and has failed under certain conditions as shown in figure 6. The failure in figure 6 can be attributed to local maxima being present at a different location because of the extreme illumination effects.

A feed forward neural network with 200 hidden layers was trained and the regression plot is as shown in figure 10 and performance plot is as shown in figure 11. The x-coordinate, y-coordinate of the regression plot represents target requirement and output=99*target+0.0029 respectively. The plot indicates a perfect fit of the data and the accuracy obtained is very high with a computed value of 98.68%.

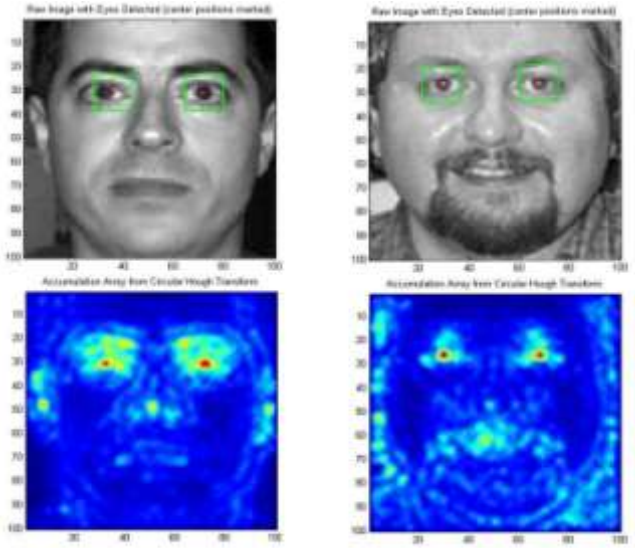


Figure 4. Eye Detections and Accumulator

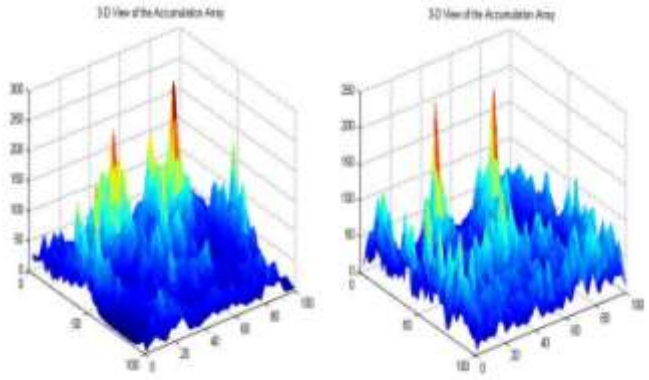


Figure 5. 3D view of accumulator array

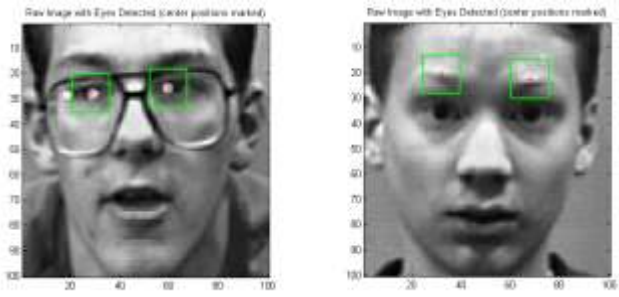


Figure 6. positive sample with occlusion and a negative sample

Figure 7. Database Images

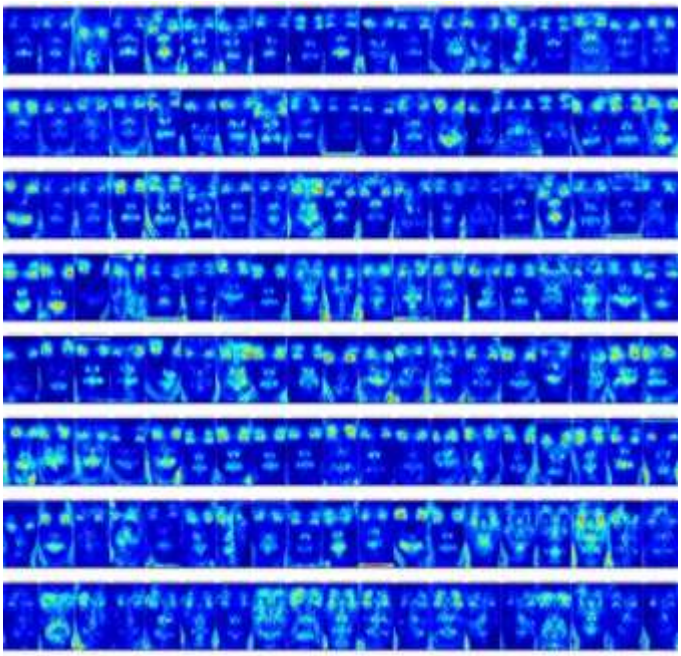


Figure 8. Accumulator Bin

Figure 9. Eye localized images (green rectangle around eyes :Output)

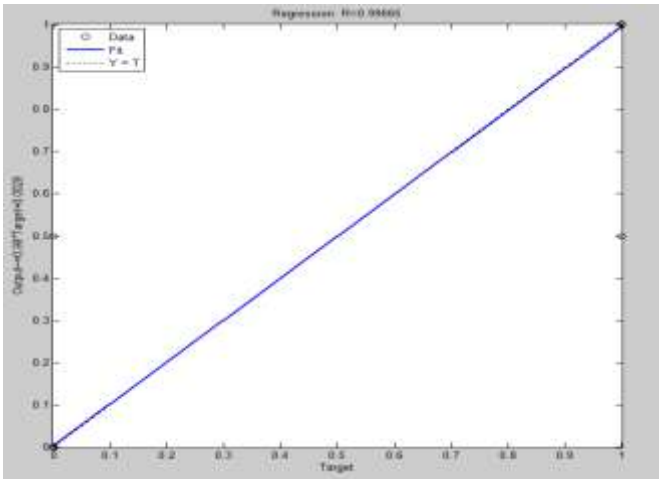


Figure 10. Regression Plot

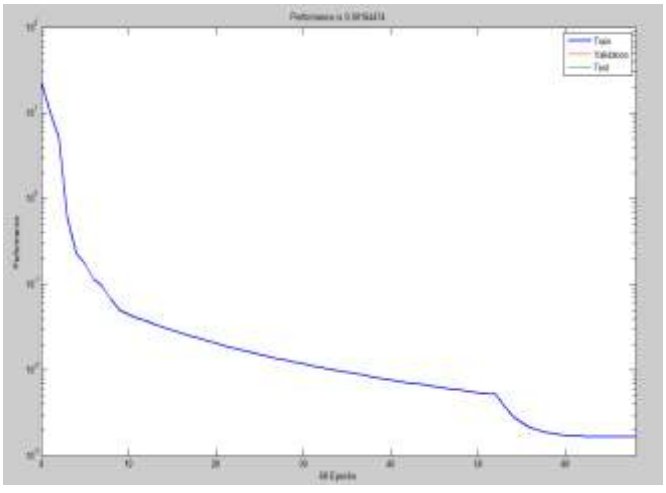


Figure 11. Performance Plot of Neural Network

TABLE I. TEST RESULTS

Database	No. of Images	No. of detections	Neural Network Results	Faulty Recognition	% Accuracy
Standard DBs	25	25	25	0	100
Built DB	127	127	125	2	98.42
Total	152	152	150	2	98.68

V. CONCLUSIONS

In this paper, we have presented a human eye localization algorithm for faces with frontal pose and upright orientation using Hough transform. This technique is tolerant to the presence of gaps in the feature boundary descriptions and is relatively unaffected by image noise. In addition Hough transform provides parameters to reduce the search time for finding lines based on set of edge points and these parameters can be adjusted based on application requirements.

The voting procedure has been successfully applied in the current method and is able to yield better results to effectively localize eyes in a given set of input images. The component can be effectively used as an input to the face recognition engine for recognizing faces with less number of false positives.

VI. FURTHER WORK

Hough Transform requires a large amount of storage and high cost in computation. To tackle these many improvements can be worked on. In addition, adaptable neural networks are coming in place which can probably save time and space complexities in computation. For face recognition like applications, future work can address to get other features under extreme conditions of images. They can be nodal points like width of nose, cheekbones, jaw line etc., which are sufficient for recognition of a particular face, within a given set of images.

ACKNOWLEDGMENT

The authors would like to extend their sincere gratitude to PES Management in all support provided for carrying out this research work.

REFERENCES

[1] Ballard DH, “Generalizing the Hough transform to detect arbitrary shapes”, In the proceedings of Pattern Recognition, pp. 111–122, 1981.

[2] J. Illingworth and J. Kittler. “A survey of the Hough Transform”, In the proceedings of International Conference on Computer Vision, Graphics and Image Processing, pp. 87–116, 1988.

[3] JR Bergen and H Shvaytser, “A probabilistic algorithm for computing Hough Transforms”, In the Journal of Algorithms, pp.639–656, 1991.

[4] N Kiryati, Y Eldar, and AM Bruckstein, “A probabilistic Hough Transform”, In the proceedings of Pattern Recognition, pp. 303–316, 1991.

[5] Duda RO, Hart PE, “Use of the Hough Transform To Detect Lines and Curves In Pictures”, In the Communications of ACM, pp. 11–15, 1972.

[6] Kimme, C., D. Ballard and J. Sklansky, “Finding Circles by an Array of Accumulators”, In the Communications of ACM, pp. 120–122, 1975.

[7] Mohamed Roushdy “Detecting Coins with Different Radii based on Hough Transform in Noisy and Deformed Image**”** In the proceedings of GVIP Journal, Volume 7, Issue 1, April, 2007.

[8] Igor Aizenberg, Claudio Moraga, Dmitriy Paliy ”**A** Feed-forward Neural Network based on Multi-Valued Neurons”, In the Proceedings of

Computational Intelligence, Theory and Applications**.** Advances in Soft Computing, Springer XIV, pp. 599-612, 2005.

[9] Eric A. Plummer "Time Series Forecasting With Feed-Forward Neural Networks: Guidelines And Limitations" A Thesis from Department of Computer Science, The Graduate School of The University of Wyoming.

[10] Sybil Hirsch, J. L.Shapiro, Peter I.Frank. "The Application of Neural Network System Techniques to Asthma Screening and Prevalence Estimation". In the *Handbook of Computational Methods in Biomaterials, Biotechnology, and Biomedical Systems*. Kluwer, 2002.

[11] Y. Liu, J. A. Starzyk, Z. Zhu, "Optimized Approximation Algorithm in Neural Network without Overfitting", In the Proceedings of IEEE Transactions on Neural Networks, vol. 19, no. 4, June, 2008, pp. 983-995.

[12] P.Wang, M. Green, Q. Ji, and J.Wayman., "Automatic Eye Detection and Its Validation", In the proceedings of *IEEE Conference on. Computer Vision and Pattern Recognition*, 2005.

[13] Z. Niu, S. Shan, S. Yan, X. Chen, and W. Gao. "2D Cascaded AdaBoost for Eye Localization". In the *Proceedings. of the 18th International Conference on Pattern Recognition*, 2006.

[14] M.R. Everingham and A. Zisserman, "Regression and Classification Approaches to Eye Localization in Face Images", In the *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FG2006)*, pp. 441-448, 2006.

[15] P. Campadelli, R. Lanzarotti, and G. Lipori. "Face localization in color images with complex background", In the proceedings of IEEE International Workshop on Computer Architecture for Machine Perception, pp. 243–248, 2005.

[16] Cheng Z. and Lin Y., "Efficient technique for ellipse detection using restricted randomized Hough transform", In the proceedings of the International Conference on Information Technology (ITCC'04), Vol.2, pp.714-718, 2004.

[17] Yu Tong, Hui Wang, Daoying Pi and Qili Zhang, "Fast Algorithm of Hough Transform- Based Approaches for Fingerprint Matching", In the Proceedings of the Sixth World Congress on Intelligent Control and Automation, Vol. 2, pp.10425- 10429, 2006.

[18] Lam, W.C.Y., Yuen, S.Y., "Efficient technique for circle detection using hypothesis filtering and Hough transform", IEE Proceedings of Vision, Image and Signal Processing, vol. 143, Oct. 1996.

[19] Torii, A. and Imiya A., "The Randomized Hough Transform Based Method for Great Circle Detection on Sphere", In the Proceedings of Pattern Recognition, pp. 1186-1192, 2007.

[20] Cheng Z. and Lin Y., "Efficient technique for ellipse detection using restricted randomizedHough transform", Proc. of the International Conference on Information Technology(ITCC'04), Vol.2, pp.714-718, 2004.

[21] Yip, R.K.K.; Leung, D.N.K. and Harrold,S.O., "Line segment patterns Hough transform for circles detection using a 2-dimensional array", International Conference on Industrial Electronics, Control, and Instrumentation, Vol.3, pp. 1361-1365, 1993.

[22] Canny, J., "A Computational Approach to Edge Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence,Vol.8, No. 6, Nov. 1986.

[23] Qi-Chuan Tian; Quan Pan; Yong-Mei Cheng and Quan-Xue Gao, "Fast algorithm and application of Hough transform in iris segmentation", Proceeding of International Conference on Machine Learning andCybernetics, Vol. 7, pp.3977-3980, 2004.

[24] Trivedi, J. A. (2011). Framework for Automatic Development of Type 2 Fuzzy , Neuro and Neuro-Fuzzy Systems. International Journal of Advanced Computer Science and Applications - IJACSA, 2(1), 131-137.

[25] Deshmukh, R. P. (2010). Modular neural network approach for short term flood forecasting a comparative study. International Journal of Advanced Computer Science and Applications - IJACSA, 1(5), 81-87.

[26] Nagpal, R., & Nagpal, P. (2010). Hybrid Technique for Human Face Emotion Detection. International Journal of Advanced Computer Science and Applications - IJACSA, 1(6).

[27] Khandait, S. P. (2011). Automatic Facial Feature Extraction and Expression Recognition based on Neural Network. International Journal of Advanced Computer Science and Applications - IJACSA, 2(1), 113-118.

AUTHORS PROFILE

Shylaja S S is currently a research scholar pursuing her Doctoral Degree at P E S Institute of Technology. She has vast teaching experience of 22 years and 3 years in research. She has published 13 papers at various IEEE / ACM conferences & Journals

K N Balasubramanyamurthy is Principal and Director of P E S Institute of Technology. He is the executive council member of VTU. He has over 30 years of experience in teaching and 18 years in research. He has published more than 60 papers at various IEEE / ACM conferences & Journals

S Natarajan is Professor in Department of Information Science and Engineering P E S Institute of Technology. He has 34 years of experience in R & D and 12 years of experience in teaching. He has published 31 papers at various IEEE / ACM conferences & Journals

Nischith, Muthuraj R and Ajay S are currently pursuing their B.E in Information Science & Engineering at P E S Institute of Technology. They are passionate towards research in Computer Vision and Image Processing.

Computer Aided Design and Simulation of a Multiobjective Microstrip Patch Antenna for Wireless Applications

Chitra Singh[1*] and R. P. S. Gangwar[2]
[1,2]Department of Electronics & Communication Engineering, College of Technology,
G.B.Pant Univ. of Agric. & Tech.,
Pantnagar, Uttarakhand, India.
*chitra.singh88@gmail.com

***Abstract*— The utility and attractiveness of microstrip antennas has made it ever more important to find ways to precisely determine the radiation patterns of these antennas. Taking benefit of the added processing power of today's computers, electromagnetic simulators are emerging to perform both planar and 3D analysis of high-frequency structures. One such tool studied was IE3D, which is a program that utilizes method of moment. This paper makes an investigation of the method used by the program, and then uses IE3D software to construct microstrip antennas and analyze the simulation results. The antenna offers good electrical performance and at the same time preserves the advantages of microstrip antennas such as small size, easy to manufacture as no lumped components are employed in the design and thus, is low cost; and most importantly, it serves multiple wireless applications.**

Keywords- Electromagnetic simulation; microstrip antenna; radiation pattern; IE3D;

I. INTRODUCTION

Microstrip antennas have become increasingly popular for microwave and millimeter wave applications, because they offer several distinct advantages over conventional microwave antennas. These advantages include small size, easy to fabricate, lightweight, and conformability with the hosting surfaces of vehicles, aircraft, missiles, and direct integration with the deriving electronics. Microstrip antennas are assigned different names, such as, printed antennas (can be printed directly onto a circuit board), planar antennas, microstrip patch antennas or simply microstrip antennas (MSA). MSA in general consists of radiating conducting patch, a conducting ground plane, a dielectric substrate sandwiched between the two, and a feed connected to the patch through the substrate. A patch radiates from fringing fields around its edges as shown in Fig. 1. As can be seen in the figure, the electric field does not stop abruptly near the patch's edges: the field extends beyond the outer periphery.

These 'field extensions' are known as 'fringing fields' and cause the patch to radiate.

Their utility and increasing demand makes it necessary to develop a precise method of analyzing these types of antennas. Several theoretical methods of determining far field patterns have been formulated, such as the aperture method, the cavity method, modal method and the transmission line method [1-3]. Although these methods are capable of providing a physical insight, are not as accurate as some numerical techniques can be because all these methods contain approximations. As full-wave solutions are more accurate and applicable to many structures, some researchers begin to study the MoM-based solution for MSAs, such as [4-13] in recent years. The integral equation formulation solved using the method of moments (MoM) [14–16] is known as an efficient tool. The application of the MoM to a planar configuration results in a dense coupling matrix. This matrix can be considered as the coefficients matrix of a few systems of linear equations which are required to be solved. With the increasing capability of today's computers, numerical techniques offer a more accurate reflection of how microstrip antennas actually perform. Their use reduces the need to modify the final dimensions using a knife to remove metal or metal tape to increase the patches.

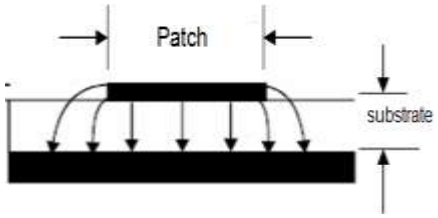


Fig. 1 A patch radiating from fringing fields around its edges

The MoM modeling program used in this paper is IE3D full wave EM solution of Zeland simulation software. IE3D is the computational power-house that when given an input file performs all the calculations necessary to obtain the fields and/or currents at each grid point in the model. Full wave EM simulation refers to the solution of Maxwell's or differential equations governing macro electromagnetic phenomenons, without assuming frequency at DC or any other assumptions. Full-wave EM simulations are very accurate from very low frequency to very high frequency.

The paper is organized in the following way. Section II presents an explanation of the method of moments as a solution to the integral equations. In Section III, the proposed microstrip antenna designing procedure and its geometry is described and discussed. In Section IV, the IE3D full wave EM solution of Zeland simulation software based on MoM modeling program is employed to construct a unique microstrip patch antenna, and

simulation results are presented and discussed. Finally, conclusions are drawn in Section V.

II. General Description of the Method of Analysis Used

The integral equations are formulated with a full dyadic Green's function and the matrix elements are computed completely numerically in the spatial domain [3]. IE3D can model truly arbitrary 3D metal structures. The method of moments (MOM) [17] expands the currents on an antenna (or scattering object) in a linear sum of simple basis functions. The approximate solution is a finite series of these basis functions [18]:

$$f_a = \sum_{i=1}^{N} a_i f_i \qquad \text{-(1)}$$

Coefficients are obtained by solving integral equations satisfying boundary conditions on the surface of the antenna (or object). The integral equation can be expressed in the form $L f_a = g$, where L is a linear operator, usually a scalar product using an integral, f_a the unknown currents given by Eq. (1), and g the known excitation or source function. The summation of Eq. (1) is substituted into the linear operator equation and scalar product integral is used to calculate the terms in a matrix equation. The solution of the matrix equation determines the coefficients of current expansion. The MOM produces filled matrices. The art of the MOM is in choosing basis functions and deriving efficient expressions for evaluating the fields using the basis function currents. Common basis functions are simple staircase pulses, overlapping triangles, trigonometric functions, or polynomials. The method does not satisfy boundary conditions at every point, only over an integral average on the boundaries. By increasing the number of basis functions, the method will converge to the correct solution. Spending excessive time on the solution cannot be justified if it greatly exceeds our ability to measure antenna performance accurately using real hardware [18].

This approach is relevant to a general planar geometry in stack or multi-layer configuration. To highlight the main idea, general description for a conducting patch on top of a single dielectric layer with a ground plate on bottom is being presented, without loss of generality, as shown in Fig. 2.

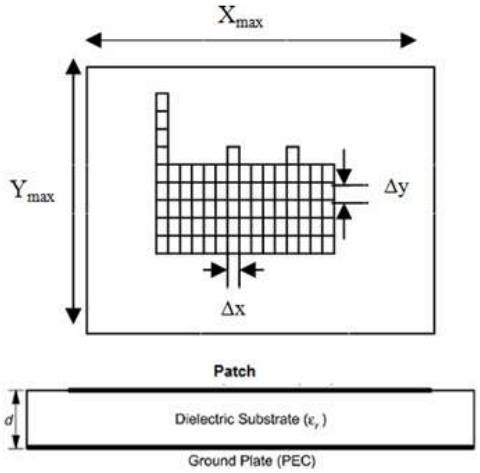


Fig. 2 MoM solution by subdividing the structure into a number of segments

Here ε_r is dielectric constant of the substrate and d is its thickness. Method of moments (MoM) is used to solve the problem and analyze the patch antenna performance. For brevity significant steps are presented here. As shown in Fig. 2, the structure is subdivided into a number of segments in first step. In second step, the unknown induced current on the structure is expanded in terms of known basis functions weighted by unknown amplitudes. The excitation current at the source locations is modeled using another set of basis functions [19].

$$J_x^i(x,y) = \sum_{n=1}^{N_x^i} A_n^i ET(x-a_n^i) R(y-b_n^i) \qquad \text{-(2)}$$

$$J_y^i(x,y) = \sum_{n=1}^{N_y^i} B_n^i ET(x-c_n^i) ET(y-d_n^i) \qquad \text{-(3)}$$

$$J_x^e(x,y) = \sum_{n=1}^{N_x^e} A_n^e OT(x-a_n^e) R(y-b_n^e) \qquad \text{-(4)}$$

$$J_y^e(x,y) = \sum_{n=1}^{N_y^e} B_n^e R(x-c_n^e) OT(y-d_n^e) \qquad \text{-(5)}$$

where the superscripts i and e mean induced and excitation currents, respectively. J_x and J_y are the x and y components of the current, respectively. N_x and N_y are the numbers of x and y directed basis functions. A_n and B_n are the amplitudes of the nth x and y directed basis functions, respectively. a_n and b_n are the x and y coordinates of the center of the nth x directed basis function. c_n and d_n are the x and y coordinates of the center of the nth y directed basis function, respectively. The even triangular, ET, odd triangular, OT, and rectangular, R, functions are defined as follows[19]:

$$ET(u) = \begin{cases} \dfrac{(u+\Delta_u)}{\Delta_u^2}; & -\Delta_u < u < 0 \\ \dfrac{(\Delta_u - u)}{\Delta_u^2}; & 0 < u < \Delta_u \\ 0; & otherwise \end{cases} \qquad \text{-(6)}$$

$$OT(u) = \begin{cases} -\dfrac{(u+\Delta_u)}{\Delta_u^2}; & -\Delta_u < u < 0 \\ \dfrac{(\Delta_u - u)}{\Delta_u^2}; & 0 < u < \Delta_u \\ 0; & otherwise \end{cases} \qquad \text{-(7)}$$

$$R(u) = \begin{cases} \frac{1}{\Delta_u}; & -\frac{\Delta_u}{2} < u < \frac{\Delta_u}{2} \\ 0; & otherwise \end{cases} \quad \text{-(8)}$$

where the variable u is replaced by either x or y, Δ_u is the width of the segment along the u direction. Boundary conditions are satisfied in third step of MoM solution procedure [18]. Assuming both, the patch and ground as perfect conductors, the tangential electric field should vanish. Expressing this boundary condition in terms of an integral equation and substituting the previous current expansion into this integral equation and applying Galerkin's testing, results in the following matrix equation:

$$\begin{bmatrix} [E_x] \\ [E_y] \end{bmatrix} = \begin{bmatrix} [Z^i_{xx}] & [Z^i_{xy}] \\ [Z^i_{yx}] & [Z^i_{yy}] \end{bmatrix} \begin{bmatrix} [A^i] \\ [B^i] \end{bmatrix} + \begin{bmatrix} [Z^e_{xx}] & [Z^e_{xy}] \\ [Z^e_{yx}] & [Z^e_{yy}] \end{bmatrix} \begin{bmatrix} [A^e] \\ [B^e] \end{bmatrix} = \begin{bmatrix} [0] \\ [0] \end{bmatrix}$$

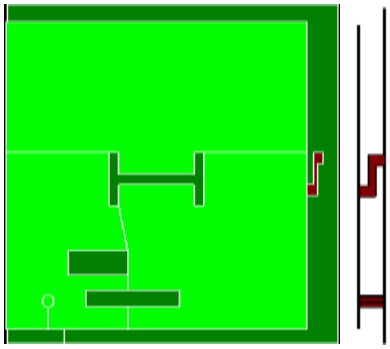
- (9)

where the elements of the vectors $[E_y]$ and $[E_x]$ are the total y and x components of the electric field at the test points. Thus, knowing the impedance submatrices, and the amplitudes of the excitation current $[A^e]$ and $[B^e]$, and, the unknown amplitudes of the induced current, $[A^i]$ and $[B^i]$ can be calculated. A number of distributions of the excitation current can be imposed and corresponding to each distribution, the induced current can be obtained from (9) by solving a system of linear equations. As a last step, a de-embedding technique is employed to obtain the S-parameters matrix which characterizes the structure. Some de-embedding techniques are given in [19]. Filling of the submatrices : $[Z^i_{xx}]$, $[Z^i_{xy}]$,$[Z^i_{yx}]$ and $[Z^i_{yy}]$ takes most of the time in electrically small and medium size structures. However, IE3D is efficient full wave electromagnetic simulation software which solves the above matrices for unknown surface currents and computes the important parameters, such as, return loss, voltage standing wave ratio, gain, directivity, radiation pattern etc. [21].

III. Antenna Geometry and Design Procedure

An effort to provide multiple functionality, usually results in increased size and additional complexity. For example, use of stacked patches/ multiple layers in reference [22], [23] and [24] increase the volume of antenna. Using patch antenna closely surrounded by patches occupies more lateral area. Similarly, air substrate has been widely used which increases thickness and thus, volume of the antenna. Approaches to enhance bandwidth include increasing substrate thickness, using electrically thick elements and parasitic element either in co-planar or stack configuration. But all these techniques increase the complexity and the size of the antenna. In our design, focus was on obtaining good electrical performance without design complexity or increased overall volume.

Physical layout of proposed antenna is given in Fig. 3. Slots of appropriate shape are incorporated on patch to perturb the surface current distribution, so as to optimize the results while maintaining single layer. The patch is shorted with the ground using a bent shorting plate on right edge. The proposed antenna's performances are analyzed using MoM based IE3D full wave electromagnetic simulation software. Since its formal introduction in 1993 IEEE International Microwave Symposium (IEEE IMS 1993), the IE3D has been adopted as an industrial standard in planar and 3D EM simulation. Since then, much improvement has been achieved in IE3D.

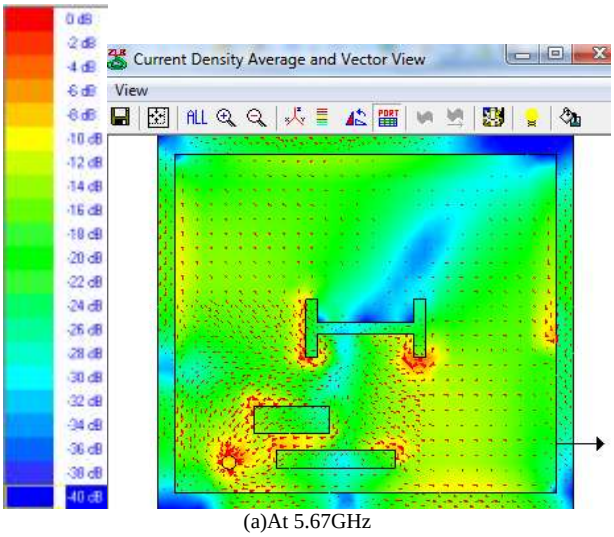


(a) Top view (b) Side view

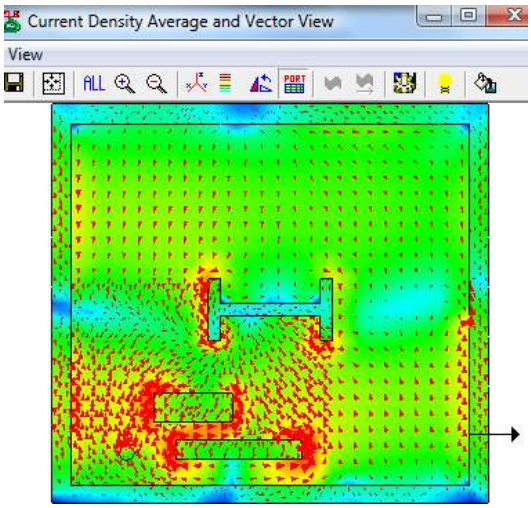
Fig. 3 Structure of the proposed antenna

IV. Results and Discussion

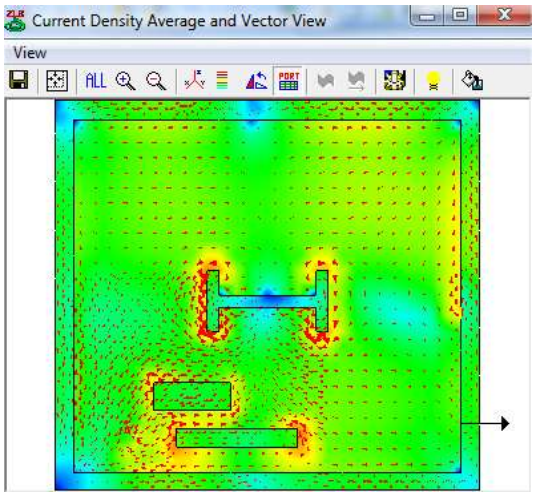
The simulator tool computes useful quantities of interest such as radiation pattern, input impedance, gain etc. Effect of shorting plates and slots can be well understood from the current density distribution as shown in Fig. 4 for different frequencies. The probe feed at one end provides current source which flows through the shorting plate to the bottom conducting plate. However, in between the feed and plate, slots are incorporated in centre with four branches to provide more and more perturbation to the surface current density distribution.



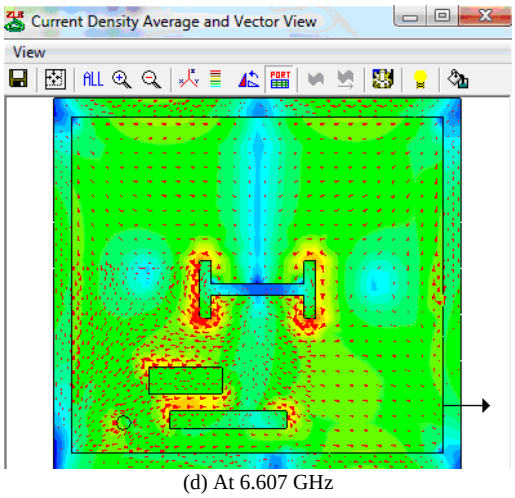
(a)At 5.67GHz



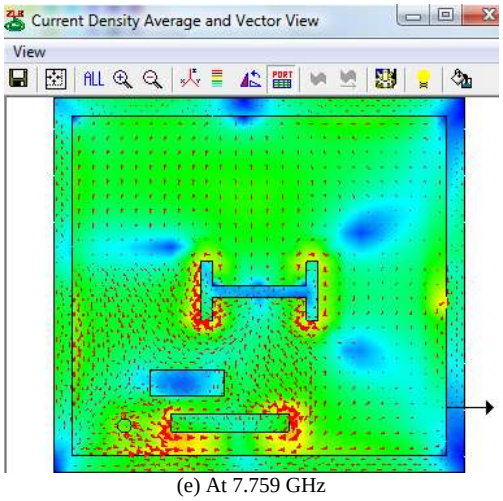
(b)At 6.12GHz



(c) At 6.386 GHz



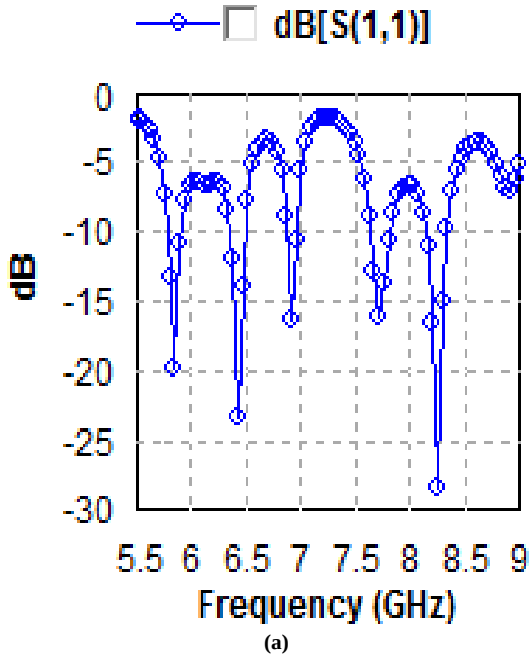
(d) At 6.607 GHz



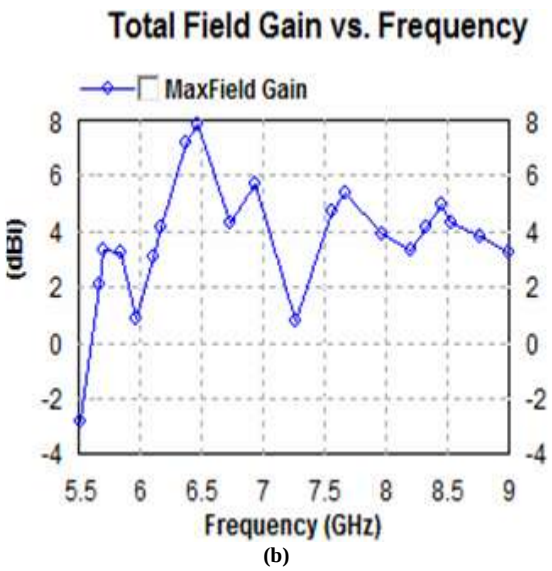
(e) At 7.759 GHz

Fig. 4 Current density distribution over the antenna surface

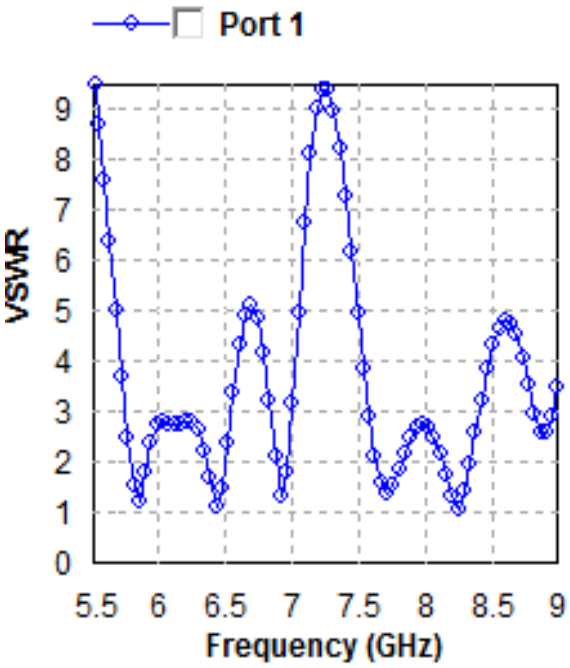
Fig. 5(a) introduces return loss vs. frequency curve for the proposed design. Return loss is defined as the ratio of the Fourier transforms of the incident pulse and the reflected signal. The bandwidth can be obtained from the return loss (RL) plot. The bandwidth is said to be those range of frequencies over which the return loss is less than -10 dB, which is approximately equivalent to a voltage standing wave ratio (VSWR) of less than 2:1. The simulated results show five bands at 5.673 GHz, 6.463 GHz, 6.880 GHz, 7.686 GHz and 8.402 GHz. The first band provides Wi-MAX, WLAN (IEEE 802.11a standard) and ISM (Industrial, Scientific and Medical) band applications. These days, WLAN is also being considered for mobile phone applications. Next three bands are useful for various wireless applications of C-Band, such as, object identification for surveillance applications, location tracking applications operating in the frequency band from 6 GHz to 9 GHz, cordless phones and other airborne applications. The last band offers some X-band applications like, ground-probing and wall-probing radar, tank level probing radar, sensors, automotive radar etc.



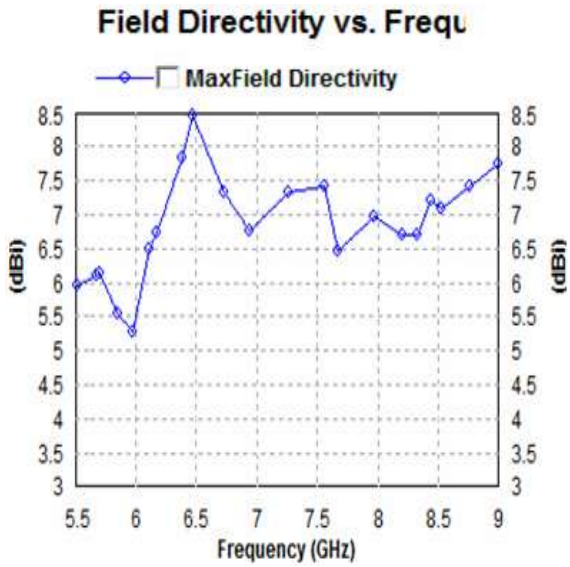
(a)



(b)

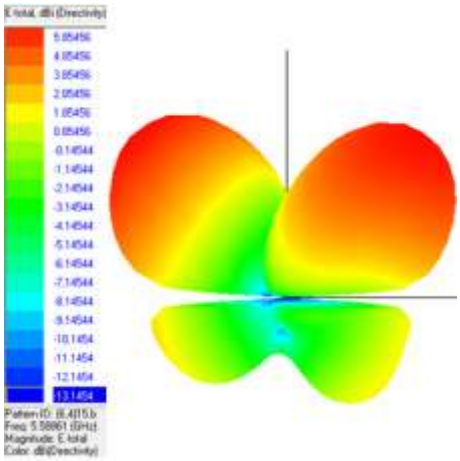


(c)

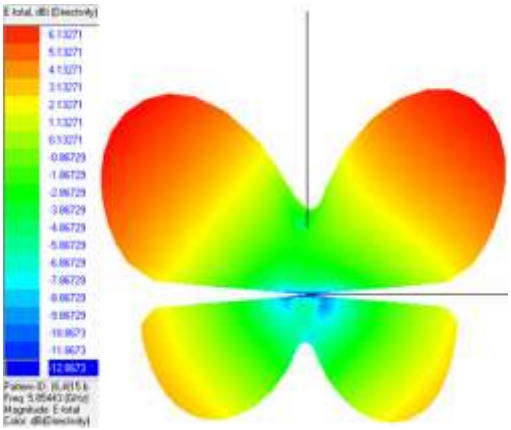


(d)

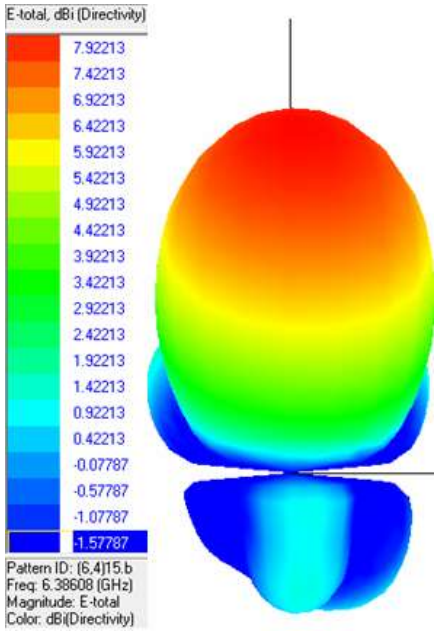
Fig. 5 (a) Return loss plot, (b) Gain, (c) VSWR, and (d) Directivity vs. frequency plot



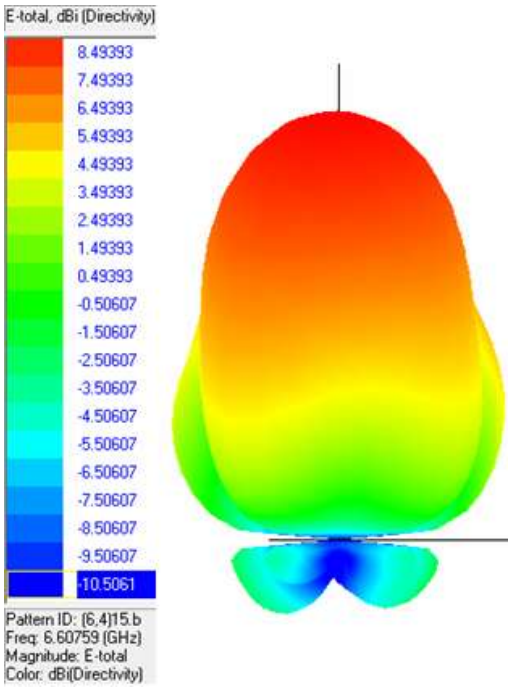
(a) At 5.588 GHz



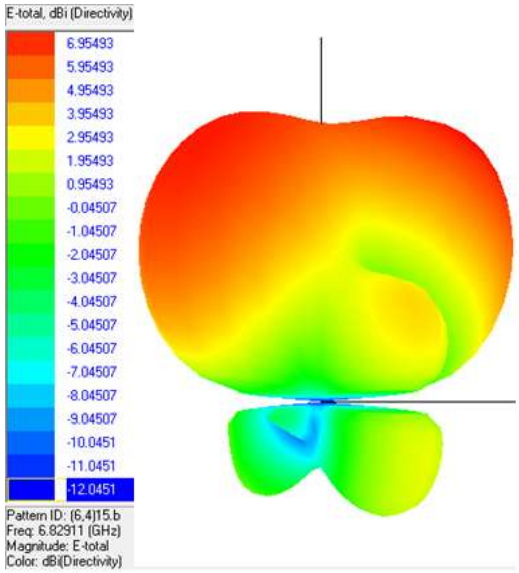
(b) At 5.8544 GHz



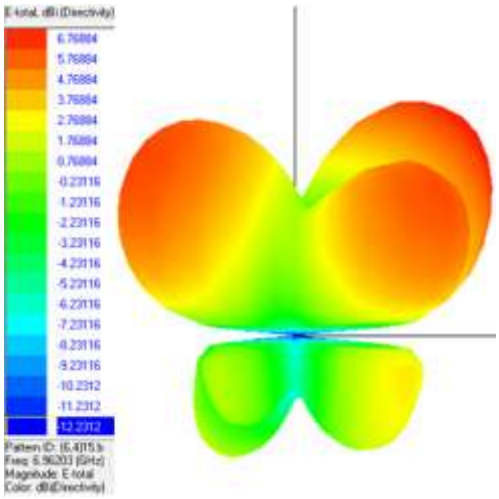
(c) At 6.38 GHz



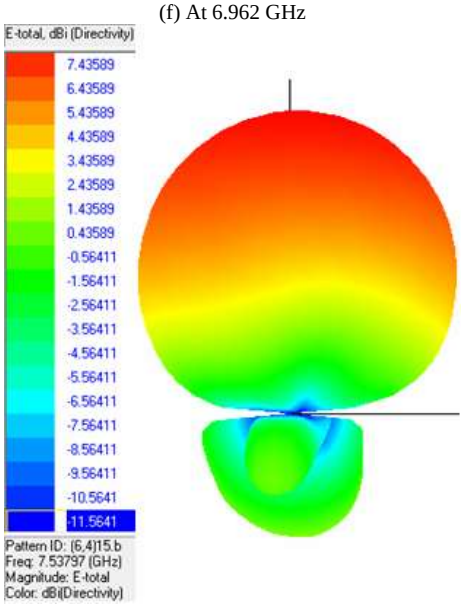
(c) At 6.607 GHz



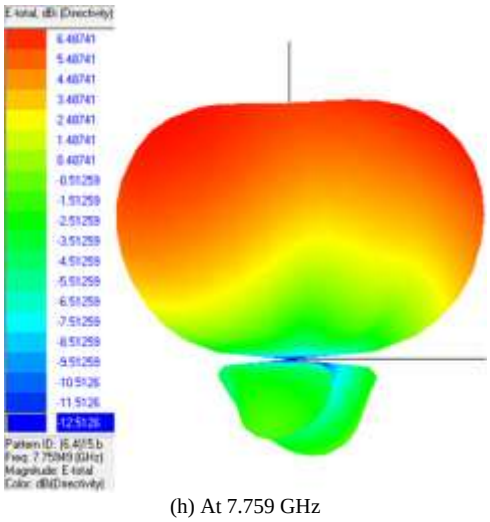
(e) At 6.829 GHz



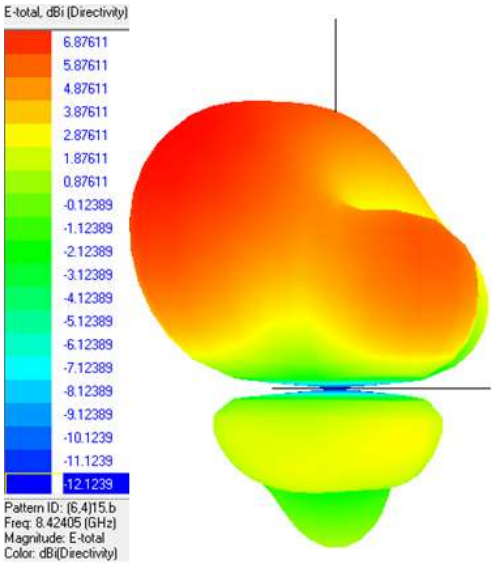
(f) At 6.962 GHz



(g) At 7.537 GHz



(h) At 7.759 GHz



(i) At 8.424 GHz

Fig. 6 True 3D radiation patterns on IE3D software

The antenna achieves a very high simulated directivity and gain as shown in Fig. 5(d) and 5(b). The peak gain reaches 8 dBi. Fig. 6 introduces simulated true 3D simulated radiation patterns at different frequencies.

V. CONCLUSION

It was shown that the method as used by IE3D can be used to efficiently model microstrip antennas. With computers getting larger, faster, and cheaper every month, large models will be able to be simulated with little difficulty. The rapid growth of wireless communications demands that antennas be low profile and allow operation at multiple frequency bands, eliminating the need for separate antennas for each application. This paper first presents a brief explanation of the method of moment. Then a compact, multipurpose and low manufacturing cost microstrip patch antenna is designed using this software and the simulation results are presented. The designed antenna possesses very high directivity and peak gain. According to the good electrical performance over the frequency range of interest, the proposed antenna should find use for several wireless applications like, Wi-MAX, WLAN, probing and automotive radar, wireless sensors, and various ground based and airborne applications in C-band.

REFERENCES

[1] Balanis C.A., "Antenna Theory-Analysis and Design"

[2] K.R.Carver,J.W. Mink "Microstrip Antenna Technology". *IEEE Trans. Antenna Propagat.*, Vol AP-29, No. 1 pp. 2-24.January 1981.

[3] R. Garg and P. Bhartia, *Microstrip antenna handbook*. Boston: Artech house, 2001

[4] Jin Sun, Chao-Fu Wang, Le-Wei Li, Mook-Seng Leong, "Further Improvement for Fast Computation of Mixed Potential Green's Functions for Cylindrically Stratified Media". *IEEE Transactions Antennas Propogat*, vol.52, pp.3026- 3036, 2004.

[5] S. Karan, V.B.Ertürk, and A. Altintas, "Closed-Form Green's Function Representations in Cylindrically Stratified Media for Method of Moments Applications". *IEEE Transactions Antennas Propogat*, vol.57, pp.1158-1168, 2009.

[6] V.B.Ertürk and R.G.Rojas, "Efficient Computation of Surface Fields Excited on a Dielectric-Coated Circular Cylinder". *IEEE Transactions Antennas Propogat*, vol.48, pp.1507-1516, 2000.

[7] V.B.Ertürk and R.G.Rojas, "Paraxial Space-Domain Formulation for Surface Fields on a Large Dielectric Coated Circular Cylinder". *IEEE Transactions Antennas Propogat*, vol.50, pp.1577–1587, 2002.

[8] V.B.Ertürk and R.G.Rojas, "Efficient Analysis of Input Impedance and Mutual Coupling of Microstrip Antennas Mounted on Large Coated Cylinders". *IEEE Transactions Antennas Propogat*, vol.51, pp.739-749, 2003.

[9] T.N.Kaifas and J.N.Sahalos, "Analysis of Printed Antennas Mounted on a Coated Circular Cylinder of Arbitrary Size". *IEEE Transactions Antennas Propogat*, vol.54, pp.2797-2807, 2006.

[10] Mang.He and Xiaowen.Xu, "Closed-Form Solutions for Analysis of Cylindrically Conformal Microstrip Antennas with Arbitrary Radii". *IEEE Transactions Antennas Propogat*, vol.53, pp.518-525, 2005.

[11] A. Svezhentsev, G.A.E. Vandenbosch, "Effcient Spatial Domain Moment Method Solution of Cylindrical Rectangular Microstrip Antennas". *IEE Proc.-Microw. Antennas Propag*, vol.153, pp.376-384, 2006.

[12] R.Cüneyt Acar, "Mutual Coupling of Printed Elements on a Cylindrically Layered Structure Using Closed-Form Green's Functions". *Progress in Electromagnetics Research*, PIER 78, pp.103-127, 2008.

[13] C.Tokgöz and G.Dural, "Closed-Form Green's Functions for Cylindrically Stratified Media". *IEEE Transactions Microwave Theory Tech*, vol.48, pp.40-49, 2000.

[14] J. R. Mosig and F. E. Gardiol, "General integral equation formulation for microstrip antennas and scatters," *Proc. Inst. Elect. Eng.*, pt. H, vol. 132, pp. 424–432, Dec. 1985.

[15] K. A. Michalski and D. Zheng, "Electromagnetic scattering and radiation by surfaces of arbitrary shape in layered media—Part I: Theory," *IEEE Trans. Antennas Propagat.*, vol. 38, pp. 335–344, Mar.1990.

[16] I. Park, R. Mittra, and M. I. Aksun, "Numerically efficient analysis of planar microstrip configurations using closed-form Green's functions," *IEEE Trans. Microwave Theory Tech.*, vol. 43, pp. 394–400, Feb. 1995.

[17] R.F.Harrington,*FieldComputation by Moment Methods*, Macmillan, New York, 1968; reprinted by IEEE Press, New York, 1993.

[18] Thomas A. Milligan, Modern Antenna Design, IEEE Press, A John Wiley & Sons, Inc., Publication, 2nd Edition

[19] I. Park, R. Mittra, and M. I. Aksun, "Numerically efficient analysis of planar microstrip configurations using closed-form Green's functions," *IEEE Trans. Microwave Theory Tech.*, vol. 43, pp. 394–400, Feb. 1995.

[20] M. I. Aksun, "A robust approach for the derivation of closed-form Green's functions," *IEEE Trans. Microwave Theory Tech.*, vol. 44, pp. 651–658, May 1996.

[21] IE3DUser's Manual,Zeland Software,Inc.,Fremont. C

[22] B. Sanz-Izquierdo, J.C. Batchelor, R.J. Langley and M.I. Sobhy, "Single and double layer planar multiband PIFAs", EEE Transactions on Antennas and Propagation, vol. 54, pp. 1416-1422, 2006.

[23] S. D. Targonski, R. B. Waterhouse, and D. M. Pozar, "Design of wideband aperture-stacked patch microstrip antennas," *IEEE Trans. Antennas Propagat.*, vol. 46, pp. 1245–1251, Sept. 1998.

[24] W.Chen, K. F.Lee, and R. Q. Lee, "Spectral-domain moment method analysis of coplanar microstrip parasitic subarrays," *Microw. Opt. Technol. Lett.*, vol. 6, no. 3, pp. 157-163, 1993.