IJACSA

WHERE WISDOM SHARES

ISSN 2158-107X(PRINT)
ISSN 2156-5570(ONLINE)

Volume 1 Issue 5, November 2010

INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS

INTERNATIONAL JOURNAL OF
ADVANCED COMPUTER SCIENCE AND APPLICATIONS

New York
United States of America

Technically Co Sponsored By:

THE SCIENCE AND INFORMATION ORGANIZATION

IJACSA Publications Indexed @

IJACSA Editorial

From the Desk of Managing Editor…

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

With each issue, IJACSA is moving towards the goal of being the most widely read and highly cited journal in the field of Computer Science and Applications

This journal presents investigative studies on critical areas of research and practice, survey articles providing short condensations of the best and most important computer science literature worldwide, and practice-oriented reports on significant observations.

IJACSA collaborates with The Science and Information Organization (SAI) that brings together relevant people, projects, and events in the domain, drawing on a wide research community studying various aspects of computer sciences in various settings like industrial and commercial organizations, professionals, governments and administrations, or educational Institutions.

In order to publish high quality papers**,** Manuscripts are evaluated for scientific accuracy, logic, clarity, and general interest. Each Paper in the Journal will not merely summarize the target articles, but will evaluate them critically, place them in scientific context, and suggest further lines of inquiry. As a consequence only 29% of the received articles have been finally accepted for publication.

IJACSA emphasizes quality and relevance in its publications. In addition, IJACSA recognizes the importance of international influences on Computer Science education and seeks international input in all aspects of the journal, including content, authorship of papers, readership, paper reviewers, and Editorial Board membership

IJACSA has made a tremendous start and has attracted the attention of Scientists and Researchers across the globe.

This has only been possible due to large number of submissions and valued efforts of the reviewers.

We are very grateful to all the authors who have submitted such work for the journal and the reviewers for dealing with the manuscripts very quickly.

We hope that the relationships we have cultivated will continue and expand.

Thank You for Sharing Wisdom!

Managing Editor
IJACSA
November 2010
editorijacsa@thesai.org

Editorial Board

IJACSA Reviewer Board

CONTENTS

Wavelet Time-frequency Analysis of Electro-encephalogram (EEG) Processing

Zhang xizheng[1],
[1]School of Computer and Communication
Hunan Institute of Engineering
Xiangtan China

Yin ling[2], Wang weixiong[1]
[2]School of Computer and Communication
Hunan University
Xiangtan, China P.R.

***Abstract*—This paper proposes time-frequency analysis of EEG spectrum and wavelet analysis in EEG de-noising. In this paper, the basic idea is to use the characteristics of multi-scale multi-resolution, using four different thresholds to wipe off interference and noise after decomposition of the EEG signals. By analyzing the results, understanding the effects of four different methods, it comes to a conclusion that the wavelet de-noising and soft threshold is a better conclusion.**

Keywords- EEG, time-frequency analysis, wavelet transform, de-noising.

I. INTRODUCTION

Electro-encephalogram (EEG) is the electrical activity of brain cell groups in the cerebral cortex or the scalp surface. The mechanism of EEG is a complex random signal within the brain activities, it is in the cerebral cortex of the synthesis of millions of nerve cells. Brain electrical activity is generated by electric volume conductor (the cortex, skull, meninges, and scalp). It reflects the electrical activity of brain tissue and brain function. Different state of mind and the cause of the cerebral cortex in different locations reflect the different EEG. Therefore, the electro-encephalogram contains plentiful physical, psychological and pathological information, analyzing and processing of EEG both in the clinical diagnosis of some brain diseases and treatments in cognitive science research field are very important.

EEG has the following characteristics[1-5]:

① EEG signal is very weak and has very strong background noise, the average EEG signal is only about 50gV, the biggest 100gV;

②EEG is a strong non-stationary random signal;

③nonlinear, biological tissue and application of the regulation function will definitely affect the eletro-physiological signal, which is nonlinear characteristics;

④EEG signal has frequency domain feathers.

As the EEG of the above characteristics, Fourier transformation and short time Fourier transformation analysis of EEG can not analyze it effectively. Therefore, this paper represents time-frequency analysis and wavelet transformation. The basic idea of wavelet transformation is similar to Fourier transformation, is using a series of basis function to form the projection in space to express signal. Classical Fourier transformation expanded the signal by triangulation of sine and cosine basis, expressed as arbitrary functions with different frequencies the linear superposition of harmonic functions, can describe the signal's frequency characteristics, but it didn't has any resolution in the time domain, can not be used for local analysis. It brought many disadvantages in theory and applications. To overcome this shortcoming, windowed Fourier transformation proposed. By introducing a time localized window function, it's improved the shortage of Fourier transformation, but the window size and shape are fixed, so it fails to make up for the defection of Fourier transformation. The wavelet transformation has good localization properties in time and frequency domain and has a flexible variable time-frequency window[6-9]. Compared to Fourier transformation and windowed Fourier transformation, it can extract information more effectively, using dilation and translation characteristics and multi-scale to analyze signal. It solved many problems, which the Fourier transformation can't solve[11,12].

Therefore, section II proposed time-frequency analysis of EEG spectrum and section III proposed EEG de-noising of the wavelet analysis method. The basic idea is to use the characteristics of multi-scale and multi-resolution, using four different thresholds to remove interference and noise decomposition of the EEG signals, final results show the de-noised signal.

II. TIME-FREQUENCY ANALYSIS

Time-frequency analysis is a nonlinear quadratic transformation. Time-frequency analysis is an important branch to process non-stationary signal, which is the use of time and frequency of joint function to represent the non-stationary signal and its analysis and processing.

A. Spectrogram

Spectrogram is defined as the short time Fourier transform modulus of the square, that is,

$$S_z(t,f)=\left|STFT_z(t,f)\right|^2=\left|\int_{-\infty}^{+\infty} z(t')\eta^*(t'-t)e^{-j2\pi ft'}dt'\right|^2$$

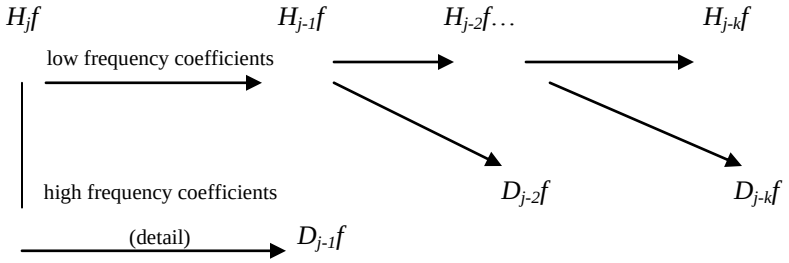
. It is real, non-negative quadratic distribution, with the following properties:

(1) time and frequency shift invariance;

(2) frequency resolution as with the short time Fourier transform limited;

(3) there is interference;

Spectrogram can be more clearly seen the emergence of some short transient pulse in the EEG signal.



B. Time-frequency Analysis in Signal Processing

EEG is a brain electrical activity of non-invasive method. Fourier transformation and the linear model have been widely used to analyze the pattern of EEG characteristics and non-transient EEG activity, but only for stationary signals' spectrogram analysis. It is not appropriate to transient spontaneous EEG and evoked potential, which are non-stationary signal. Therefore, it's necessary to use time-frequency analysis.

EEG often has some short transient pulse, which contains some important pathological information, and some belong to interference. As the EEG is highly non-stationary, using time-frequency analysis toolbox tfrsp function to analyze the spectrum is a good way.

III. WAVELET TRANSFORM ATION

Wavelet transformation is a time-scale analysis method and has the capacity of representing local characteristics in the time and scale (frequency) domains. In the low frequency, it has a lower time resolution and high frequency resolution, the high frequency part has the high time resolution and lower frequency resolution, it is suitable for detection of the normal signal, which contains transient anomalies and shows their ingredients.

A. The Basic Principle of Wavelet Transformation

Telescopic translation system { $\phi_{a,b}$ } of basic wavelet $\phi(t)$ is called wavelet function, denoted

$$\psi_{a,b}(t)=\frac{1}{\sqrt{a}}\psi(\frac{t-b}{a}) \quad (1)$$

type of a,b(including the subscript a,b) are called scale parameters and positional parameters respectively. Wavelet transformation of any function $f(t)$ is called the inner product of function $f(t)$ and wavelet function.

$$W_f(a,b)=\{f(t),\psi_{a,b}(t)\} \quad (2)$$

Wavelet transformation is a time-frequency analysis and reflected the state of function $f(t)$ in the scale(frequency) and position(time). If the check scale is $a=2t$, $j\in z$, that is a dyadic wavelet transformation. Usually Mallat tower algorithm proposed discrete dyadic wavelet transformation calculation, discrete signal sequence of function $f(t)$ is $f(n)$ n=1,2…n, and its discrete dyadic wavelet transform is as follows:

$$C_{J+1}(n)=\sum_{k\in z}h(k-2n)C_i(k) \quad (3)$$

$$D_{i+1}(n)=\sum_{k\in z}g(k-2n)C_j(k) \quad (4)$$

type of the above formulas:$h(k)$ and $g(k)$ is the wavelet function ϕ_{2j} ,the conjugate orthogonal $b(t)$ set the filter coefficients $g(k)=(-1)h(1-k)g(k)$, C and D are called the approximation signal at scale parts and detail parts. When the original signal can be seen as an approximation of scale J=0, that is $c(n)=f(n)$. Discrete signal decomposition by the scale j=1,2,3,r…j, get D_1, D_2, D_3 , …, D_j, C_j.

B. Multi-resolution of Wavelet Transformation

Multi-resolution analysis decomposes the processed signal to the approximation signal and detail signal at different resolutions with orthogonal transformation. Multi-resolution analysis can express the following formula:

$$V_0=V_1\oplus W_1=V_2\oplus W_2\oplus W_1=V_3\oplus W_3\oplus W_2\oplus W_1=\cdots \quad (5)$$

Mallat tower algorithm can represent the original signal with detail signal in a series of different resolutions. The basic idea: The energy limited signal H_jf approximating in the resolutions 2^j can be further decomposed into approximation $H_{j-1}f$, which is under the resolution2^{j-1}, and the details $D_{j-1}f$ in the resolution 2^{j-1} and 2^j, the decomposition process are shown in Figure1.

C. Wavelet De-noising

Signal de-noising actually inhibit the useless part and restore the useful part. According to Mallat signal decomposition algorithm, it can remove the corresponding high-frequency of noise and low-frequency approximation of the relevant part of signal and then reconstruct to form the filtered signal [14,15].

There are many types of wavelet functions, the article is using Daubechies wavelet function, wavelet decomposition using db signal. Daubechies wavelet is a compactly supported wavelets, the majority does not have symmetry.

This paper uses four different de-noising methods, including wavelet de-noising, the default threshold de-noising, soft threshold and hard threshold. In engineering technology, if the received signal is $X(t)$, which generally contains two components: one is a useful signal $S(t)$, through analyzing and studying of the signal, we can understand the nature of object; the other is the noise $N(t)$, which has intensity spectrum distributing in the frequency axis, it is hindered us to understand and master the $S(t)$.

To illustrate the extent of the problem, expressed as the limited noise signal:

$$X_i(t) = S_i(t) + N_i(t) \quad (i = 1,2,3,\cdots n)$$

The basic purpose of signal processing is making the maximum extent possibility to recover the effective signal from the contaminated signal $X_i(t)$, maximum suppression or elimination of noise $N_i(t)$. If $\tilde{S}$ is expressed as signal, which is processed after de-noising, TH is threshold value, wavelet transformation of X,S are expressed as $X_i(t)$and $S_i(t)$ respectively, so the Donoho nonlinear threshold described as follows:

(1) after wavelet transformation, signal $X_i(t)$, obtained as X;

(2) in the wavelet transformation domain, threshold is processed in wavelet coefficients.

Soft - Threshold

$$\tilde{s} = \begin{cases} \mathrm{sn}g(x)(|x| - TH) & |x| \geq TH \\ 0 & |x| \leq TH \end{cases}$$

Hard - Threshold

$$\tilde{s} = \begin{cases} x & |x| \geq TH \\ 0 & |x| \leq TH \end{cases}$$

(3) Wavelet inverse transformation calculation is obtained $s_i^*(t)$ (* in order to distinguish it from $s_i(t)$).

It can be seen that different threshold values are set at all scales, then the wavelet transformation coefficients compared with the threshold values, if less than this threshold, we think that the noise generated and set to zero, if more than, to retain its value, thus achieving the purpose of de-noising. Clearly, the crucial point is how to choose threshold value between preserving signal details and selecting the de-noising capacity, to some extent, it is directly related to the quality of the signal de-noising.

Generally, Th is taken as: Th $=\sigma\sqrt{2\log n}$, also in the resolution of the wavelet transformation coefficients, taking a percentage of maximum value or absolute value as threshold.

IV. Experimental Results

This is a spectrogram analysis of EEG data. From the experimental results, it can be seen that there were many time-domain waveform pulse signal, but we cannot determine the frequency range, we also cannot rule out the interference caused by transient pulse. From the EEG signal spectrogram, it can be seen mainly in the 10 Hz or so, but still not make sure the exact range. Therefore, we calculated this spectrogram as shown in Figure4 and Figure5 to show transient pulses existing in the 0.9s to 1.1s and 1.4s to 2.0s. So we can better extract the pathological information from the transient pulse signal.

Following the results of wavelet de-noising analysis, four de-noising methods are used in this paper. Figure6 is original EEG waveform, in order to comparing with the filtered signals.

Wavelet de-noising is the most important aspect in signal processing. From Figure7, it can be seen that EEG signals largely restore the original shape, and obviously eliminates noise cause by interference. However, compared with original signal, the restored signal has some changes. This is mainly not appropriate to choose wavelet method and detail coefficients of wavelet threshold.

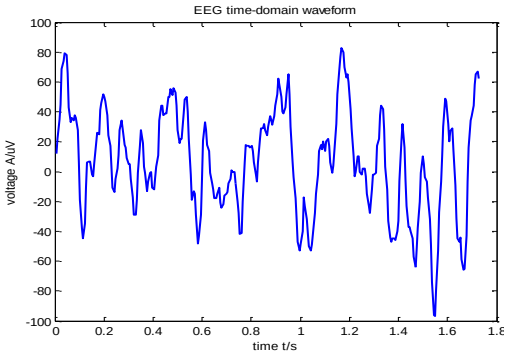


Figure 2 EEG time-domain waveform

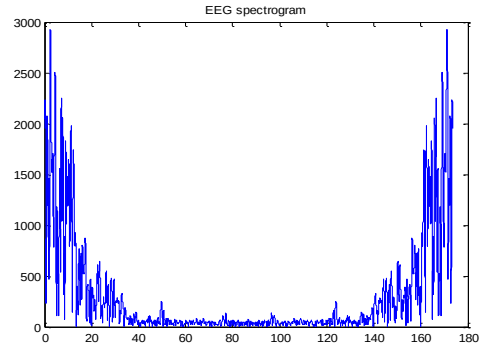


Figure 3 EEG spectrogram

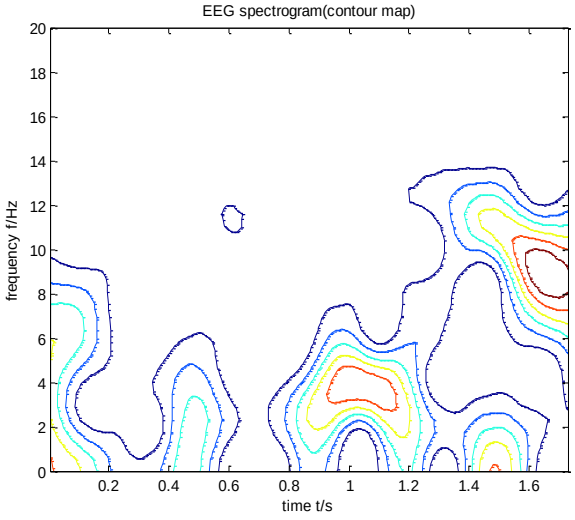


Figure 4 EEG spectrogram (contour map)

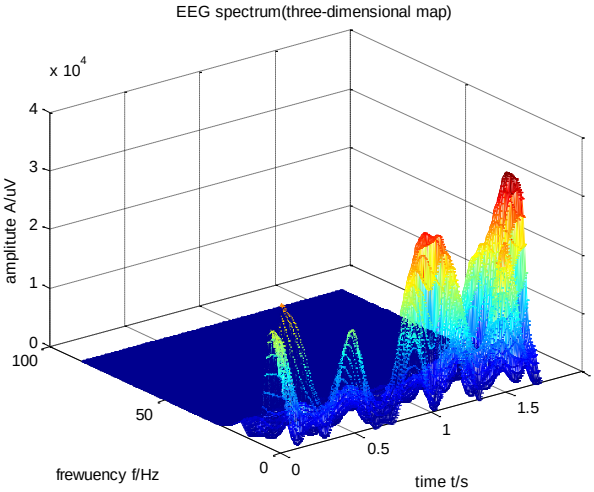


Figure 5 EEG spectrogram (three-dimensional map)

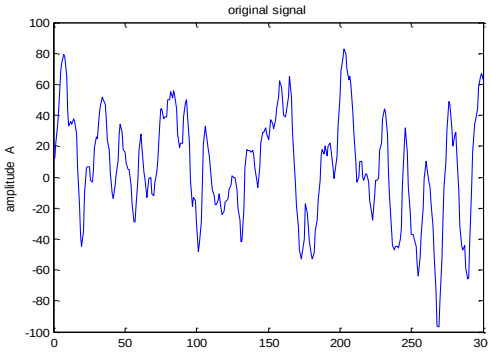


Figure 6 original signal

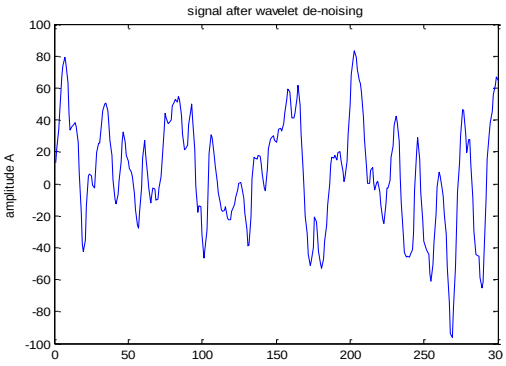


Figure 7 signal after wavelet de-noising

After de-noising with the default threshold, the signal is smooth, but may lose some useful signal components.

After hard threshold de-noising, the restored signal is almost the same with the original signal, it is indicated that hard threshold is not a good method.

Soft threshold de-noising eliminates noise effectively and has very good retention of the useful signal components.

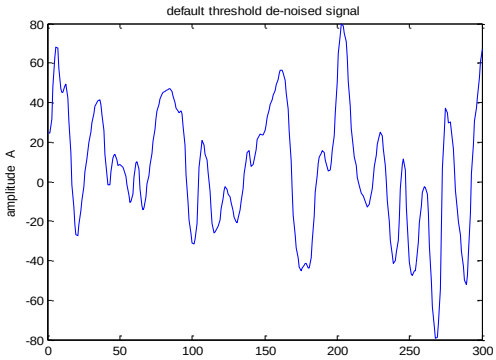


Figure 8 the default threshold de-noised signal

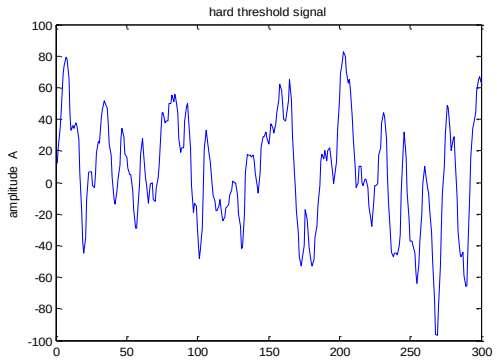


Figure 9 signal after hard threshold de-noising

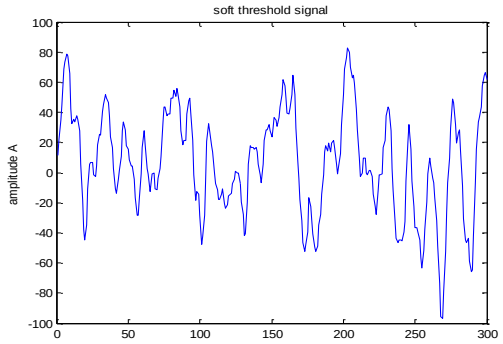


Figure 10 signal after soft threshold de-noising

V. CONCLUSION

In this paper, time-frequency analysis toolbox function tfrsp is used in analysis spectrogram of EEG. As can be seen from the spectrum and spectrogram, analyzing spectrogram can be known the specific time period of useful transient information. Thus, it can be very easy to extract useful diagnostic information through the analysis of pathological in medicine. There are four de-noising methods, including wavelet de-noising, default threshold, hard threshold and soft threshold, wavelet de-noising is to choose wavelet function db5 and the level of decomposition 3. To ensure signal without distortion, it is better to choose wavelet de-noising and soft threshold de-noising. So, they are widely used in signal processing.

ACKNOWLEDGEMENT

The authors are grateful to the anonymous reviewers and to the Natural Science Foundation of Hunan Province for supporting this work through research grant JJ076111, and the Student Innovation Programme of Hunan Province through research grant 513.

REFERENCES

[1] G. Jianbo, H. Sultan, H. Jing, T. Wen-Wen,"Denoising Nonlinear Time Series by Adaptive Filtering and Wavelet Shrinkage: A Comparison", IEEE Signal Processing Letters, 2010, vol.17, no.3, pp.237 – 240.

[2] G. Zhexue, C.Zhongsheng, "Matlab Time-frequency Analysis and Its Application (Second Edition) ," Beijing: Posts & Telecom Press. 2009.

[3] J.J. Kierkels, G.M.Boxtel, L.L.Vogten, "A model-based objective evaluation of eye movement correction in EEG recordings," IEEE Trans.Bio-Med.Eng., 2006, vol. 53, no.5, pp.246-253.

[4] A.R.Teixeira, A.M.Tome, E.W.Lang, *et a1*. "Automatic removal of high-amplitude artefacts from single-channed eletro-encephalograms". Comput Methods Program Biomed, 2006, vol. 83,no.2, pp.125-138.

[5] S. Romero, M.A.Mafianas, M.J.Barbanoj. "A comparative study of automatic techniques for ocular artifact, reduction in spontaneous EEG signals based on clinical target variables: a simulation case," Computer Biology Medicine, 2008, vol.38, no. 3, pp.348-360.

[6] D. Yao, L. Wang, R. Ostenveld, *et a1*. "A comparative study of different references for EEG spectral mapping the issue of neutral reference and the use of infinity reference". Physiological Measurement, 2005, vol. 26, no.1, pp.173-184.

[7] D. Yao. "High-resolution EEG mapping:an equivalent charge-layer approach," Physics in Medicine and Biology, 2003, vol.48, no.4, pp.1997-201l.

[8] V.Sampsa, V.Juha, K.Kai,"Full-band EEG(FbEEG): an emerging standard in electroeneep halography". Clinical Neurophysiology, 2005, vol. 116, no.1, pp.1-8.

[9] A. Phinyomark, C. Limsakul, P. Phukpattaranont, "EMG denoising estimation based on adaptive wavelet thresholding for multifunction myoelectric control," CITISIA 2009, 25-26 July 2009, pp.171 – 176.

[10] G.Umamaheswara, M. Muralidhar, S. Varadarajan,"ECG De-Noising using improved thresholding based on Wavelet transforms", International Journal of Computer Science and Network Security, 2009, vol.9, no.9, pp. 221-225.

[11] M.Kania, M.Fereniec, and R. Maniewski, "Wavelet denoising for multi-lead high resolution ECG signals", Measurement Science review, 2007, vol.7.pp. 30-33.

[12] Y. Ha,"Image denoising based on wavelet transform", Proc. SPIE, Vol. 7283, 728348 (2009).

[13] A. Borsdorf, R. Raupach, T. Flohr, and J. Hornegger, "Wavelet Based Noise Reduction in CT-Images Using Correlation Analysis," IEEE Transactions on Medical Imaging, 2008, vol. 27, no. 12, pp. 1685–1703.

[14] Y. Lanlan,"EEG De-Noising Based on Wavelet Transformation", 3rd International Conference on Bioinformatics and Biomedical Engineering, 2009 11-13 June 2009 On page(s): 1 – 4.

[15] M.K. Mukul, F. Matsuno,"EEG de-noising based on wavelet transforms and extraction of Sub-band components related to movement imagination", ICCAS-SICE, 2009, 18-21 Aug., pp.1605 – 1610.

Xizheng Zhang received the B.S. degree in control engineering and the M.S. degree in circuits and systems in 2000 and 2003, respectively, from Hunan University, Changsha, China, where he is currently working toward the Ph.D. degree with the College of Electrical and Information Engineering.

He is also with the School of Computer and Communication, Hunan Institute of Engineering, Xiangtan, China. His research interests include industrial process control and artificial neural networks.

Requirements Analysis through Viewpoints Oriented Requirements Model (VORD)

Ahmed M. Salem
Computer Science Department
California State University, Sacramento
Sacramento, CA 95819 USA
Email: salema@ecs.csus.edu

***Abstract*— This paper describes an extension to the Viewpoints Oriented Requirements Definition (VORD) model and attempts to resolve its lack of direct support for viewpoint interaction. Supporting the viewpoint interaction provides a useful tool for analyzing requirements changes and automating systems. It can also be used to indicate when multiple requirements are specified as a single requirement. The extension is demonstrated with the bank auto-teller system that was part of the original VORD proposal.**

Keywords: Software Requirements; Requirements Modeling; Functional Requirements

I. INTRODUCTION

The Viewpoints Oriented Requirements Definition (VORD) was proposed by [1] by Kotonya and Somerville as a method to tackle requirements engineering from a viewpoint level. A significant part in the software development process today, is not anymore programming, designing or testing, but requirement analysis. Interactive systems, whose operations involve a degree of user interaction, have a serious problem in identifying all the clients' needs. At the same time the analyst has to be sure that all needs are recognized in a valid way. The VORD method is useful in detecting these user needs, and also identifying the services that a user expects from the system [10]. It provides a structured method for collecting, documenting, analyzing, and specifying viewpoints and their requirements. Viewpoints map to classes of end-users of a system or to other systems interfaced to it. The viewpoints that make up the core model are known as direct viewpoints. To allow organizational requirements and concerns to be taken into account, viewpoints concerned with the system's influence on the organization are also considered. These are known as indirect viewpoints [2]. When describing the VORD model, its creators identified a limitation that it does not explicitly support the analysis of interaction across and within all of the viewpoints. This paper will expand upon the Viewpoints Oriented Requirements Definition (VORD) process by modifying the approach; so that it supports viewpoint interaction without any degradation to the existing framework. In addition to proposing a theoretical solution to support viewpoint interaction, we will show how the solution works by using practical examples. Much of the breadth of this paper consists of working out the practical application of our model extension.

II. BACKGROUND

In a follow-up piece, [2] on the VORD model, one of the original authors noted that the viewpoint interaction limitation still existed. After researching other papers on viewpoint interaction, we were able to find one that closely expanded on this topic, however the VORD model was not updated directly. The paper [3] was on the VISION model proposed by Araújo and Coutinho. Their approach took ideas from the VORD and PREVIEW [4] models, then incorporated viewpoint associations, UML modeling, and aspectual use cases. They have identified some basic methods of identifying and documenting viewpoint relationships that we would like to incorporate into our enhancement. We will expand the viewpoint interaction portion of their solution, and then use it to enhance the VORD method.

Other authors have also used the VORD process model during requirement analysis; a paper by author Zeljka Pozgaj, clearly explains the three steps of the VORD using a Stakeholder example. The UML diagrams illustrate the viewpoint and service interactions [10]. There have been other methods of incorporating multiple viewpoints into requirements engineering other than the VORD model. Lee's Proxy Viewpoints Model-based Requirements Discovery (PVRD) [5] methodology is especially designed for working from "legacy" SRS documents. Others use viewpoints as a tool for addressing possible inconsistencies in requirement specifications such as Greenspan, Mylopoulos, and Borgida [6] and Nuseibeh and Easterbrook [7].

III. VORD PROCESS MODEL

To gain an appropriate understanding of the process extension, it is necessary to provide a brief description of the VORD process model itself. The VORD process model is designed to elicit, analyze, and document the requirements for a Service-Oriented System (SOS). It specifically attempts to look at all the entities that will interact or otherwise use the services of the system. The requirement sources may be from stakeholders, other systems that interface with the proposed system, or other entities in the environment of the proposed

system that may be affected by its operation. Each requirement source is then considered to be a *viewpoint*.

VORD defines two classes of viewpoints:

1) Direct viewpoints: These correspond directly to clients in that they receive services from the system and send control information and data to the system. Direct viewpoints are either system operators/users or other sub-systems, which are interfaced to the system being analyzed.

2) Indirect viewpoints: Indirect viewpoints have an "interest" in some or all of the services which are delivered by the system but do not interact directly with it. Indirect viewpoints may generate requirements which constrain the services delivered to direct viewpoints [9]. Each viewpoint has a relationship with the proposed system based upon its needs and interactions with the system. The model assumes that if all the viewpoints have been analyzed and specified, then all the system's requirements would also have been analyzed and specified.

The VORD process model is shown in Fig. 1. The first three iterative steps are:

1) *Viewpoint identification and structuring*
2) *Viewpoint documentation*
3) *Viewpoint requirements analysis and specification*

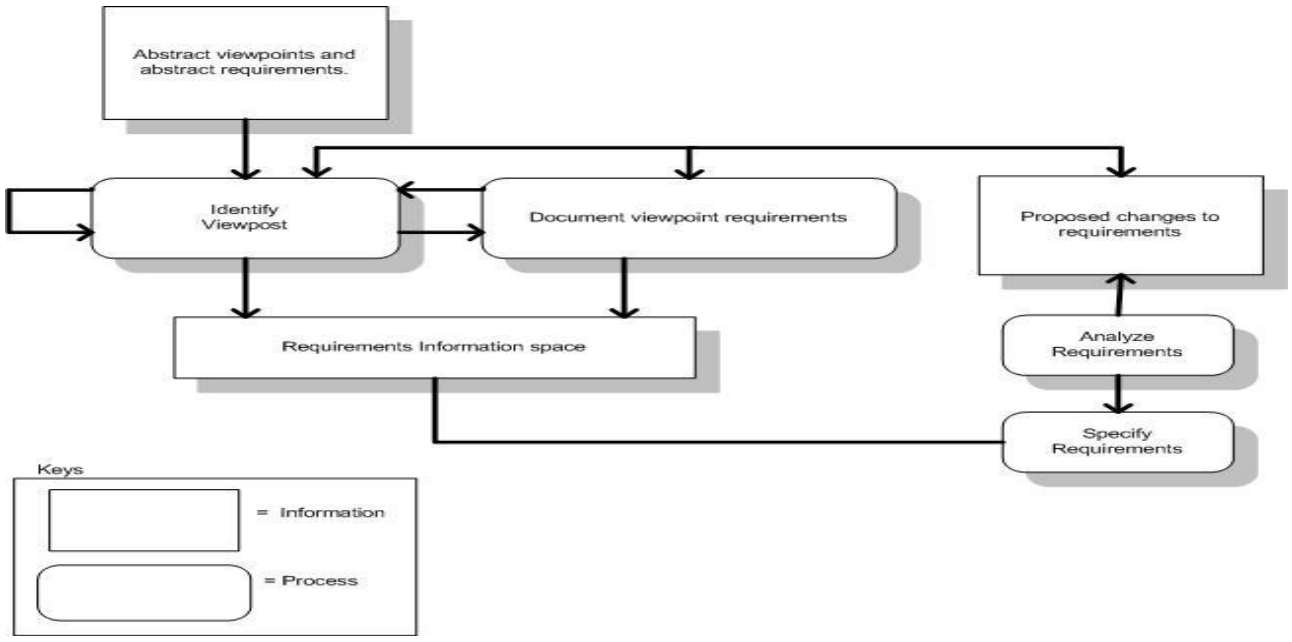


Figure 1. The VORD Process Model

The first step, viewpoint identification and structuring, is concerned with identifying relevant viewpoints in the problem domain and structuring them. The second step is concerned with documenting the viewpoints identified in step 1. Viewpoint documentation consists of documenting the viewpoint name, requirements, constraints on its requirements and its source. Viewpoint requirements include a set of required services, control requirements and set of non-functional requirements. The last step is concerned with analyzing, and specifying the functional and non-functional viewpoint requirements in an appropriate form [9].

IV. Proposed VORD Process Model Extension

Our extension is an iterative process that takes place after step three. Our extension has three steps:

1) *Requirement to viewpoint mapping*
2) *Viewpoint interaction analysis*
3) *Viewpoint interaction documentation (interaction matrix)*

The first step of our model extension is to map each requirement to its associated viewpoints. This is actually backward from the VORD model listing viewpoints first and requirements second. This is needed since we assume that the most reliable method of identifying interactions is at the level of required services, control requirements, or set of non-functional requirements. This step is done by first listing all the labeled requirements, both functional and non-functional. Then the associated viewpoints for each requirement are listed.

The second step of our model extension is to determine if any viewpoint interaction exists for each requirement. This is done by analyzing the list created in step one, along with the specification for each requirement. If a requirement has only one associated viewpoint, then we can assume that no

viewpoint interaction takes place (for that requirement), and no further analysis is needed. If there are two or more viewpoints listed, then further analysis is needed. This analysis consists of analyzing the requirement specification and determining if the first viewpoint listed interacts with the second viewpoint listed. This process is continued until all the viewpoints for the requirement have been compared against all the other viewpoints for the requirement. If any interaction is discovered, then the interaction type must be determined. If there is a transitive relationship, i.e. the viewpoints interact but only through another viewpoint, then it is considered to be indirect interaction. If the viewpoints interact directly, then it is considered to be direct interaction. If there is no interaction, then this may be an indication of a *compound requirement*. In this case, the requirement should probably be split up into two or more requirements.

The third step of our model extension is to document the viewpoint interactions discovered in step two. The results are displayed in an interaction matrix that has rows and columns for each viewpoint and the requirement name(s) listed in the corresponding "box". Note that there may be more than one requirement listed in a box since two viewpoints may have interactions in more than one requirement. There may be boxes with no requirements listed since two viewpoints may not always have an interaction.

V. ACTICAL APPLICATION AND ANALYSIS

We devised several examples of viewpoint interactions, and then applied our model extension to them. The reason for this is two fold. The first reason is to provide a means of explaining our proposal. Using examples provide a clear understanding of the theoretical model. The second reason is to actually "test" the practicality of our proposal. Using examples, not only help to show the proposal's strengths and limitations, but also its ease of use.

By continuing the ATM machine case study, we feel that our paper compliments the original proposal; similarly, our theoretical model compliments the original model. The requirements chosen to extend the ATM case study to demonstrate viewpoint interactions are listed in Table I. These requirements are specified to a much less thorough extent than they would be if they were actually part of an SRS document. However, the purpose here is to provide just enough narrative to convey the essence of the requirement.

A. *Example 1: Banking*

The original VORD proposal listed the following viewpoints in the case study.

- Bank manager
- ATM operator
- Home customer
- Customer database
- Foreign customer
- Security officer
- System developer
- Bank policy

For the purposes of clarity, we modified the viewpoint list.

The home customer and foreign customer viewpoints are combined into a single "bank customer" viewpoint. The reason is that no distinction is necessary in the requirement examples that we have chosen. We also added the Bank employee viewpoint since the ATM operator viewpoint is only concerned with stocking the ATM with cash and starting and stopping its operation. The result of applying step one of our model extensions is listed in Table II.

Step two requires more detailed analysis. Within each requirement, each viewpoint is analyzed and compared to the other viewpoints. Listed below is the viewpoint interaction analysis for each requirement. The general approach is to compare each viewpoint to all the other viewpoints in the requirement. Then, determine if any interaction takes place. If interaction takes place directly, then it is considered a direct interaction. If two viewpoints interact, but only through other viewpoints, then it is considered an indirect interaction.

TABLE I. EXTENDED CASE STUDY REQUIREMENTS

Requirement Name	Description
R1	Bank manager approves the bank customer to withdraw funds above the daily withdraw limit.
R2	Bank employee reset's bank customer's forgotten PIN number.
R3	Bank employee provides replacement card to a bank customer. (Card is lost, stolen, damaged etc.)
R4	Bank customer notifies bank employees of ATM machine problems. (Out of money, malfunctioning etc.)
R5	Bank employee issues a ATM card to a new customer.
R6	Bank customer reports unauthorized withdraw from ATM to bank employee.
R7	Bank employee notifies bank customer of a newly added ATM machine.
R8	Bank customer makes a deposit.

1) *R1 Viewpoint Interaction Analysis*

The case study of the original VORD proposal listed bank policy as the viewpoint handling for a variety of business rules. Although, not explicitly stated in the original proposal, we are

assuming that a daily ATM withdraw limit per bank customer is part of bank policy.

The bank customer's withdraw is limited by the bank policy. The bank customer requests a bank employee, to permit him to withdraw funds beyond the daily limit. The bank employee then informs the bank manager of the request, along with any relevant justification. The bank manager then modifies the bank policy (granting additional amount to the customer's daily withdraw limit), which allows the bank customer to withdraw additional funds. The bank customer directly interacts with the bank employee with the request to withdraw additional funds.

The bank employee and the bank manager directly interact with the bank customer's request. If the bank manager grants the customer request, then the bank manager interacts directly with the bank policy. The bank customer indirectly interacts with the bank manager through the bank employee.

The bank employee indirectly interacts with the bank policy through the bank manager.

2) R2 Viewpoint Interaction Analysis

The bank customer contacts a bank employee to have a PIN number reset. The bank employee then resets the PIN number. Therefore, the bank employee and bank customer directly interact.

3) R3 Viewpoint Interaction Analysis

The bank customer contacts a bank employee to request a new ATM card. The bank employee then provides the bank customer with a new card. Therefore, the bank employee and bank customer directly interact.

4) R4 Viewpoint Interaction Analysis

The bank customer notifies the bank employee of an ATM machine problem. The bank employee in turn notifies the ATM operator about the problem. After the problem is resolved, the ATM operator notifies the bank employee, who then informs the bank customer who reported the problem. Direct interaction takes place between the bank customer and bank employee (notification). Direct interaction takes place between the bank employee and ATM operator (notification). The bank customer initiates the action that results in the ATM operator performing a function. However, this is done though the bank employee viewpoint. Therefore, the bank customer and the ATM operator indirectly interact.

5) R5 Viewpoint Interaction Analysis

The bank customer opens an account and the bank employee provides an ATM card. Therefore, the bank employee and bank customer directly interact.

6) R6 Viewpoint Interaction Analysis

The bank employee and bank customer directly interact with the notification and collection of facts. The bank employee and the bank manager directly interact with data exchange. The bank manager and bank policy directly interact to determine if the account should be credited for the unauthorized withdraw. The bank policy and the bank customer directly interact with a notification of the manager's decision. The bank customer initiates action that results in the bank manager performing a function. However, this is done through the bank employee viewpoint. Therefore, the bank manager and the bank customer indirectly interact. The bank employee and the bank policy interact indirectly through the bank manager.

7) R7 Viewpoint Interaction Analysis

The bank employee notifies that bank customer of a newly added ATM machine. Therefore, the bank employee and bank customer directly interact.

8) R8 Viewpoint Interaction Analysis

The bank customer is the only viewpoint listed, so it does not interact with any other viewpoints. This requirement was put in place to show that not all requirements will contain viewpoint interaction.

The third step in our model extension is to document the analysis performed in step two. This is displayed in a matrix that consists of columns and rows for each viewpoint. The first column and first row are considered as the "headers" of the matrix as they list the viewpoints used in the example. The intersection of each row and column is a box that represents the interaction between the corresponding viewpoints depicted in the headers. Since each viewpoint is listed twice (first column and first row), there will be two corresponding "boxes" for

TABLE II. REQUIREMENTS TO VIEWPOINT MAPPING

Requirement Name	Description
R1	Bank customer, bank employee, bank manager, bank policy.
R2	Bank employee, bank customer.
R3	Bank employee, bank customer.
R4	Bank customer, bank employee, ATM operator.
R5	Bank employee, bank customer.
R6	Bank customer, bank employee, bank manager, bank policy.
R7	Bank employee, bank customer.
R8	Bank customer.

each viewpoint interaction. The corresponding "box" for each viewpoint interaction lists the requirement name (that is the requirement in which the viewpoints interacted), along with an interaction type designation, "D" for direct and "I" for indirect interaction. Note that the matrix design will include two "boxes" that correspond to the interaction of a single viewpoint (i.e.: bank customer to bank customer). These will be left blank since a viewpoint did not interact with itself in any instance.

The viewpoint interaction matrix created from the analysis in step two is displayed in Table III. The matrix was created using the methodology described in the previous paragraph. The first column and first row list all the defined viewpoints. The corresponding "box" for each viewpoint interaction list all the requirements where the viewpoints interact, along with a designation of "D" for direct and "I" for indirect. For the purposes of brevity, only the viewpoints used in R1 – R8 are shown.

B. Example 2: Library System

By considering the Library System (LS) case study, we propose that our paper compliments the original proposal; similarly, our theoretical model compliments the original model. The requirements chosen to extend the LS case study to demonstrate viewpoint interactions are listed in Example 2 – Table IV. These requirements are briefly described than they would be as part of an SRS document. However, the purpose here is to provide just enough narrative to convey the essence of the requirement.

Viewpoints in Case Study:

- Library Member
- Library Staff
- Library Manager
- Library Policy
- Library IT Administrator
- Library Database

The result of applying step one of our model extensions is listed in Example 2 - Table V.

The library system is then analyzed and the interactions between the viewpoints are listed. The interaction matrix in Example 3 - Table III is then derived from the defined viewpoint interactions.

1) R1 Viewpoint Interaction Analysis

Library staff informs the new library members about the services offered in the library. Library staff and member involve in a direct interaction with each other and therefore direct relationship exists for this requirement.

2) R2 Viewpoint Interaction Analysis

Library member requests the library staff for a book available in another library. Library Staff informs the Library Manager about the request, who in turn refers to the library policy. If available, the library manager arranges for the book and ensures that the member receives it through the library staff. Here, the library member and library manager interact through another viewpoint, that is the library staff and therefore they maintain an indirection relationship. Library member and staff maintain a direct relationship because of direct interaction. The Manager maintains a direct interaction with the library policy since he refers to it.

TABLE III. VIEWPOINT INTERACTION MATRIX

	Bank Customer	**Bank Employee**	**Bank Manager**	**Bank Policy**	**ATM Operator**
Bank Customer		R1 D R2 D R3 D R4 D R5 D R6 D R7 D	R1 I R6 I	R1 D R6 D	R4 I
Bank Employee	R1 D R2 D R3 D R4 D R5 D R6 D R7 D		R6 D	R1 I R6 I	R4 D
Bank Manager	R1 I R6 I	R6 D		R1 D R6 D	
Bank Policy	R1 D R6 D	R1 I R6 I	R1 D R6 D		
ATM Operator	R4 I	R4 D			

TABLE IV. EXTENDED CASE STUDY REQUIREMENTS

Requirement Name	Description
R1	Library staff, library member.
R2	Library staff, library member, library manager, library policy.
R3	Library staff, library member, library manager, library policy.
R4	Library staff, library member, library database.
R5	Library staff, library member, library IT administrator.
R6	Library staff, library member, library database.
R7	Library member.
R8	Library member, library policy.

TABLE V. REQUIREMENT TO VIEWPOINT MAPPING

Requirement Name	Description
R1	Library staff informs new library members services offered in the library.
R2	Library members request books from another library.
R3	Library manager approves library members for checking out books more than prescribed limit.
R4	Library members informs non availability of the books to the staff
R5	Library members notifies library staff problem in library website.
R6	Library staff updates member record in the database after member pays back late fee.
R7	Library member checks in the book.
R8	Library member talking in his/her phone in silent place of library.

3) R3 Viewpoint Interaction Analysis

Library member requests the library staff to check out more books than the prescribed limit. Library staff informs the above issue to the manager who in turn refers to the library policy and allows the member to checkout more books if library policy permits it under special circumstances. Here, the library member and manager maintain an indirect relationship whereas the library manager and library policy, library member and library staff, maintain a direct relationship. Since the policy is referred to, to clarify the doubt of the library member, the member and staff indirectly interact with the policy as well.

4) R4 Viewpoint Interaction Analysis

Library member informs the non availability of the book to the staff. Library staff checks in the library database about the availability and informs the member about the availability of the book. The library member and staff thus have a direct interaction. Here, the library member and library database have an indirect relationship because they interact through the library staff.

5) R5 Viewpoint Interaction Analysis

Library member notifies the problem in the library website to the library staff who in turn contacts the Library IT administrator to resolve the issue. Here, the library member and IT administrator interact indirectly through the library staff and thus maintain an indirect relationship.

6) R6 Viewpoint Interaction Analysis

Library staff updates the member record in the database after the member pays back the late fee. Here, library member and database involve in an indirect interaction. The Library member pays the late fee directly to the staff and therefore the relationship is direct.

7) R7 Viewpoint Interaction Analysis

Library member checks in the book. Here, the only viewpoint is the library member, who in this case does not interact with any other view points. This requirement was put in place to show that not all requirements contain viewpoint interaction.

8) R8 Viewpoint Interaction Analysis

Library member talking over his/her phone in the silent study place of the library is the compound requirement because requirement's viewpoints such as member and policy do not interact with each other. In order to resolve this issue, the compound requirement needs to be split into multiple requirements. The above requirement can be split into the following:

Library member talking outside library.	Library member.
Maintaining silence in silent study place.	Library policy.

VI. ANALYSIS

The lack of explicit support for viewpoint interaction is listed in the limitations section of the original proposal as an area for further research. Our extension to VORD provides that direct support for viewpoint interaction. Due to the complex nature of requirements engineering, viewpoint interaction is a common occurrence. Therefore, our proposal strengthens the VORD model's ability to cope with a problem that it previously could not directly address.

By providing direct support for viewpoint interaction, our model extension would be useful for automating legacy systems. For example, a bank may wish to lower operating costs by automating certain customer operations. By analyzing the viewpoint interaction matrix, all interactions to the bank employee viewpoint are clearly defined. The bank may choose to automate certain operations provided by the employee such as the notification to the bank manager in R6. A possible cost effective solution may be used to provide a web-based user interface that allows the bank customer to send the relevant information to the bank manager. This can be observed in terms of the Library System example as well; by analyzing the viewpoint interaction matrix, all interactions between each of the viewpoints are clearly defined. The interaction matrix gives an overall view of the direct and indirect interactions between the viewpoints. By analyzing the interaction matrix, the library may choose to automate certain operations that would aid in reducing the number of interactions yet perform the required task. For instance, the library may choose to automate operations such as the interactions between the library employee and library manager in R2 and R3. This interaction is merely for a notification purpose and hence can be done through other means such as sending an e-mail. By automating such actions, the entire system can be reduced of a number of interactions. Thus overall, we see the library system uses 11 direct interactions and 8 indirect interactions between its viewpoints. By our simple analysis, we observed that automation of certain processes could reduce the number of interactions, making the system run more efficiently and in a less complex manner. The proposed extended VORD process model simplifies the entire process; it not only links the viewpoints but also presents the interaction matrix.

This model would also be useful when analyzing the effect of modifying legacy systems. It could help to determine which viewpoints may be affected if a specific requirement is modified. At the very least, it would list the viewpoints to re-analyze.

Another result of our extension is the ability to expose probable compound requirements. By analyzing the type of interaction between viewpoints (direct/indirect), it can be determined if two viewpoints would not interact at all within a certain requirement. This may be an indication that the specified requirement may actually contain multiple requirements. The remedy would be to specify the requirement further into two or more requirements. This effect is limited to those requirements that have two or more viewpoints. Our extension will not indicate if a compound requirement is specified for requirements that only affect one viewpoint.

VII. LIMITATIONS AND FUTURE WORK

The original VORD model was deliberately restricted to a service-oriented view of systems. Therefore any restrictions of the service-oriented systems (SOS) paradigm will also affect our extension of VORD. The VORD authors did not consider this to be a serious limitation, as they believed that most systems can be regarded as providing services of some kind to their environment.

One limitation of the original VORD model that was not

TABLE VI. VIEWPOINT INTERACTION MATRIX

	Library Member	**Library Staff**	**Library Manager**	**Library Policy**	**Library Database**	**Library IT Administrator**
Library Member		R1 D R2 D R3 D R4 D R5 D R6 D	R2 I R3 I	R2 I R3 I	R4 I	R5 I
Library Staff	R1 D R2 D R3 D R4 D R5 D R6 D		R2 D R3 D	R2 I R3 I	R4 D R6 D	R5 D
Library Manager	R2 I R3 I	R2 D R3 D		R2 D R3 D		
Library Policy	R2 I R3 I	R2 I R3 I	R2 D R3 D			
Library Database	R4 I	R4 D R6 D				
Library IT Administrator	R5 I	R5 D				

addressed with our extension is the control issues associated with concurrency. The VORD model addresses the process of requesting and responding to services as a linear flow. However, it does not address services provided concurrently to separate entities at the same time. Since, we did not address this issue with our model extension, this limitation still exists.

Since the practical examples continue upon the ATM machine case study, we cannot accurately predict at this time how our model extension will work with other types of systems. However, we feel that the proposed extension should work with other types of systems that use the SOS paradigm.

This paper does not address conflict resolution with viewpoint interaction. That is, if any viewpoint interactions resulted in a conflict, that conflict would need to be resolved when specifying the requirements. The original VORD proposal directly addressed conflict analysis; however, it was limited to conflicts within a single viewpoint not directly addressing conflicts across viewpoints. We recommend that further research be performed in this area. A possible place to start is with goal-oriented analysis [8]. Goal-oriented analysis provides a mechanism for finding alternatives in requirements. It also provides a method for "weighing" each alternative to determine which one best fits the customer's goals. This is useful for finding a "middle ground" (skewed toward the customer's needs) when conflicts occur.

VIII. CONCLUSIONS

The VORD model ensures that system requirements rather than high-level system specification or designs are derived [1]. The model is highly regarded in requirements engineering as demonstrated by the frequency that the original proposal is cited. The authors of the VORD model expressed the lack of viewpoint interaction analysis as a limitation of the model. We used this limitation as a starting point for our research. We devised an extension to the VORD model by providing a method to explicitly support viewpoint interaction. We then demonstrated the practicality of this method by applying it to several examples.

REFERENCES

[1] G. Kotonya and I. Sommerville. Requirements Engineering with Viewpoints. Software Engineering Journal 1996.

[2] G. Kotonya. Practical Experience with viewpoint-Orientated Requirements Specification. Requirements Engineering Journal 1997.

[3] J. Araújo and P. Coutinho. Identifying Aspectual Use Cases Using a Viewpoint-Oriented Requirements Method, Early Aspects 2003: Aspect-Oriented Requirements Engineering and Architecture Design. Workshop of the 2nd International Conference on Aspect-Oriented Software Development, Boston, USA, 17th March 2003.

[4] I. Sommerville, P. Sawyer, and S. Viller Viewpoints for Requirements Elicitation: A Practical Approach. IEEE International Conference on Requirements Engineering (ICRE 1998).

[5] P. Lee. Viewpoints Model-based Requirements Engineering. 2002 ACM Symposium on Applied computing (2002).

[6] S. Greenspan, J. Mylopoulos, and A. Borgida. On Formal Requirements Modeling Languages: RML Revisited, Proceedings of the 16th international conference on Software engineering (1994).

[7] B. Nuseibeh and S. Easterbrook. The Process of Inconsistency Management: A Framework for Understanding, Proceedings of the First International Workshop on the Requirements Engineering Process (REP'99), 2-3 September 1999.

[8] J. Mylopoulos et al. Exploring Alternatives During Requirements Analysis, IEEE Software, January/February 2001, pp. 2 – 6.

[9] Gerald Kotonya and Ian Sommerville. Integrating Safety Analysis and Requirements Engineering, Proceeding of ICSC'97/APSEC'97, pp.259-270

[10] Zeljka Pozgaj. Requirement Analysis Using VORD, 22nd International Conference. Information Technology interfaces, June 13-16, 2000, Pula, Croatia. University of Zagreb, Croatia.

A Model for Enhancing Requirements Traceability and Analysis

Ahmed M. Salem
Department of Computer Science, California State University, Sacramento
Sacramento, CA 95819
916-278-3831 – Fax 916-278-6774
salema@ecs.csus.edu

Abstract—**Software quality has been a challenge since the inception of computer software. Software requirements gathering, analysis, and specification; are viewed by many as the principle cause of many of the software complex problems. Requirements traceability is one of the most important and challenging tasks in ensuring clear and concise requirements. Requirements need to be specified and traced throughout the software life cycle in order to produce quality requirements. This paper describes a preliminary model to be used by software engineers to trace and verify requirements at the initial phase. This model is designed to be adaptable to requirement changes and to assess its impact.**

Keywords- Requirements Traceability; Software Faults; Software Quality.

I. INTRODUCTION

Consistent and traceable software requirements are critical elements in today's complex software projects. Requirements include business requirements, user requirements, functional requirements, non-functional requirements, and process requirements. It is well documented such that most of the errors in software development occur in the requirements phase. With the complexity of software systems and the interdependencies of requirements, requirement traceability models and tools become very critical for improving software fault detection and the overall software quality.

Requirements traceability can be defined as the ability to describe and follow the life of a requirement, in both a forward and backward direction, i.e. from its origins, through its development and specification, to its subsequent deployment and use, and through periods of ongoing refinement and iteration in any of these phases. The requirements traceability is a characteristics of a system in which the requirements are clearly linked to their sources and to the artifacts created during the system development life cycle based on these requirements. [10]

It provides an efficient method for the detection of software faults, which are the static defects that occur due to an incorrect state or behavior of the system. Through traceability we can track which part of the code is linked to the requirements and which is not, this helps us remove discrepancies if any. These discrepancies if not detected can be really expensive at later stages and can lead to faults and failures.

The benefits of requirements traceability are the most obvious when the system changes. When high-level requirements change, it is implied that lower-level objects need to be changed. This issue alone justifies the need for requirements traceability. Testing and software quality also benefit greatly from requirements traceability. If a low-level requirement should fail during requirements testing, the software engineer would know which high-level requirements will not be met. Furthermore, if there is a defect, all of the segments that will be affected based on the requirements traceability can be identified, documented and reviewed.

II. BACKGROUND

Requirements engineering has played a vital role in the development of software in recent years. Ever since, requirements traceability has become an important issue as it can help provide answers to a variety of questions, such as: "Is the implementation compliant with the requirements?" or "Is the implementation even complete?" [2]. Many such questions can be answered depending on the completeness of the traceability links between the requirements and other software artifacts. However, in practice, a variety of traceability problems occur which generally include the use of informal traceability methods, failure in the cooperation between people responsible for coordinating traceability, difficulty in obtaining necessary information in order to support the traceability process, and lack of training of personnel in traceability practices.

To deal with such challenges and the additional burden of today's globally distributed development environment, some researchers have introduced an Event-Based method [1]. In this approach, the author proposes a methodology in which requirements and other software engineering artifacts can be linked through publish-subscribe relationships. This type of relationship is widely used in systems such as news service subscriptions and hardware management. In this Event-Based Traceability (EBT) system, requirements and other software artifacts that may induce changes are considered to be publishers while artifacts dependent on such changes act as subscribers. Hence the requirements are published and the performance models subscribe to the system.

A change in requirements will cause events to be published to an event server, which in turn will send out notifications to all dependent subscribers. The "publish-subscribe" model used

within EBT allows automatic linkage and validation of values within the requirements that are established with the EBT system. The EBT system then acts as an automated change notification agent. When changes are made, the affected artifacts and models are logged, and the developer can determine which artifacts to update, and which to ignore. This is done based on factors such as the criticality of the requirement changed. The developers, can then make the necessary changes. The major advantage of this system is its support for managing changes as well as its ability to support the identification and maintenance of artifacts. Another research project presented the Information Retrieval (IR) approach as the key to recover traceability links between code and documentation [3]. This research introduces a methodology where Information Retrieval techniques are used to link artifacts to each other as well as requirements, through a mechanism that queries the set of artifacts. Next, by using an indexing process, artifact information and semantics are parsed and used in order to rank each artifact on its relevance to a given query. The rankings are then used to create links between artifacts, which are returned to the user. The user can use the rankings in order to understand the relationships between artifacts and even requirements in order validate the links that have been generated by the system.[8] Establishing these traceability links can help support tasks such as: program comprehension, maintenance, requirement tracing, impact analysis and reuse of existing software[5]. The premise of this research relies on the probability that most programmers tend to use meaningful names for program items, such as: function names and variables. Concepts implemented in the code are suggested by the use of such names. By using such correlations, queries can be constructed from modules in the source code. This proposed model concludes that IR provides a practical solution to the problem of semi-automatically recovering traceability links. In a different research work titled "Rule-Based Generation of Requirements Traceability Relations", the authors address the challenges that arise when analyzing requirements [6]. This approach uses traceability rules to support the automatic generation of traceability relations. The generation of such relations is based on the *requirement-to-object-model* traceability rules. They help to trace the requirements and use case specification documents to an analysis object model, and *inter-requirements* traceability rules to trace requirement and use case specification documents to each other.[9] Throughout this approach, the authors focus on the challenge that requirements are expressed in natural language and often contain ambiguity.

Other new models have also been proposed to support the ideology of requirement traceability, one such model is the Conceptual Trace Model. It consists of three parts; A fine-grained trace model, which determines the types of documentation entities and relationships to be traced to support impact and implementation of system requirements changes; A set of process descriptions that describe how to establish traces and how to analyze the impacts of changes and the third part is tool support that provides (semi-) automatic impact analyses and consistency checking of implemented changes. [7] Our proposed model shares some of these concepts, but with a unique approach to requirement traceability.

III. PROPOSED TRACEBILITY MODEL

The proposed model is an extension to a previously suggested traceability model [4] which allows the software developer to achieve traceability at the source code level. The model focused on keeping track of the sets of working modules specific to satisfying the requirements. This model is the base for our extension and the new model thus offers a number of enhancements and features. There are two types of users for requirements traceability, high-end users and low-end users. Low-end users tend to consider traceability only because it is required, while high-end users recognize the importance of traceability [4]. This new model is simple for low-end users, yet comprehensive for high-end users.

It is composed of a Traceability Engine Component (TEC), a Traceability Viewer Component (TVC), and a Quality Assurance Interface (QAI). The first component, the TEC, is used to assist developers correlate source code elements with the software requirements. It functions by first reading in the requirements data from the requirements database, analyzes the source code and corresponding requirements, and creates its own internal traceability matrix. The TEC supplies this data to the QAI for evaluation, and is then updated with the results. The data that the TEC receives and the results of its own analysis are kept in a Traceability Database where it is accessible for re-evaluation at each stage of software development.

The TEC also contains an interface that enables the developer to indicate flags relating each piece of code or file to a specific requirement. When the code is checked into the CVS (Concurrent Version System), a version control system which is used to record source code history and documents, the TEC detects any change that has been made, and will prompt the developer to indicate the specific requirements related to each piece of code. Once all these relationships have been entered, the QAI is notified that there is data that needs to be verified. In addition, once the QAI has completed its process, the TEC will be able to determine which pieces of code do not have corresponding requirements and which requirements have no corresponding code.

The TVC acts as the 'client' portion of the proposed model. The TVC provides the software engineer with a unique way to view all the information that the TEC has gathered. It will have the ability to provide custom data such as: a detailed list of all requirements, reports regarding which requirements have been met and in which modules they are implemented, and the results of the verification and validation completed by the QAI.

The business analyst must insert the requirements into a spreadsheet, which is then imported into the database tables using a specialized tool. The interface added is called the Quality Assurance Interface (QAI), which a quality assurance specialist may use to verify that the code being checked meets the corresponding requirements. The importing of requirements and the QAI will be discussed in greater detail in the sections to follow.

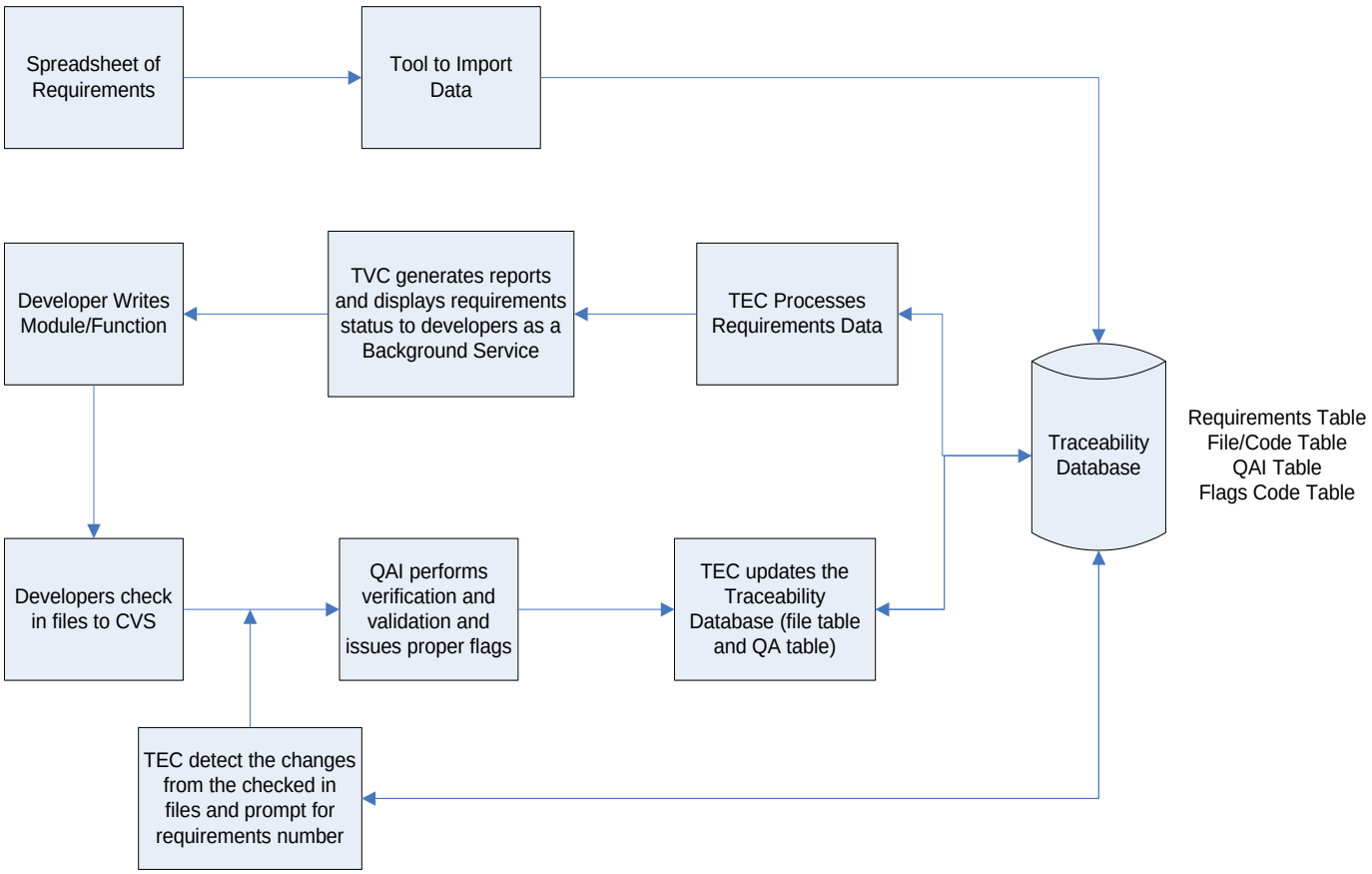


Figure 1: Proposed Traceability Model

TABLE 1: Requirements Flags

Flag Name	Description	Comments
NCF	Code Not Checked Flag	File name/line number and requirement set should be evaluated
CVF	Checked and Valid Flag	Signifies a good file name/line number and requirement set
CNF	Checked and Not Valid Flag	Developer of the code is working on his/her code to satisfy the requirement
CNFDATA	Checked and Not Valid Flag (DATA Error)	Developer of the code is working on his/her code to fix the data error and satisfy the requirement
CNFLOGIC	Checked and Not Valid Flag (Coding Bug)	Developer of the code is working on fixing a bug in his/her code to satisfy the requirement
CRF	Code with no Requirement Flag	There is a significant amount of code that is not assigned a requirement match
RCF	Requirement with no Code Flag	A certain requirement has not been met with any of the source code from the project
RCFF	Requirement Changed Flag	Indicates that the requirement has been modified
RRCF	Related Requirement Changed Flag	Indicates that a requirement related to this requirement has been changed

A. Quality Assurance Interface (QAI)

The proposed model provides a mechanism to address the issue of validating and verifying that the requirements are actually being met. This model allows the software engineer to choose either to add the tags directly to the source code, or, to choose which requirement is being met from a list of all requirements. The QAI will be able to perform aspects of verification and validation such as to "double-check" that a requirement has actually been met. The QAI performs a dual job, which is to insure that the requirements are accomplished, and to report requirements or parts of source code that do not match.

The table (See table 1) shows the requirements flags that are used in the QAI to indicate the status of the verification and validation of the requirements and the code. When the source code is checked into a version control system (CVS), a table will be populated with all the file names and line numbers which satisfy certain requirements. Each of these file name/line number and requirement matches are assigned a flag, NCF, which signifies that they have not been verified by the QAI. In addition to this, each requirement is assigned an NCF flag to indicate that it does not have any corresponding code that meets the requirement or that the validity of the correlation has not been verified. To insure that the requirements are truly met, the QAI will parse this table and go through a six step process to determine if the requirements have been met.

First, the QAI will look up the requirement from the database in the initial list of requirements that were given for the project. Second, it will read in the description of the requirement that has been linked with the requirements in the database. Third, the QAI will find the file and the line from the file name/line number combination from the match. Fourth, it will read and evaluate the source code. Fifth, it will determine if the match is a good match and if the requirement is actually met. Lastly, the sixth step is to create a flag for this file name/line number and requirement match that signifies that the match has been checked. Furthermore, it will indicate that either the requirement has been accomplished or it has not been sufficiently met. If it has not been met an email will be sent to the developer with the message that the source code should be re-done.

The second major task of the QAI is to handle the flags for the software requirements. This will help solve the problem of reporting which requirements or parts of the source code do not have matches. There are nine different flags that can be assigned; see Table 1. The first of these flags is the Not Checked flag (NCF). This signals to the QAI that the file name/line number and requirement set should be evaluated as outlined above. The second flag is the Checked and Valid flag (CVF); this signifies that a file name/line number and requirement set are valid. The third flag is the Checked and Not Valid flag (CNF); this means that the developer of the code is working on the code to satisfy the requirements. The fourth flag is the Checked and Not Valid Flag as a result of a Data Error (CNFDATA); this indicates that the developer needs to work on the code to fix the data error in order to fulfill the requirement. The fifth flag is the Checked and Not Valid Flag as a result of a coding bug (CNFLOGIC); this indicates that there is a bug in the code that needs to be resolved before the requirement can be met. The sixth flag is the Requirement with No Code flag (RCF); this signifies that a certain requirement has not been met with any of the source code from the project. This should signal that the requirement needs to be completed. The seventh flag is the Code with No Requirement flag (CRF); this indicates that there is a significant amount of code that is not assigned a related requirement has changed since it was imported into the database.

The QAI also has the ability to handle changes made to the requirements. When a requirement is changed or simply modified, the corresponding flag field in all records in the QAI table containing this requirement will be reset to the Requirement Changed Flag (RCHF). In addition, all requirements listed in the related requirements field will have their flag in the QAI reset to Related Requirement Changed Flag (RRCF). With the use of these two flags, when requirements change, only the code that could possibly relate to the requirements will need to be reviewed to ensure that it still satisfies the new requirement. This model enables changes to be made to the requirements without a need for all requirements and code to be rechecked. Only the changed requirement and requirements that relate to it need to have their corresponding code reviewed by the QAI.

B. Traceability Database Tables

The Traceability Database contains five tables. The Requirements Table (See table 2) is populated with the requirements that were imported from the business analyst's spreadsheet. The table will contain the following fields: the requirements key, which is the primary key for the table, the person adding the requirement, and a description of the requirement.

The related requirement field contains any requirements that are directly related to the listed requirement; any changes made to the requirement or its related requirements can be indicated in the QAI table (See table 5).

The Code/File Change Table (See Table 3) contains the following fields: the code key, which is the primary key for the table, the file path, the file name, the method name, the class name, the date the change was entered, and the name of the coder.

When the coder checks the code into requirements without a need for all requirements and code to be rechecked. Only the changed requirement and CVS, changes that have been made are tracked by assigning a new code key. For example, if Door.cs and abc.config, the new record will look like table 2. After the new record has been added, the TEC will prompt for the requirements key from the requirements table (Table 2).

TABLE 2: Requirements Table

RequirementKey (Primary Key)	AddedBy	Description	Related Requirements
1000	BA1	Display Name in the header.	1010
1001	BA2	Display Date.	1010
1002	BA1	Be able to select style.	1011
1003	BA1	To store names.	1012, 1013
1004	BA1	To change names.	1012, 1013
1005	BA2	To delete names.	1012, 1013

TABLE 3: Code/File Change Table

CodeKey (Primary Key)	FilePath	FileName	Method Name	Class Name	Date	Coded by
1000	\Main\Project1\	Display.cs	GetName()	Display	10/1/06	Coder1
1001	\Main\Project1\	Render.cs	GetLineNumber()	Render	10/12/06	Coder2
1002	\Main\Project2\	Cashier.cs	NULL	Cashier	10/13/06	Gary
1003	\Main\Project1\	Interface.cs	NULL	Interface	10/12/06	Coder4
1004	\Main\Project3\	Window.cs	NULL	Window	10/13/06	Coder5
1005	\Main\Project5\	Door.cs	Main()	Door	10/18/06	Gary
1006	\Main\	abc.config	NULL	NULL	10/18/06	Gary

The Code Flags table (See Table 4) contains fields for the Flag Key, which is the primary key, the flag name, the description of the flag, and the purpose of the flag. When flags are assigned in the QAI table (See table 5), the Flag Key is used; the Flag Name is used for display purposes only. In addition, by using the key in place of the name, it will be more functional for purposes such as grouping records. The QAI can also add custom-defined code flags to the system.

The QAI table (Table 5) is a link between the Requirements records and the Code/File records. After a new record has been added to the Code/File Change table, the TEC will prompt for a requirement key to correspond to the entry. Once it has been entered, the TEC will create a new record in the QAI table to show which files or lines of code correspond to which requirements. he fields in this table are: the QAIKey, the primary key for the relationship; the requirements key, a foreign key to the Requirements table; the Code/File Key, a foreign key to the Code/File Change Table; the Flag Key, a foreign key to the Code Flags Table (discussed later); the date the record was entered into the database; and the person who performed the QA. Based on table 2, once the TEC has received the requirements key input, a table will be displayed as in table 4. The QA analyst then performs the verification and validation, and the flags will be assigned to each record through the QAI interface. After this has been completed, the TEC gathers the data and imports it into the QAI table as reflected in table 6.

TABLE 4: Code Flags Table

FlagKey (Primary Key)	Flag	Description of Flag	Purpose
1000	NCF	Code Not Checked Flag	File name/line number and requirement set should be evaluated
1001	CVF	Checked and Valid Flag	Signifies a good file name/line number and requirement set
1002	CNF	Checked and Not Valid Flag	Developer of the code is working on his/her code to satisfy the requirement
1003	CNFDATA	Checked and Not Valid Flag (DATA Error)	Developer of the code is working on his/her code to satisfy the requirement
1004	CNFLOGIC	Checked and Not Valid Flag (Coding Bug)	Developer of the code is working on his/her code to satisfy the requirement
1005	CRF	Code with no Requirement Flag	There is a significant amount of code that is not assigned a requirement match
1006	RCF	Requirement with no Code Flag	A certain requirement has not been met with any of the source code from the project
1007	RCHF	Requirement Changed Flag	Indicates that the requirement has been modified
1008	RRCF	Related Requirement Changed Flag	Indicates that a requirement related to this requirement has been changed

TABLE 5: QAI Table Default Entry

QAIKey (foreign key to Requirements Table)	RequirementKey (foreign key to Requirements Table)	Code/File (foreign key to the Code File Table)	Flag (foreign key to Code Flags Table)	Date	QA by
1001	1001	1001	1000 (NCF, Default)	10/1/06	QA1
1002	1001	1002	1000 (NCF)	10/12/06	QA1
1003	1001	1003	1000 (NCF)	10/13/06	Gary
1004	1001	1004	1000 (NCF)	10/12/06	QA1
1005	1002	1004	1000 (NCF)	10/13/06	QA2
1007	1003	1005	1000 (NCF)	NULL	NULL
1008	1004	1006	1000 (NCF)	NULL	NULL

TABLE 6: QAI Table after QAI is Performed

RequirementKey (foreign key to Requirements Table)	Code/File (foreign key to the Code File Table)	Flag (foreign key to Code Flags Table)	Date	QA by
1001	1001	1000 (NCF, Default)	10/1/06	QA1
1001	1002	1000 (NCF)	10/12/06	QA1
1001	1003	1000 (NCF)	10/13/06	Gary
1001	1004	1000 (NCF)	10/12/06	QA1
1002	1005	1000 (NCF)	10/13/06	QA2
1003	1005	1001 (CVF)	10/19/06	Dan
1004	1006	1003 (CNFDATA)	10/19/06	Dan
1005	NULL	1006 (RCF)	10/19/06	Dan

The TEC also has the ability to indicate which pieces of code do not have corresponding requirements and which requirements do not have corresponding code. The TEC takes the keys from the Code/File Change table and the Requirements table and verifies that each exists in a record in the QAI table. If they do not exist, then a record is created in the QAI table with a null value in the field of the item that does not exist, as in the last record in table 6.

C. Populating the Database

Another advantage of this model is that it considers the initial state of the database as shown in figure 1. But the question that follows is: how does the database get populated initially? Using a tool to fill the database with requirements can be one technique. Most current Database Management Systems such as SQL Server and Oracle, have a specialized tool to allow the importation of data from other types of file formats. One example is the Data Transformation Service (DTS) that is available with the more recent versions of SQL Server. This service allows the user to import data using a convenient user interface in the form of a wizard. Even though this wizard is convenient, it is not likely that the business analysts will have direct access to the database. Therefore they will not be able to use this wizard and must list the requirements in an organized manner. The business analysts will have to give the data to the developers in a format that can be imported into the database.

In general, business analysts are skilled in working with spreadsheets. Therefore, the business analysts should list the requirements in a spreadsheet with each row containing one requirement. This spreadsheet will be comprised of the following columns: requirement number, requirement name, description of the requirement, and related requirements. The requirement number must be unique and can serve as a key. This number will be chronological and the business analysts will be given a block of numbers from the developer that he or she will be able to use for that particular spreadsheet. The related requirements column will give the requirement number for any requirements that are related to it. This list of requirement numbers must be a set of valid values that are separated by commas. This would make it simple for the developer to parse these fields and extract the requirement numbers. The columns in this spreadsheet must be named as follows:

- RequirementKey,
- AddedBy
- RequirementName,
- Description,
- RelatedReqNumbers

The columns must have these names to maintain database consistency and to allow the tool to recognize which column in the spreadsheet corresponds to which column in the database. There are tools available for SQL Server, Oracle, MySQL, and Access to import data from Excel spreadsheets, therefore spreadsheets created in Microsoft Excel would be the most versatile.

IV. CONCLUSION AND FUTURE WORK

As previously noted, requirements traceability in the early stages plays a crucial role in the software development lifecycle. This model provides a very intuitive and dynamic way of requirements traceability. It provides a formal and measurable process to carry out traceability which can really be critical in exposing the defects at very early stages of the lifecycle. The suggested model amalgamates the features of the event based tracking and information retrieval tracking and adds new features in the design which makes it a very efficient method for requirement traceability. However, in order for this

approach to be successful it does require commitment and support from the quality assurance group. The proposed model will prove beneficial to the software engineer and the software quality assurance process and will help in optimizing the process as a whole. The TEC, TVC, and QAI components provide a very efficient way of tracking and tracing requirements which can be quite tedious to detect considering the complex nature of requirements and its relationship. The Quality Assurance Interface can facilitate the verification of the code to the requirements by following a six-step process that we have prescribed. The process ensures all the changes in the requirements are accessed thoroughly and their impact are foreseen clearly. Furthermore, a mechanism to import requirements into a database was outlined with detailed account of all the objects that will be used such as the tables . Finally, a flagging procedure is designed using Requirements Flags to provide traceability between the requirements and the source code. The flagging procedure clearly demarcates which parts of the code is linked to the requirements and which are not. This preliminary model provides a simple interface that allows developers to seamlessly locate the correct requirements and link them to the correct source code elements, thus providing a very dynamic and intuitive method of requirement traceability during the software development process.

Several directions for future work are possible. First and foremost, a tool implementing this model and its corresponding database will be useful in determining the feasibility of the proposed system. Case studies need to be conducted to further evaluate the effectiveness of such an approach. Further add-ons to the TEC and TVC can be done to make it more flexible and generic. The database can be further developed to accommodate more flags and features that helps in more detailed description of mapping attributes. The QAI can be further developed to be more dynamic and effective. Finally, it will be important to incorporate the tracing of software design documentation into this traceability model. The ultimate goal will be to provide traceability over every software artifact of the software development lifecycle.

REFERENCES

[1] Cleland-Huang, J., Chang, C.K., and Christensen, M. “Event-based traceability for managing evolutionary change”. IEEE Transactions on Software Engineering, Volume: 29, Issue: 9, Sept. 2003, Pages: 796 – 810.

[2] Requirements Tracing – An Overview. Retrieved on March 9, 2005, from http://www.sei.cmu.edu/str/descriptions reqtracing_body.html

[3] Antoniol, Giuliano, Gerardo Canfora, Gerardo Casazza, Andrea De Lucia, and Ettore Merlo.“Recovering Traceability Links between Code and Documentation”. IEEE Transactions on Software Engineering, VOL. 28, NO. 10, OCTOBER 2002

[4] Ahmed Salem and Johnny Lee “Requirements Traceability Model for the Coding Phase”, Conference on Computer Science, Software Engineering, Information Technology, E-Business and Applications, 2004

[5] Ramesh, Balasubramaniam. “Factors Influencing Requirements Traceability Practice”. Association for Computing Machinery, Inc., 1998.

[6] Spanoudakis, George, Andrea Zisman, Elena Perez- Minana, and Paul Krause. “Rule-based generation of requirements traceability relations”. Software Engineering Group, Department of Computing, City University, Northampton Square, August 2003.

[7] Mirka Palo. "Requirements Traceability", Department of Computer Science, University of Helsinki, October 2003.

[8] Grant Zemont. “Towards Value-Based Requirements Traceability”, DePaul University, Chicago Illinois, March 2005 .

[9] George Spanoudakis, Andrea Zisman, Elena Pérez-Miñana and Paul Krause. **“**Rule-based generation of requirements traceability relations”, Journal of Systems and Software, July 2004.

[10] Gotel and A. Finkelstein. “An Analysis of the Requirements Traceability Problem,” Proceedings of the First International Conference on Requirements Engineering, Colorado Springs, April 1994

The Result Oriented Process for Students Based On Distributed Data Mining

Pallamreddy.venkatasubbareddy[1]
Associate Professor, CSE Dept
QIS College of Engg & Technology
Ongole, India

Vuda Sreenivasarao[2]
Professor & HOD, CSE Dept
St.Mary's College of Engg & Technology,
Hyderabad, India.

***Abstract* - The student result oriented learning process evaluation system is an essential tool and approach for monitoring and controlling the quality of learning process. From the perspective of data analysis, this paper conducts a research on student result oriented learning process evaluation system based on distributed data mining and decision tree algorithm. Data mining technology has emerged as a means for identifying patterns and trends from large quantities of data. It is aimed at putting forward a rule-discovery approach suitable for the student learning result evaluation and applying it into practice so as to improve learning evaluation of communication skills and finally better serve learning practicing.**

Key words: Learning result evaluation system; distributed data mining; decision tree.

I. INTRODUCTION

With the accelerating development of society and the well-known knowledge explosion 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. For a long time, people limited learning to the transfer of culture and knowledge; research on learning was confined to the fields of education research within the existing traditions of classroom learning. To understand learning within the new context of the 21st century, we need professionals from psychology, sociology, bran science, computer science, economics, to name a few. We must extend out understanding about human learning from macro levels to micro levels and from history to current conditions. At present, the most urgent need is to synthesize all the findings on human learning and integrate them into a united framework to guide the practice of learners. A research on result oriented learning process evaluation from the perspectives of philosophy and culture has emerged as a major challenge to educators, and invent a new and integrated learning culture.

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 following university 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. Education reform for the 21st century has generated various models of learning result evaluation that have emerged over time.

The assessment of student learning is an essential component of University's efforts in evaluating overall institutional effectiveness. The value of research on student learning result evaluation system is to help teachers and students surpass the ivory-towered and alienating traditional classroom teaching model, and face the rapidly developing real-life environment and the ill-structured learning environment and adapt to current teaching realities. Distributed Data Mining (DDM) aims at extraction useful pattern from distributed heterogeneous data bases in order, for example, to compose them within a distributed knowledge base and use for the purposes of decision making. A lot of modern applications fall into the category of systems that need DDM supporting distributed decision making. Applications can be of different natures and from different scopes, for example, data and information fusion for situational awareness; scientific data mining in order to compose the results of diverse experiments and design a model of a phenomena, intrusion detection, analysis, prognosis and handling of natural and man-caused disaster to prevent their catastrophic development, Web mining ,etc. From practical point of view, DDM is of great concern and ultimate urgency.

The remaining sections of the paper are organized as follows. In Section II we describe the distributed data mining. . In Section III we describe decision tree algorithm. In Section IV we describe Student Learning Result Evaluation System and verify the efficiency of this system. Section V concludes the paper.

II. DISTRIBUTED DATA MINING

Data mining technology has emerged as a means for identifying patterns and trends from large quantities of data. Data mining and data warehousing go hand-in-hand: most tools operate on a principal of gathering all data into a central site, then running an algorithm against that data (Figure 1). There are a number of applications that are infeasible under such a methodology, leading to a need for distributed data mining.

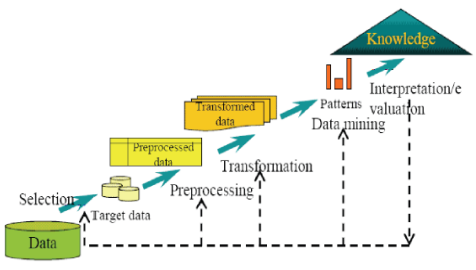


Fig 1 the transaction from raw data to valuable knowledge.

Distributed Data Mining Distributed data mining (DDM) considers data mining in this broader context. As shown in figure(2), objective of DDM is to perform the data mining operations based on the type and availability of the distributed resources. It may choose to download the data sets to a single site and perform the data mining operations at a central location.

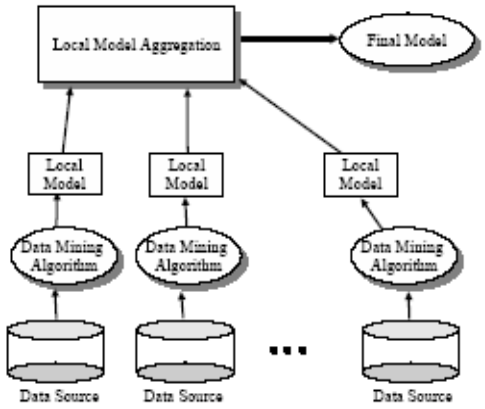


Fig (2). Distributed Data mining Frame work

Data mining is a powerful new technology with great potential to help companies focus on the most important information in the data they have collected about the behavior of their customers and potential customers. Data mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data set. These tools can include statistical models, mathematical algorithm and machine learning methods. It discovers information within the data that queries and reports can't effectively reveal.

III. DECISION TREE ALGORITHM

A decision tree is a classification scheme which generates a tree and a set of rules, representing the models of deferent classes, from a given data set. The set of records available for developing classification methods is generally divided in to two disjoint subsets-a training set and a test set. The decision tree approach is most useful in classification problems, with this technique, a tree is constructed to modal the classification process.

Decision tree is a decision support tool in the field of data mining that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility . Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal. Another use of decision trees is as a descriptive means for calculating conditional probabilities. Each node in a decision tree represents a feature in an instance to be classified, and each branch represents a value that the node can assume. Instances are classified starting at the root node and sorted based on their feature values.

A .Definition:

Given a data base D ={t_1,.....t_n}where

ti ={ti1,……tih} and the data base schema contains the following attributes {A1,A2,A3,……Ah}.Also given is a set of classes C={C1,C2,C3……..Cm}.

A decision tree (DT) or classification tree is a tree associated with D that has the following properties:-

Each internal node is labeled with an attribute, Ai

Each arc is labeled with a predicate that can be applied to the attribute associated with the parent.

Each leaf node is labeled with a class, Cj

The basic algorithm for decision tree induction is a greedy algorithm that constructs decision trees in a top-down recursive divide-and-conquer manner. The algorithm is shown in fig.3.

create a node N

if

samples are all of the same class C

then

return N as a leaf node labeled with the class C

if

attribute-list is empty

then

return N as a leaf node labeled with the most common class

in samples

label node N with test-attribute

for each known value a_i of test-attribute

grow a branch from node N for the condition test-attribute=a_i

let s_i be the set of samples for which test-attribute= a_i

if

s_i is empty

then

attach a leaf labeled with the most common class in

samples

else

Attach the node returned by Generate_ decision_ tree

(s_i, attribute-list_ test-attribute)

Fig. 3 Algorithm for decision tree .

IV. RESULT ORIENTED LEARNING EVALUATION SYSTEM FOR STUDENT

Result Oriented Learning Evaluation System for Student (ROLES) is composed of five parts. Data Collection Module collects data. Data Processing Module converts rude data to the regular mode. Data Analysis Module use Decision Tree to compute regular data and give the corresponding result. Graph Visualizing Module is used to show data analysis result through graphic mode. Database is used to stored rude

data, regular data and data analysis result. The structure of is shown in fig.4.

Fig.5 shows the Decision Tree used in ROLES. Each internal node tests an attribute, each branch corresponds to attribute value, each leaf node assigns a classification.

Fig.4 shows the accuracy of decision tree learning.

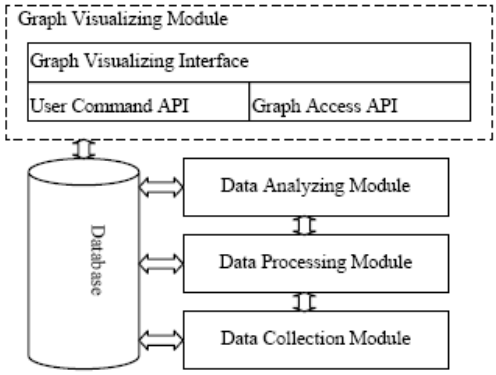


Fig.4: Structure of ROLES.

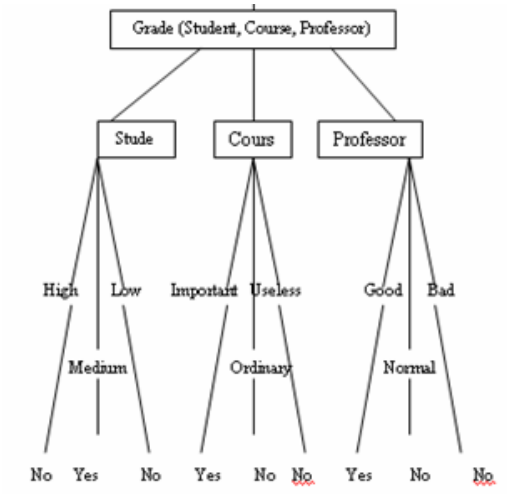


Figure No- 5 Decision tree used in ROLES.

No	Student	Course	Professor	Grade	Result
1	Ramu	DW	Srinivas	70	No
2	Ramya	DB	Swamy	69	No
3	Srinu	MS	Trirumala	95	Yes
4	Smith	DBMS	Raghavendra	97	Yes
…….	…….	…….	……	…….	……
n	Jone	CO	Prathibha	65	No

Table 1: Form of training examples

Table 1 shows the form of training data. 1000 student score records are used for training.

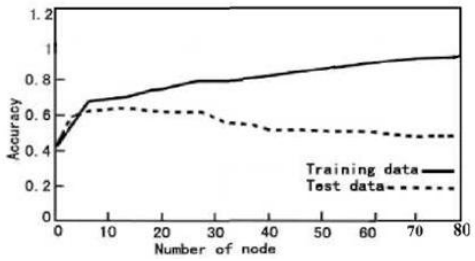


Figure 6: Accuracy of Decision tree learning.

V. CONCLUSION

The Student result oriented learning process system for assess learning environments is an image of what makes sense today. As time goes on, new features will be added and others dropped. With this model in practice, student learning can become more energetic, more interesting, more challenging, and more suited to the times.

REFERENCES

[1] Sally A. Gauci, Arianne M. Dantas, David A. Williams, and Robert E. Kemms, "Promoting Student-Centered Active Learning in Lectures with a Personal Response System," Advances in Physiology Education, vol. 33 (March 2009), pp. 60–71.

[2] Margie Martyn, "Clickers in the Classroom: An Active Learning Approach," EDUCAUSE Quarterly, vol. 30, no. 2, (April–June 2007), pp. 71–74.

[3] Y. Lindell and B. Pinkas Secure Multiparty Computation for Privacy-Preserving Data Mining Journal of Privacy and Confidentiality, Vol. 1, No. 1, pp. 59-98, 2009.

[4] T.C. Fu and C.L Lui, "Agent-oriented Network Intrusion Detection System Using Data Mining Approaches," International Journal on Agent-Oriented Software Engineering, Vol. 1, No. 2, pp. 158-174, 2007.

[5] Sang,X.M.: On the Edge of Learning Theory and Practice in the Digital Age. China Central Broadcasting and TV University Press, Beijing, 2000.

[6] R.A.Burnstern and L.M. Lederman, "Comparison of Different Commercial Wireless Keypad Systems", The Physics Teacher, Vol.41, Issure 5, 2003, pp.272-275.

[7] D.English, "Audiences Talk Back: Response Systems Fill Your Meeting Media with Instant Data", AV Video Multimedia Producer, Vol.25, No.12, 2003,pp.22-24.

AUTHOR PROFILE

PALLAMREDDY.VENKATA SUBBA REDDY

received his M.Tech degree in Computer Science & Engg from the Satyabama University, in 2007 and Master of Science in Information Technology from Madras University with First class in may 2003 .Currently working as Assoc. Professor in CSE Dept in QIS College of Engg & Technology, Ongole. His main research interests are Data Mining and Network Security. He has got 7 years of teaching experience. He is currently pursuing the PhD degree in CSE Depart in Bharath University, Chennai.

VUDA SREENIVASARAO received his M.Tech degree in Computer Science & Engg from the Satyabama University, in 2007.Currently working as Professor & Head in the Department of CSE & IT at St.Mary's college of Engineering & Technology, Hyderabad, India.. His main research interests are Data Mining, Network Security, and Artificial Intelligence. He has got 10years of teaching experience .He has published 20 research papers in various international journals. He is a life member of various professional societies like MIACSIT, MISTE.

An New Filtering Methods in the Wavelet Domain for Bowel Sounds

Zhang xizheng[1],
[1]School of Computer and Communication
Hunan Institute of Engineering
Xiangtan China

Rui Yuanqing[2], Wang weixiong[1]
[2] School of Physics
South China University of Technology
Guangzhou ,China

***Abstract*—Bowel sounds signal (BS) is one of the important human physiological signals, analysis of the BS signal can then study gastrointestinal physiology and implement direct and effective diagnosis of gastrointestinal disorders. Use different threshold denoising methods denoising the original bowel sounds, simulated in the environment of MATLAB, then compared the denoising results and analyzed the advantages and disadvantages of those threshold denoising methods.**

Keywords-bowel sounds; de-noising; wavelet transform; threshold

I. INTRODUCTION

Bowel sounds signal (BS) is the human digestive organs particularly the bowel movements and intestinal narrowing of energy in the intestinal wall pressure release form, which produces small vibration in the abdominal body surface. Through the sensing device of the abdominal skin vibration (ASV) system can capture these quivers. Analysis of the BS signal can then study gastrointestinal physiology and offer effective diagnosis for gastrointestinal disorders directly. However, there are many difficulties when extract BS signal from the data captured by ASV. As the body's own signal is weak, the signal is very susceptible to be contaminated by other signal source. In the case of sensor sensitivity and signal amplification and conditioning circuit performance is limited, the difficulties and key to extract the true BS signal is to reduce the extent of noise pollution or even eliminate the noise [1-5].

To solve this difficult problem, the conventional method of signal filtering is often ineffective or powerless. Wavelet transform is a new signal processing tools which developed over a decade, because of its time and frequency domain also has good localization properties, multi-resolution analysis, "centralized" in nature and easy to combine with auditory and visual characteristics of human, therefore, wavelet is very suitable for non-stationary signals such as separation of signal and noise. The technology of BS signal filtering in wavelet domain is not mature, and left a lot of work and issues to be resolved. Specific to the country, it has not seen the relevant technical literature and reports appear. Therefore, from the theoretical and academic point of view, study the formation mechanism, signal characteristics and filtering techniques in wavelet domain of BS this complex and highly unstable medical signal, can greatly enrich the theory of wavelet filtering itself, and is one of the most likely to occur theoretical breakthrough and technological innovation in the next several years[6-9].

This paper using several methods to implement filtering for the collected BS signal in the wavelet domain, through qualitative and quantitative analysis, the optimal filtering method is proposed, with this kind of filtering method implement filtering can eliminate noise and retain the useful component effectively, and recovered the features of BS signal and can provide some help for the clinical diagnosis.

II. BASIC THEORY OF WAVELET ANALYSIS

A. *Continuous Wavelet Transform(CWT)*

$\forall f(t) \in L^2(R)$, the CWT of $f(t)$ is defined as [10,11]:

$$WT_f(a,b) = |a|^{-1/2} \int_{-\infty}^{\infty} f(t) \overline{\psi\left(\frac{t-b}{a}\right)} dt, \quad a \neq 0 \qquad (1)$$

Or use its inner product form:

$$WT_f(a,b) = \langle f, \psi_{a,b} \rangle \qquad (2)$$

In the above formula $\psi_{a,b}(t) = |a|^{-1/2} \psi\left(\frac{t-b}{a}\right)$

In order to make inverse transform exist, $\psi(t)$ must meet the below requirement:

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty \qquad (3)$$

In the above formula, $\hat{\psi}(\omega)$ is the FT of $\psi(t)$.

Then, the inverse transform is

$$f(t) = C_\psi^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi_{a,b}(t) WT_f(a,b) db \frac{da}{|a|^2} \qquad (4)$$

In MATLAB, we can use the function of *cwt* to implement CWT for continuous signal [2].

B. Discrete Wavelet Transform(DWT)

In practice, especially when implemented on a computer, continuous wavelet transform must be discrete. So, it is necessary to discuss the discretization of continuous wavelet $\boldsymbol{\Psi}_{a,b}(t)$ and continuous wavelet transform $W_f(a,b)$. Need to stress that the above-mentioned discretization is towards continuous scale parameter *a* and continuous translation parameter *b*, but not time parameter *t*.

In the continuous wavelet, consider the function:

$$\boldsymbol{\Psi}_{a,b}(t) = |a|^{-1/2}\boldsymbol{\Psi}(\frac{t-b}{a}) \tag{5}$$

And $b \in R, a \in R^+$, $a \neq 0$, $\boldsymbol{\Psi}$ is allowed, for convenience, *a* must be limited to positive in discretization, and now the compatibility requirement can be represented as

$$C_\Psi = \int_0^\infty \frac{|\hat{\boldsymbol{\Psi}}(\varpi)|}{|\varpi|} \mathrm{d}\varpi < \infty \tag{6}$$

Generally, the discretization formula of continuous wavelet transform scale parameter *a* and translation parameter *b* can be taken as $a = a_0^j$ and $b = ka_0^j b_0$, and $j \in Z$,expansion step $a_0 \neq 1$ is a fixed value, for convenience , always assume that $a_0 > 1$ .So the corresponding discrete wavelet function $\boldsymbol{\Psi}_{j,k}$ can be writed as

$$\boldsymbol{\Psi}_{j,k}(t) = a_0^{-j/2}\boldsymbol{\Psi}(\frac{t-ka_0^j b_0}{a_0^j}) = a_0^{-j/2}\boldsymbol{\Psi}(a_0^{-j/2}t - kb_0) \tag{7}$$

And discretization wavelet transform coefficient can be expressed as

$$C_{j,k} = \int_{-\infty}^{\infty} f(t)\boldsymbol{\Psi}_{j,k}^*(t)\mathrm{d}t = \langle f, \boldsymbol{\Psi}_{j,k} \rangle \tag{8}$$

The corresponding reconstruction formula is

$$f(t) = C\sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty} C_{j,k}\boldsymbol{\Psi}_{j,k}(t) \tag{9}$$

C is constant independent of $f(t)$.

In MATLAB, the function of *dwt* and *dwt*2 can use for DWT.

C. Multi-resolution Analysis

In order to understand multi-resolution, we will choose a decomposition of three to clarify it; the wavelet decomposition tree is shown as Figure 1.

It is clear from Figure 1 that the further decomposition of multi-resolution analysis is towards low frequency, while not consider the part of high frequency. The relation of decomposition is S=A3+D3+D2+D1. In addition to stress that here is one level decomposition, if it need further decomposition, the part of low frequency A3 can bread down into part of low frequency A4 and part of high frequency D4, and so on.

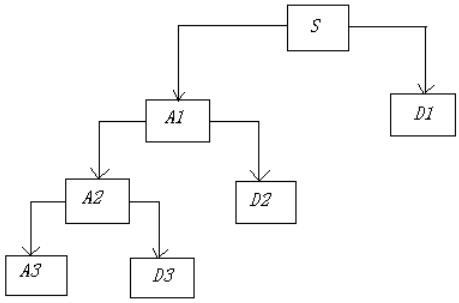


Figure 1 three multi-resolution analysis tree

In order to understand multi-resolution analysis, we must firmly grasp the point that the ultimate purpose of decomposition is seek to construct a orthogonal wavelet base which highly approach the space of $L^2(R)$ in frequency. These orthogonal wavelet bases of different frequency resolution are equal to band pass filter of different bandwidth. As can be seen from Figure 1, multi-resolution analysis on further decomposition of low frequency space, which made the frequency resolution, became more and more high.

III. PRINCIPLE OF WAVELET THRESHOLD DENOISING

A. The Process of One-dimensional Signal Denoising

(1)Wavelet decomposition of one-dimensional signal. Choose a wavelet and determine the level of a wavelet transformn *N*, then implement *N* layers wavelet decomposition of signal *S*.

(2)Threshold quantization of high-frequency coefficient of wavelet transform. Selecting a threshold implement soft threshold quantization for high-frequency coefficient of 1 to N layer.

(3)Reconstruction of one-dimensional wavelet. Implement wavelet reconstruction for low-frequency of NO.N and high-frequency of 1 to N layer which experienced the process of quantization.

In these three steps, the most critical is how to select the threshold and how to quantity the threshold. To some extent, it is directly related to the quality of the signal de-noising [10-12].

B. Principle of Threshold Selection in Wavelet Denoising

The following threshold can be chosen:

tptr='rigrsure' is an adaptive threshold selection based Stein's unbiased likelihood estimation principle. For a given threshold *t*, getting its likelihood estimation first, then minimize the nonlikelihood *t*, so the threshold has been obtained. It is a software threshold estimator.

tptr='sqtwolog' is using in the form of fixed threshold, the threshold generated is "sqrt(2*log(length((X)))"

tptr='heursure' is integration of the first two threshold, and also the best predict variable threshold selection. If the signal to noise ratio is small, SURE estimation will have lots of noise. If this occurs, we can use this fixed threshold.

tptr='minimaxi' is also using a fixed threshold, it produces an extremum with minimum mean square error, but not without error. Statistically, this maximum principle is used for designing estimator. Because the signal needed de-noising can be seen similar to the estimation of unknown regression function, this extreme value estimator can realize minimize of maximum mean square error for a given function[12].

IV. SIMULATION ANALYSIS

In the MATLAB environment, first of all, implement low-bass filtering for original collection BS signal. As far as the results of foreign literature, BS spectrum is distributed in 50~1500Hz frequency range. As the acquisition process of original signal can not prevent some other band noise interference, it is necessary to implement low-pass pre-filtering for the original signal; this can greatly enhance the effect of noise elimination, and at the same time reduce the burden on the wavelet filter. Then, use four threshold methods of the function *wden* and the default threshold de-noising of the function *wdencmp*. Select the waveletof *db*4 and the decomposition level of 5.

Denoising command under MATLAB is:

Four threshold denoising code:

```
xd1=wden(y,'rigrsure','s','one',5,'db4');

xd2=wden(y,'heursure','s','one',5,'db4');

xd3=wden(y,'sqtwolog','s','one',5,'db4');

xd4=wden(y,'minimaxi','s','one',5,'db4');
```

The default threshold denoising method code:

```
[thr,sorh,keepapp]=ddencmp('den','wv',y);

xd5=wdencmp('gbl',y,'db4',5,thr,sorh,keepapp);
```

Where *y* is a signal through low-pass filter.

xd1~xd5 are denoising results, the results are shown as Figure (d) ~Figure (h). And Figure (a) is the original signal, Figure (b) is its spectrum and Figure (c) is the low-pass filtering result.

From the above graph you can see, the spectrum of original signal distributed in 0~4000Hz, the main component distributed in 0~2000Hz, and there is a maximum at 228Hz, that verify the normal BS signal have a maximum peak between 160Hz and 260Hz[3].

Compare the above chart with Figure (a), this signal after low-pass filtered slightly different than the original signal, because the main spectrum of original signal distributed in 0~2000Hz, while the low-pass filter retain the signal spectrum range of 0~1500Hz, so get this result.

Compare Figure (d) ~Figure (g) with Figure (c), and Figure (d) ~Figure (g) are the results of the four threshold denoising. Compared to the signal through low-pass pre-filtered, the four kinds of threshold methods have removed some noise. Compare the Figure (d) and Figure (g) with Figure (e) and Figure (f), the naked eye can see that the first two denoising are better than the latter two, that's to say that the *rigrsure* and *minimaxi* threshold method can effectively remove noise while retain the characteristics of BS. The denoising results are very similar with the results of previous studies. While the *heursure* and *sqtwolog* threshold denoising method at the same time may remove some of the useful component.

Compare the above chart with Figure (c), we can see that the default threshold denoising method have removed some noise, but the denoising effect is not significant and can not recover bowel sounds better.

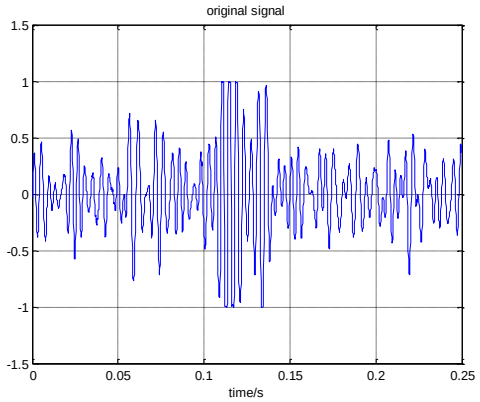


Figure (a)

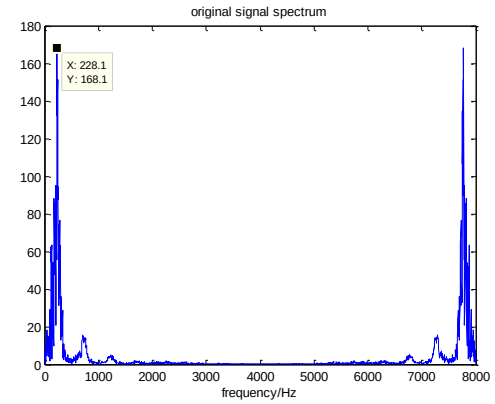


Figure (b)

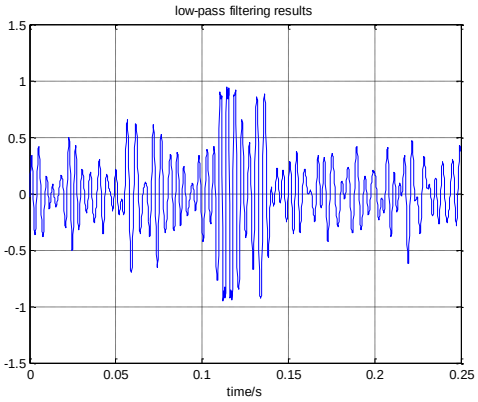


Figure (c)

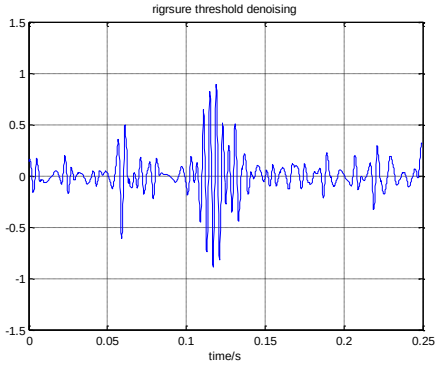


Figure (d)

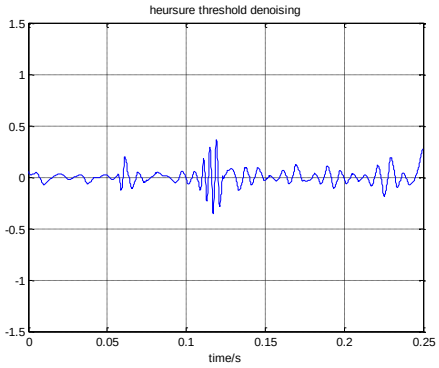


Figure (e)

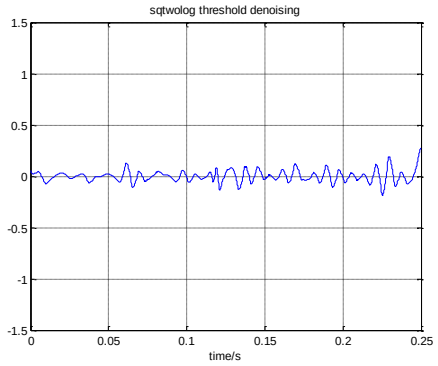


Figure (f)

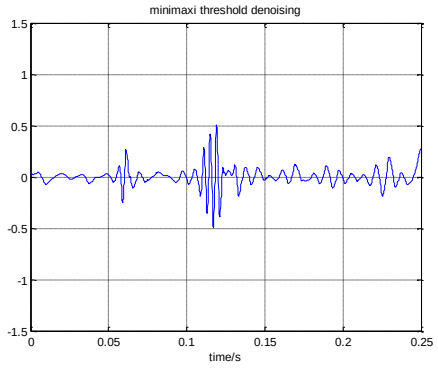


Figure (g)

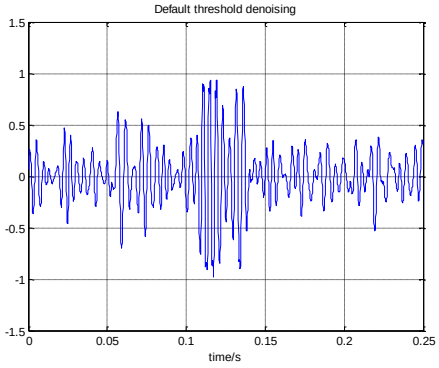


Figure (h)

In order to represent the denoising effect more accurately, we can calculate the *signal to noise ratio* and the *root mean square error* for the four types of threshold denoising and the default threshold denoising [4]. Use the original signal $x(n)$ and the filtered signal $x(n)'$, we can define the *signal to noise ratio* as [15]:

$$\mathrm{SNR} = 10\log\left[\frac{\sum_n x(n)^2}{\sum_n [x(n)-x(n)']^2}\right] \quad (10)$$

The *root mean square error* of original and denoised signal is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_n [x(n)-x(n)']^2} \quad (11)$$

The higher *signal to noise ratio*, the smaller the *root mean square error*, the closer the denoising signal with the original signal, the better the denoising effect [13-5].

Shown in Table 1, we can see SNR and RMSE which used *db*4 wavelet.

Table 1 the SNR and RMSE after denoising (*db*4)

	rigrsure threshold	*heursure* threshold	*sqtwolog* threshold	*minimaxi* threshold	default threshold
SNR(db)	3.5471	0.7926	0.2283	1.1793	16.3385
RMSE	0.2093	0.2875	0.3068	0.2749	0.0480

If use the *symlet*4 wavelet, the SNR and RMSE after denoising are shown in Table 2. But the denoising results are omitted.

Table 2 the SNR and RMSE after denoising (*sym*4)

	rigrsure threshold	*heursure* threshold	*sqtwolog* threshold	*minimaxi* threshold	default threshold
SNR(db)	1.9583	0.6806	0.1985	1.0726	16.4907
RMSE	0.2514	0.2912	0.3078	0.2783	0.0472

SNR and RMSE calculated above is to test the effect of wavelet filtering, so the original signal $x(n)$ is equal to signal y after low-pass filtered, while $x(n)$' is equal to wavelet denoising signal *xd*.

From Table 1 and Table 2 shows, four threshold denoising and default threshold denoising methods are improved SNR, and have removed some noise. From the denoising effect, the principle of *rigrsure* threshold and *minimaxi* threshold denoising are better than the other three methods, because the two denoised SNR higher than *heursure* and *sqtwolog* denoising methods and its RMSE are smaller. While the default threshold method have the highest SNR and smallest RMSE, however, we can see from the waveform after denoising that the denoising effect is not obvious, there is no effective removal of noise. While the *heursure* threshold and *sqtwolog* threshold principle will remove useful signal as noise. If there is high frequency information signal, the principle of the two threshold denoising results will not be satisfactory. Compare Table 1 with Table 2, we can be seen that using the db wavelet and the sym wavelet, the denoising effect of the former is better than the latter.

V. Conclusion

From the simulation analysis, the wavelet transform in signal denoising in particular nonstationary signals, can effectively remove noise and improve SNR. With regard to the bowel sounds which complex and highly nonstationary, wavelet transform denoising can play the advantages compared to traditional denoising, wavelet can better demonstrate its advantages. From the simulation results, we can obtain that use the principle of *rigrsure* threshold and *minimaxi* threshold can effectively reduce noise, and can retain a useful component of bowel sounds. Use *heursure* and *sqtwolog* threshold to denoising, it will remove a useful component of bowel sounds at the same time, this method more harm than good. Use the default threshold method to denoising can not effectively remove the noise, unable to obtain the signal characteristics of bowel sounds.

Acknowledgements

The authors are grateful to the anonymous reviewers and to the Natural Science Foundation of Hunan Province for supporting this work through research grant JJ076111, and the Student Innovation Programme of Hunan Province through research grant 513.

References

[1] G. Jianbo, H. Sultan, H. Jing, T. Wenwen, "Denoising Nonlinear Time Series by Adaptive Filtering and Wavelet Shrinkage: A Comparison", IEEE Signal Processing Letters, 2010, vol.17, no.3, pp.237-240.

[2] G. Zhexue, C.Zhongsheng, "Matlab Time-frequency Analysis and Its Application (Second Edition) ," Beijing: Posts & Telecom Press. 2009.

[3] M.Kania, M.Fereniec, and R. Maniewski, "Wavelet denoising for multi-lead high resolution ECG signals", Measurement Science review, 2007, vol.7., no.1, pp. 30-33.

[4] Zhang Hehua. Bowel sounds acquisition system design and the signal analysis methods research. Master's degree thesis of Third Military Medical University. 2009.5.1.

[5] Wu Wei,Cai Peisheng,"Wavelet denoising simulation based on MATLAB. Signal and Electronic Engineering,2008,vol. 6, no.3, pp.220-222.

[6] Qin Yahui,Feng Jinghui,Chen Liding, "Study of signal denoising methods based on wavelet transform. Information Technology, 2010, vol.5, no.2, pp.53-55.

[7] C.Dimoulas ,G.Kalliris ,G.Papanikolaou ,V.Petridis ,A.Kalampakas,"Bowel-sound pattern analysis using wavelets and neural networks with application to long-term,unsupervised, gastrointestinal motility monitoring", Expert Systems with Applications, 2008, vol.38, no.1, pp.26-41.

[8] A. Borsdorf, R. Raupach, T. Flohr, and J. Hornegger, "Wavelet Based Noise Reduction in CT-Images Using Correlation Analysis," IEEE Transactions on Medical Imaging, 2008, vol.27, no.12, pp.1685–1703.

[9] R. Ranta, V. Louis-Dorr, Ch. Heinrich, D. Wolf, "Guillemin Digestive activity evaluation by multi-channel abdominal sounds analysis", IEEE Transactions on Biomedical Engineering, in press.

[10] Z. Hehua, W. Baoming, Z. Lianyang, et.al. ,"Study on Adaptive Filter and Methods of Bowel Sounds Features Extraction", Journlal of Medical Equipments, 2009, vol.26, no.3, pp.58-62.

[11] C.Fu, C, Lifei, "Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry", Expert Systems with Applications, 2008, vol.34, no.1, pp. 280-290.

[12] R. Ranta, R. Louis-Dorr, V. Heinrich, et.al., "Digestive Activity Evaluation by Multichannel Abdominal Sounds Analysis", IEEE Transactions on medical Engineering, 2010, vol.57, no.6, pp.1507-1519.

[13] D.Matsiki, "Wavelet-based analysis of nocturnal snoring in apneic patients undergoing polysomnography", Conf Proc IEEE Eng Med Biol Soc, Thessaloniki, Greece, 2007, pp.1912-1915.

[14] I. K. Kitsas, "Classification of transmembrane segments in human proteins using wavelet-based energy", Proc IEEE Eng Med Biol Soc Conf, hessaloniki, Greece, 2007, pp.1225-1228.

[15] W. Wei, Z.Z. Ge, H. Lu, et al. , "Purgative bowel cleansing combined with simethecone improves capsule endoscopy imaging", American Journal of Gastroenterol, 2008, vol.103, no.1, pp.77-82.

Xizheng Zhang received the B.S. degree in control engineering and the M.S. degree in circuits and systems in 2000 and 2003, respectively, from Hunan University, Changsha, China, where he is currently working toward the Ph.D. degree with the College of Electrical and Information Engineering.

He is also with the School of Computer and Communication, Hunan Institute of Engineering, Xiangtan, China. His research interests include industrial process control and artificial neural networks.

2- Input AND Gate For Fast Gating and Switching Based On XGM Properties of InGaAsP Semiconductor Optical Amplifier

Vikrant k Srivastava
Electronics & Instrumentation Department
Indian School of Mines
Dhanbad 826 004 ,INDIA.
srivastavavikrant@hotmail.com

Devendra Chack
Electronics & Communication Deptt
BCT- Kumaon engineering College
Dwarahat Almora UK
263653 ,*INDIA.*
devendra.chack@gmail.com

Vishnu Priye
Electronics & Instrumentation Department
Indian School of Mines
Dhanbad 826 004 ,INDIA.
vpriye@gmail.com

***Abstract*— We report on an all-optical AND-gate using simultaneous Four-Wave Mixing (FWM) and Cross-Gain Modulation (XGM) in a semiconductor optical amplifier (SOA). The operation of the proposed AND gate is simulated and the results demonstrate its effectiveness. This AND gate could provide a new possibility for all-optical computing and all-optical routing in future all-optical networks. In an AND (AB) gate, Boolean is firstly obtained by using signal B as a pump beam and clock signal as a probe beam in SOA-1. By passing signal A as a probe beam and as a pump beam through SOA-2, Boolean AB is acquired. Proposed optical logic unit is based on coupled nonlinear equations describing XGM and FWM effects. These equations are first solved to generate the pump, probe and conjugate pulses in a SOA. The pulse behavior are analyzed and applied to realize behavior of all-optical AND gate and its function is verified with the help of waveform and analytical assumption.**

Keywords: Optical Logic Gates, Semiconductor Optical Amplifier, SOA, Four Wave Mixing, Cross Gain Modulation

I. INTRODUCTION

A lot of schemes based on cross-gain modulation (XGM) have been reported, such as AND gates [1], NAND gate [2], NOR gates [3, 4], XOR gate [5] etc. Cross-gain modulation (XGM), one of several wavelength technique methods based on SOAs, is simple to implement and has shown impressive operation for high bit rates [6–8]. Electrons in the SOA are placed in the excited states when electrical current is applied to an SOA. An incoming optical signal stimulates the excited electrons , and settled to the ground states after the signal is amplified. This stimulated emission continues as the input signal travels through the SOA until the photons exit together as an amplified signal. However, the amplification of the input signal consumes carriers thereby transiently reducing a gain, which is called gain saturation. The carrier density changes in an SOA will affect all of the input signals, so it is possible that a signal at one wavelength affect the gain of signal at another wavelength. This nonlinearity property is called XGM based on an SOA. When a pulse exists for the pump signal passing through an SOA, it causes carrier depletion in the SOA. The carrier depletion leads to gain saturation in the SOA causing the marked intensity reduction of an incoming probe signal. Therefore, the marked intensity reduction of the probe signal in the SOA leads to no pulse existence for output signal. When a pulse does not exist for the pump signal, there is no effect on the gain of probe signal in the SOA. Therefore, output signal has the same pulse as the probe signal.

In this paper the same effect is analyzed properly and then applied to get the desired result.

II. SIMULATION METHOD

In our approach the reference equations are taken from Ref. [9] and different parameters which are taken into consideration are tabulated below in Table I. It is assumed that input pump, and probe pulses have the same temporal width as well as perfect pulse overlap, and in all of the cases, their powers are set to a ratio of 10:1. Numerical simulations have been undertaken to investigate the amplification of strong picosecond optical pulses in semiconductor optical amplifiers (SOAs), taking into account carrier heating, spectral holeburning, carrier-carrier scattering (CCS) and carrier photon scattering (CPS). The result of interference of two copolarized pulses when propagating into SOA,one pump pulse at central frequency ω1 and the other probe pulse at central frequency ω0 ,induce a bit of carrier density pulsation at the frequency detuning Ω=ω1-ω0. This results a generation of a new frequency pulse at ω2= ω0- Ω=2ω0-ω1.The new pulse is phase conjugate replica of the probe pulses, and can be extracted from the input pulses using an optical filter. Here , A_j (Z, t) , j=0,1,2, correspond to the slowly varying envelopes of the pump, the probe, and the conjugate pulses, respectively, and Ω=ω1-ω0, is the frequency detuning.

The following Equation has been taken into consideration [9].

$$A_0(L, t) = A_0(0, t)\, e^{(\frac{1}{2})(1-i\alpha)h} \quad (1)$$

Where

A_0 (0, t), is the input pump pulse amplitude at any end of SOA,
A_0 (L, t), is input pump pulse amplitude at any length L of SOA,
L is length of SOA. t is time . Rest parameters are already defined in Table I.

$$A_1(L,t) = A_1(0,t)e^{\frac{1}{2}[(1-i\alpha)h-\eta_{10}|A_0(0,t)|^2(e^h-1)]} \times \cosh[(\frac{1}{2})\sqrt{\eta_{02}\eta_{01}^*}|A_0(0,t)|^2(e^h-1)] \quad (2)$$

Where

A_1 (0, t), is input probe pulse amplitude at any end of SOA.

A_1 (L, t), is input probe pulse amplitude at any length L of SOA.
L= length of SOA. t is time. Rest parameters are already defined in Table I.

$$A_2(L,t) = \frac{-A_1^*(L,t)A_0(L,t)}{A_0^*(L,t)}\sqrt{\frac{\eta_{01}}{\eta_{02}^*}} \times \sinh[(\frac{1}{2})\sqrt{\eta_{01}\eta_{02}^*}|A_0(L,t)|^2 e^{-h}(e^h-1)] \quad (3)$$

Where

A_2 (0, t) is input conjugate pulse amplitude at any end of SOA.

A_2 (L, t) is input conjugate pulse amplitude at any length L of SOA.
L= length of SOA, t is time. Rest parameters are already defined in Table I.

$$\eta_{01}=\eta_{01}^{CD}+\eta_{01}^{CH}+\eta_{01}^{SHB} \quad (4)$$

Where,

$$\eta_{01}^{CD}=\varepsilon_{cd}\frac{1-i\alpha}{((1+i\Omega\tau_1)+(1+i\Omega\tau_s))}$$

$$\eta_{01}^{CH}=\varepsilon_{t}\frac{1-i\alpha_T}{((1+i\Omega\tau_h)+(1+i\Omega\tau_1))}$$

$$\eta_{01}^{SHB}=\varepsilon_{shb}\frac{1-i\alpha_{shb}}{1+i\Omega\tau_1}$$

The amplification function h and coupling coefficient η_{ij} are defined in [9]. The effects of carrier depletion, carrier heating, spectral hole burning, two-photon absorption, and ultrafast nonlinear refraction are taken into account, leading to a successful description of wave mixing for optical pulses with strong pulse energy, and/or with pulsewidths larger than several hundred femtoseconds , as well as in active or passive optical waveguides. The parameters α , τ , and ε determine the strength and nature of the wave mixing process caused by each of the intraband mechanisms and their relative significance. The values of α , τ , and ε cannot be determined unanimously.

TABLE I: PARAMETERS USED IN SIMULATION WORK

Parameters	***Symbols***	***Values***	***Unit***
Length of the amplifier	L	450	μ *m*
Small signal gain	G	$1.54*10^{-4}$	m^{-1}
Carrier lifetime	τ_s	300	*ps*
Nonlinear gain compression for carrier heating	ε_t	0.13	w^{-1}
Non linear gain compression for spectral hole burning	ε_{shb}	0.07	w^{-1}
Traditional linewidth enhancement factor	α	5.0	
Temperature linewidth enhancement factor	α_T	3.0	
Linewidth enhancement factor for spectral hole burning	α_{shb}	0.1	
Time for carrier-carrier scattering	τ_1	50	fs
Time for carrier photon scattering	τ_h	700	fs
Carrier depletion coefficient	ε_{cd}	47	w^{-1}

III. RESULTS AND DISCUSSION

Truth table for AND Logic can be given as

A	B	AB
0	0	0
0	1	0
1	0	0
1	1	1

Column AB of the above truth table indicates logic behaviour of AND gate. The full design needs two SOAs.

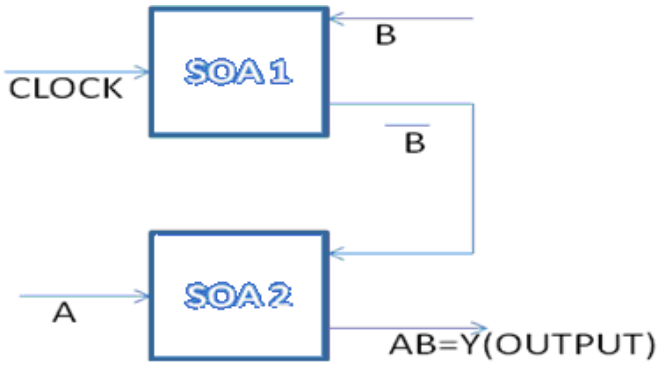


Figure 1.

Fig. 1 shows that in an AND (AB) gate, Boolean is firstly obtained by using signal B as a pump beam and clock signal as a probe beam in SOA-1. Next, by passing signal A as a probe beam and as a pump beam through SOA-2, Boolean AB is acquired.

The following waveforms shows different outputs of simulation of above equations.

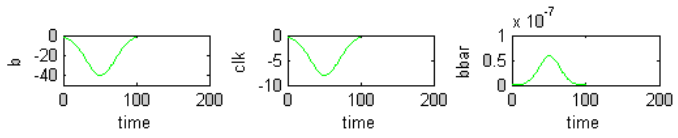


Figure 2 .(Pump b=0, Probe clk=0)

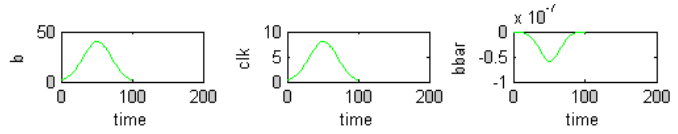


Figure 3. (Pump b=1, Probe clk=1)

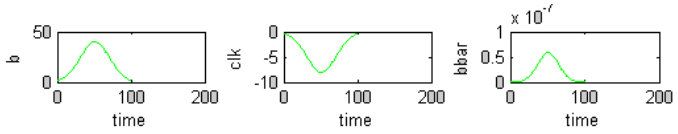


Figure 4. (Pump b=1, Probe clk=0)

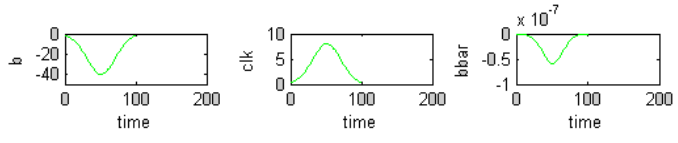


Figure 5. (Pump b=0, Probe clk=1)

Analysing Fig. 2, Fig. 3, Fig. 4,and Fig.5 ,we observe an interesting result.The result shows that on proper manipulation of pump ‘b’,and probe ‘clk’.we can get desired outputs. In Fig. 2 and Fig. 3 the input pump get inverted whereas in Fig. 4 and Fig. 5, the outputs are same as that of inputs. The generated signal when filtered out and when applied as a pump with a different probe named ‘A’, in SOA 2 ,results the desired logic of an AND Gate.

IV. CONCLUSIONS

In this paper, it is shown that using analytical solutions of nonlinear effects in semiconductor optical amplifier, we can model an AND Circuit. This research has guided readers to design all-optical AND logic circuits so that anyone can construct any types of all-optical different logic circuits by utilizing the detailed process of designing the AND Gate..

REFERENCES

[1] X. Zhang, Y. Wang, J. Sun, D. Liu, and D. Huang, "All-optical AND gate at 10 Gbit/s based on cascaded single-port-couple SOAs," Opt. Express **12**, 361-366 (2004).

[2] S.H. Kim, J.H. Kim, B.G. Yu, Y.T. Byun, Y.M. Jeon, S. Lee, D.H. Woo, and S.H. Kim, "All-optical NAND gate using cross-gain modulation in semiconductor optical amplifiers," Electronics Letters **41**, 1027-1028(2005).

[3] A. Hamie, A. Sharaiha, M. Guegan, and B. Pucel, "All-optical logic NOR gate using two-cascaded semiconductor optical amplifiers," IEEE Photon. Technol. Lett. **14**, 1439-1441(2002) .

[4] C. Zhao, X. Zhang, H. Liu, D. Liu, and D. Huang, "Tunable all-optical NOR gate at 10 Gb/s based on SOA fiber ring laser," Opt. Express **13**, 2793-2798 (2005).

[5] J.H. Kim, Y.M. Jhon, Y.T. Byun, S. Lee, D.H. Woo, and S. H. Kim, "All-optical XOR gate using semiconductor optical amplifiers without additional input beam," IEEE Photon. Technol. Lett. **14**, 1436 - 1438(2002).

[6] Kim, J.H., et al.: ‘All-optical XOR gate using semiconductor optical amplifiers without additional input beam’, IEEE Photonics Technol. Lett.,2002, 14, pp. 1436–1438

[7] Kim, J.H., et al.: ‘All-optical AND gate using cross-gain modulation in semiconductor optical amplifiers’, Jpn. J. Appl. Phys., 2004, 43, pp. 608–610

[8] Ellies, A.D., et al.: ‘Error free 100 Gbit=s wavelength conversion using grating assisted cross-gain modulation in 2 mm long semiconductor optical amplifier’, Electron. Lett., 1998, 34, pp. 1958–1959

[9] S Junqlang et al., “Analytical solution of Four wave mixing between picosecond opticalpulses in semiconductor optical amplifiers with cross gain modulation and probedepletion”, Microwave and Optical technology Letters, Vol. 28, No. 1, pp 78 – 82, 2001.

[10] Chan, L.Y., Qureshi, K.K., Wai, P.K.A., Moses, B., Lui, L.F.K., Tam, H.Y., Demokan, M.S.: ‘All-optical bit-error monitoring system using cascaded inverted wavelength converter and optical NOR gate’, IEEE photonic technology letters, 2003, 15, (4), pp.593-595.

[11] Martinez, J.M., Ramos, F., Marti, J.: ‘All-optical packet header processor based on cascaded SOA-MZIs’, Electronics letters, 2004, 40,(14), pp.894-895

[12] G. Berrettini, A. Simi, A. Malacarne, A. Bogoni, and L. Potí, *Member, IEEE:'* Ultrafast Integrable and Reconfigurable XNOR, AND,NOR, and NOT Photonic Logic Gate’, IEEE Photonic technology letters, Vol. 18, No. 8, april 15, 2006

Analysis and Enhancement of BWR Mechanism in MAC 802.16 for WiMAX Networks

R. Bhakthavathsalam
SERC, Indian Inst of Science
Bangalore, India.
bhaktha@serc.iisc.ernet.in

R. ShashiKumar, V. Kiran, Y. R. Manjunath
SJCIT, Chikballapur
Karnataka, India
rshashiku@gmail.com

***Abstract*— WiMAX [Worldwide Interoperability for Microwave Access] is the latest contender as a last mile solution for providing broadband wireless Internet access and is an IEEE 802.16 standard. In IEEE 802.16 MAC protocol, Bandwidth Request (BWR) is the mechanism by which the Subscriber Station (SS) communicates its need for uplink bandwidth allocation to the Base Station (BS). The performance of the system is affected by the collisions of BWR packets in uplink transmission, that have the direct impact on the size of the contention period in uplink sub frame, uplink access delay and uplink throughput. This paper mainly deals with Performance Analysis and Improvement of Uplink Throughput in MAC 802.16 by the Application of a New Mechanism of Circularity. The implementation incorporates a generic simulation of the contention resolution mechanism at the BS. An analysis of the total uplink access delay and also uplink throughput is performed. A new paradigm of circularity is employed by selectively dropping appropriate control packets in order to obviate the bandwidth request collisions, which yields the minimum access delay and thereby effective utilization of available bandwidth towards the uplink throughput. This new paradigm improves contention resolution among the bandwidth request packets and thereby reduces the delay and increases the throughput regardless of the density and topological spread of subscriber stations handled by the BS in the network.**

Keywords- WiMAX, MAC 802.16 , BWR, NS-2, Tcl

I. INTRODUCTION

The IEEE 802.16 MAC layer plays an important role in the OSI model. It is a sub layer of the data link layer, the other being the logical link control layer (LLC), and acts as a link between the lower hardware oriented and the upper software oriented layers.

IEEE 802.16 consists of the access points like BSs (Base Station) and SSs (Subscriber Stations). All data traffic goes through the BS, and the BS can control the allocation of bandwidth on the radio channel. According to demand of subscribers stations Base Station allocates bandwidth. Hence 802.16 are a Bandwidth on Demand system. Basically there are two modes of operation in 802.16 i.e. P2MP (Point to Multi Point) and Mesh Mode of operation. In P2MP mode of operation a base station can communicate with subscriber station and/or base stations. Each SS is identified by a 48-bit IEEE MAC address and a BS is identified by a 48-bit Base Station ID (Not MAC address). Each connection with BS and SS in a session is identified by a 16-bit CID (Connection Identifier) In P2MP mode of operation communication between BS and SS is established based on Req / Grant mechanism (which is CSMA/CA in case of 802.11). In Mesh Mode operation a subscriber station can directly communicate with another subscriber's station within its communicating range. It is ad-hoc in nature [**1**].

We have considered the point-to-multipoint mode of operation of the IEEE 802.16 network. In this mode, the communication occurs only between the SS and the BS. The BS directly controls the data that is transmitted between the different SSs. In addition to the control of communication between SSs, a BS can send information to other BSs as well. This allows SSs that are not connected to the same BS, to exchange data. Initially when a SS is switched on, it performs the network entry procedure [**2**].

The procedure can be divided into the following phases:

- Scan for downlink channel and establish synchronization with the BS
- Obtain transmit parameters (from UCD message)
- Perform ranging
- Negotiate basic capabilities
- Authorize SS and perform key exchange
- Perform registration
- Establish IP connectivity
- Establish time of day
- Transfer operational parameters
- Set up connections

The following figure 1 shows the network entry procedure

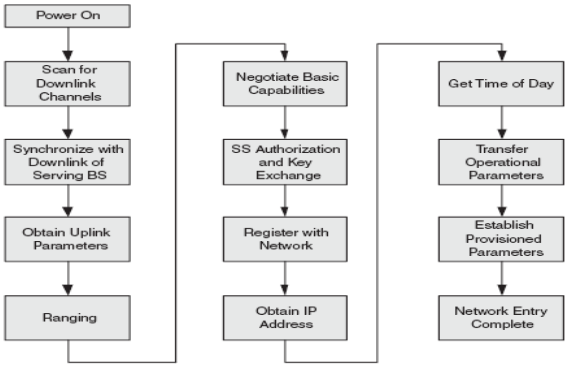


Figure 1

The BS in turn returns certain response messages in order to complete the different steps. The first such mechanism is Initial Ranging. Here, a series of Ranging-Request and Ranging-Response messages are exchanged. Also the processes of Negotiating Basic Capabilities and Registration involve their own request messages. These are called the SBC-REQ and REG-REQ messages respectively. After a SS has gained entry into the network of a BS, if it has data to send, it starts the Bandwidth Request (BWR) procedure. In our project, we focus on the contention-based BWR mechanism and its performance metrics [**3**].

A. Salient features of WiMAX

WiMax basically offers two forms of wireless service [**4**]:

1. **Non-line-of-sight**: This service is a WiFi sort of service. Here a small antenna on your computer connects to the WiMax tower. In this mode, WiMax uses a lower frequency range -- 2 GHz to 11 GHz (similar to WiFi).
2. **Line-of-sight**: In this service, where a fixed dish antenna points straight at the WiMax tower from a rooftop or pole. The line-of-sight connection is stronger and more stable, so it's able to send a lot of data with fewer errors. Line-of-sight transmissions use higher frequencies, with ranges reaching a possible 66 GHz.

The entire WiMax scenario is as shown in Figure 2.

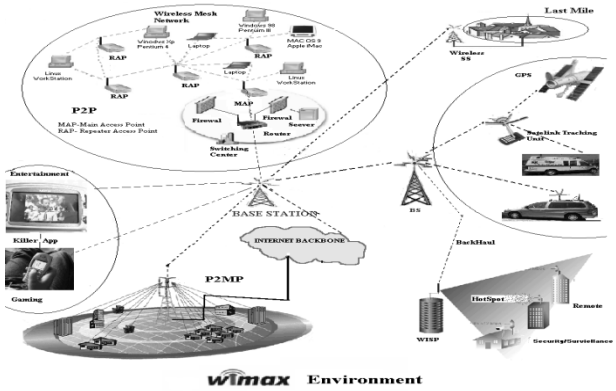


Figure 2

WiMax is a wireless broadband solution that offers a rich set of features with a lot of flexibility in terms of deployment options and potential service offerings. Some of the more salient features that deserve highlighting are as follows:

- OFDM-based physical layer
- Very high peak data rates
- Scalable bandwidth and data rate support
- Scalable bandwidth and data rate support
- Link-layer retransmissions
- Support for TDD and FDD
- WiMax uses OFDM

Flexible and dynamic per user resource allocation

II. STATEMENT OF THE PROBLEM

Bandwidth Request (BWR) is the mechanism by which the SS communicates its need for uplink bandwidth allocation to the BS. A single cell in WiMax consists of a base station (BS) and multiple subscriber stations (SSs). The BS schedules the traffic flow in the WiMax i.e., SSs do not communicate directly. The communication between BS and SS are bidirectional i.e., a downlink channel (from BS to SS) and an uplink channel (from SS to BS). The downlink channel is in broadcast mode. The uplink channel is shared by various SS's through time division multiple access (TDMA). The subframe consists of a number of time slots. The duration of subframe, slots and the number are determined by the BS scheduler. The downlink subframe contains uplink map (UL map) and downlink map (DL map). The DL map contains information about the duration of sub frames and which time slot belongs to a particular SS as the downlink channel. The UL map consists of information element (IE) which includes transmission opportunities [**5**]. Bandwidth Request refers to the mechanism by which the SS ask for bandwidth on the uplink channel for data transmission. Compared to RNG-REQ, these request packets are of multiple types

They can be stand-alone or piggybacked BWR message. The requests must be made in terms of the number of bytes of data needed to be sent and not in terms of the channel capacity. This is because the uplink burst profile can change dynamically. There are two methods for sending bandwidth request. Contention Free method and Contention Based method. In contention free bandwidth request method, the SS will receive its bandwidth automatically from BS. In contention based bandwidth request method, more than one subscriber can send their request frames to same base station at same time. Hence there will be chances of collision, which is resolved by Truncated Binary Exponential Backoff Algorithm. BE and nrtPS services are two cases of contention based request method. In the uplink subframe of the TDD frame structure of 802.16, there are a number of Contention Slots (CSs) present for the purposes of BWR. In contention based bandwidth request mechanism, an SS attempts to send its BWR packet in one of these contention slots in the uplink subframe to the BS. If a BWR packet is successfully received by the BS and bandwidth is available, then the bandwidth is reserved for the SS and it can transmit its data without contention in the next frame. In case multiple subscriber stations send their request messages to the same BS at the same time, there will be a collision. So, here the contention is resolved by using the truncated binary exponential backoff procedure. The minimum and maximum backoff windows are decided by the BS. Also, the number of bandwidth request retries is 16. The backoff windows are always expressed in terms of powers of two. This algorithm solves retransmission strategy after a collision. It keeps track of the number of collisions [**6**].

A Subscriber Station can transmit its bandwidth request

using bandwidth request contention slots known as BWR contention slots. Subscriber stations can access the channel using RACH method i.e. Random Access Channel method. A subscriber station that wants to transmit data first enters into the contention resolution algorithm. In this scheme we consider different number of contention slots for each frame and each subscriber station has 16 transmission opportunities at maximum. The Initial Backoff window size is 0-7 and the final maximum backoff window size is 0-63. While accessing the channel randomly this way, there will be collisions among these request packets.

Suppose the backoff window at a certain time for an SS is 0 to 15 (0 to 2^4-1) and the random number picked is 8. The SS has to defer a total of 8 BWR Contention Slots before transmitting the BWR packet. This may require the SS to defer BWR Contention Slots over multiple frames. In case a collision is detected, the backoff window is doubled. Now a random number is picked between 0 and 31 and the deferring is continued. This procedure can be repeated for a maximum of 16 times after which the data that was to be sent is discarded and the whole process is restarted from the very beginning. Base station can allocate bandwidth to subscriber station in three ways: GPC (Grant per Connection Mode), GPSS (Grant per Subscribers Station), and Polling. GPC is a bandwidth allocation method in which bandwidth is allocated to a specific connection within a subscriber station. GPSS is a bandwidth allocation method in which requests for the bandwidth is aggregated for all connections within a subscriber station and is allocated to the subscriber station as that aggregate. The bandwidth requests are always made for a connection. Polling is the process by which the BS allocates to the SSs bandwidth specifically for the purpose of making bandwidth requests. These allocations may be to individual SSs or to groups of SSs. Polling is done on either an SS or connection basis. Bandwidth is always requested on a CID basis and bandwidth is allocated on either a connection (GPC mode) or SS (GPSS mode) basis, based on the SS capability [**7**].

III. DESIGN SPECIFICATIONS

A. Analysis mode module

The purpose of this module is to analyze the communication between BS and SS's, and to obtain the access delay and Throughput of the network.

The following Network Architecture parameters are used for the simulation process.

- Number of Base Stations -1
- Number of Sink Nodes -1
- Number of Subscriber Stations – 6,12,…,60
- Traffic Start Time – 20 Sec
- Traffic Stop Time – 40 Sec
- Simulation Start Time – 0 Sec
- Simulation Stop Time – 50 Sec
- Channel Type – Wireless Channel
- MAC Type – Mac/802.16/BS

Along with the above network parameters, we also need to give the seed value for the simulation purpose. Seed value is actually taken by the Random Number Generation (RNG) process. There are 64 well-defined seed values available. We can find these seed values in network simulator application. By using these values we will carry out the simulation.

The parameters mentioned above are used in Tool Command Language script. Tcl scripting is used to design and simulate the WiMAX networks with varying architectures. Tcl gives us a lot of options that allow us to have a great degree of control over the simulation of networks. We use WiMAX Control Agent in order to produce a detailed account of the activities going on during the simulations.

Subscriber Stations (SSs) use the contention minislots for transmitting their requests for bandwidth. The Base Station allocates bandwidth for the SSs in the next frame, if bandwidth is available. Because of the reservation procedure, the SSs are guaranteed collision free data transmission. The contention is resolved in the uplink channel for bandwidth request based on a truncated binary exponential backoff, with the initial backoff window and the maximum backoff window controlled by the base station (BS). The values are specified as part of the Uplink Channel Descriptor (UCD) message, describe the physical layer characteristics of an uplink and are equal to power-of-two. The SS will enter a contention resolution process, if it has data for transmission.

Figure 3 shows the contention resolution process for this module.

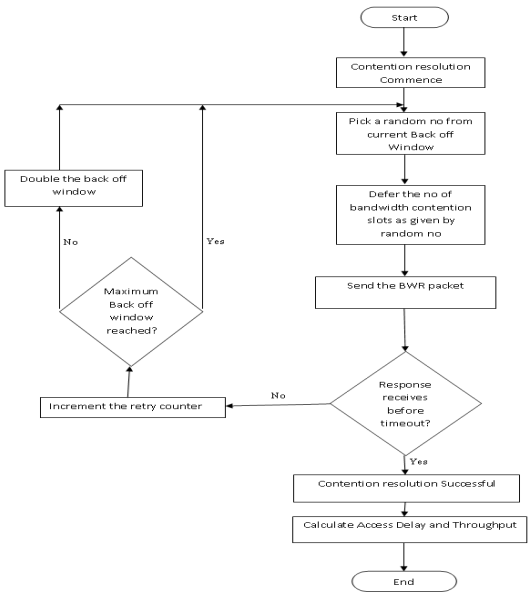


Figure 3

B. Enhancement mode module

The purpose of this module is to analyze the communication between BS and SS's, and obtain the delay incurred in BWR and Throughput of the network by using Circularity Principle. For this Module, the Network Architecture parameters are the

same as we set up in Analysis mode. We use these parameters in Tool Command Language. In addition to this, we have to input, Circularity value which is needed to carry out the enhancement process. Here we need to set this value at the back-end of the network simulator.

Figure 4 shows the contention resolution process for this module. The procedure is same as we see in the previous module but only difference is, here we are enclosed the new Circularity principle. Circularity concentrates on the modification of existing BWR scheme so as to minimize BWR collisions and hence improving the throughput [**8**].

Circularity is defined as a number which enables the identification of specific groups of events or packets. Each event or packet under consideration is numbered in a sequential manner. An event or packet with a number which is a multiple of the circularity value is said to be circularity-satisfied. So we are setting a counter to find Circularity satisfied packet and we are dropping it. So, a finite delay is introduced before the occurrence of circularity satisfied events or the sending of circularity satisfied packets. This additional delay reduces the probability of packet collisions. The selfless behavior of certain SSs may increase the individual BWR delays but on the whole the delay incurred in the entire network will be reduced. All the other calculation remains the same as we see in Analysis module.

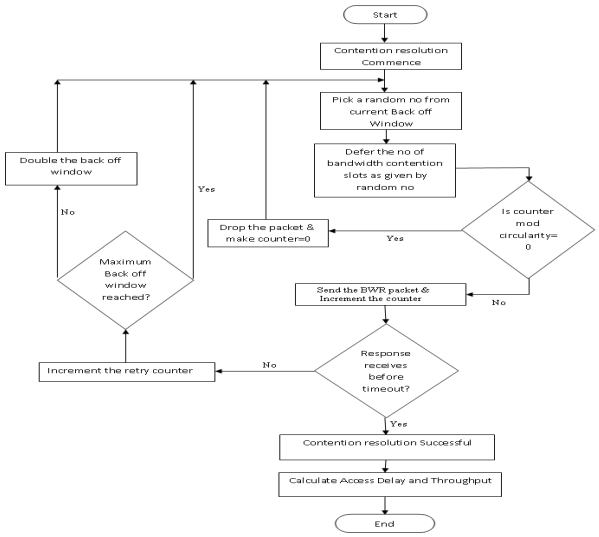


Figure 4

C. Comparison mode module

The purpose is to analyze the graph which shows that how the Enhancement mode is better than the existing scheme. For this Module, the Network Architecture parameters are the same as we set up in Analysis mode. We use these parameters in Tool Command Language. In addition to this, we have to input, Circularity value which is needed to carry out the enhancement process. Here we need to set this value at the back-end of the network simulator.

In this module, we are comparing the results of two previous modules. Before the processing of this module, we need to run the simulation by setting different number of mobile nodes for analysis and enhancement modules and note down the results. The obtained delay and throughput results are written in a file. The extension of this file should be filename.xg and need to input to the Xgraph package to plot the graph. The output file will be an image file of the extension png. Figure 5 shows the flow chart for comparing the Access delay results of two previous modules.

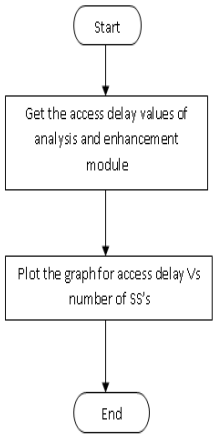


Figure 5

Figure 6 shows the flow chart for comparing the Throughput results of two previous modules.

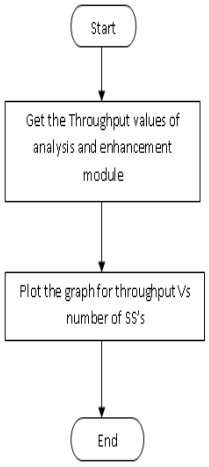


Figure 6.

IV. HARDWARE AND SOFTWARE REQUIREMENTS

A. Hardware Interface

- Pentium 4 processor.
- 512 MB RAM.
- 20 GB Hard Disc Memory

B. Software Interface

- Network Simulator 2 (ns-2)
- TCL (Tool Command Language)

1) The Network Simulator 2 (NS2): NS2 is a tool that helps you better understand certain mechanisms in protocol definitions, such as congestion control, that are difficult to see in live testing [**9**]. It is recommended NS2 as a tool to help understand how protocols work and interact with different network topologies. We can also patch the source code in case the protocol you want to simulate is not supported and live with low-quality graphics tool [**10**]. Figure 7 shows the network animator.

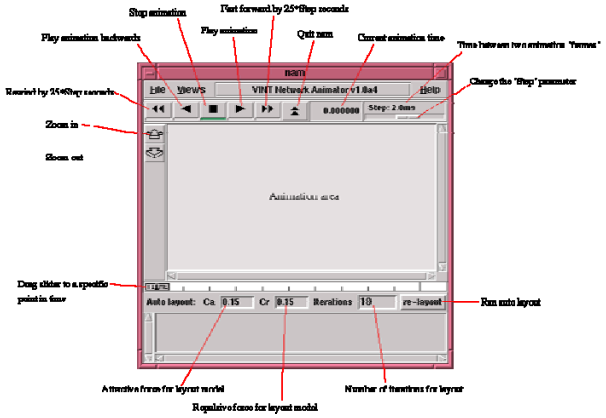


Figure 7

2) TCL (Tool Command Language)

Tcl is a small language designed to be embedded in other applications (C programs for example) as a configuration and extension language [**11**]. Tcl is specially designed to make it extremely easy to extend the language by the addition of new primitives in C. Like Lisp, Tcl uses the same representation for data and for programs.

V. SIMULATION SETUP

The Simulation consists of two parts:

- Simulation of BWR scheme without Circularity using NS-2
- Changing the backend and re-simulating (With circularity)

The following parameters are used for the simulation of the existing Bandwidth Request scheme.

- Channel Type – Wireless Channel
- Radio Propagation Model– TwoRayGround
- Network Interface Type Phy/WirelessPhy/OFDM
- MAC Type – 802_16
- Interface Queue Type – DropTail Priority Queue
- Link Layer Type – LL
- Antenna Model – Omni Antenna
- Maximum Packets in Interface Queue – 50
- Routing Protocol – DSDV (Routing is done through the Base Station)
- Routing Protocol- AODV (Routing is done through the Base Station)
- Network Architecture Parameters:
- Number of Base Stations – 1
- Number of Sink Nodes – 1
- Number of Subscriber Stations – Varied from 4 to 64
- Base Station Coverage – 20 meters
- Traffic Start Time – 20
- Traffic Stop Time – 40
- Simulation Stop Time – 50

A. Simulation of BWR Scheme without Circularity

The parameters mentioned are used in the Tool Command Language (TCL) script that we have written. This script also uses the WiMAX Control Agent in order to produce a detailed account of the activities going on during the simulations. In the resulting output file, we search for the timing details of specific events in order to extract the Bandwidth Request delay.

The start and stop traffic times of the BWR procedure for all the Subscriber Stations (SS) in the scenario are stored in files. Using a C program, we find the average BWR delay per node, after calculating the total time taken by all the nodes to complete their respective BWR processes.

Such simulations can be carried out for different numbers of SS each time by the use of shell scripts. Then the average BWR delay is recorded along with number of SS involved in each such simulation.

B. Changing the backend and re-simulating (With circularity)

In order to enhance the Bandwidth Request scheme, we make some modifications to the backend of ns-2, which is implemented in C++ language. During the BWR procedure, there will be many SS contending to send their requests to join the network. The packets sent by different SS may collide at some instants and they will have to be resent. We try to reduce the collisions between packets of different SS by making the SSs less selfish. We have made two such changes, one in each of the files mentioned above [**12**].

VI. RESULTS

We have run the simulation for 6 mobile nodes and we can see the output as in the figure 8. The calculator function takes the traced file, track the sent times details and stored in the intermediate file by name sent times-f. The figure 9 shows the screenshot of sent time details.

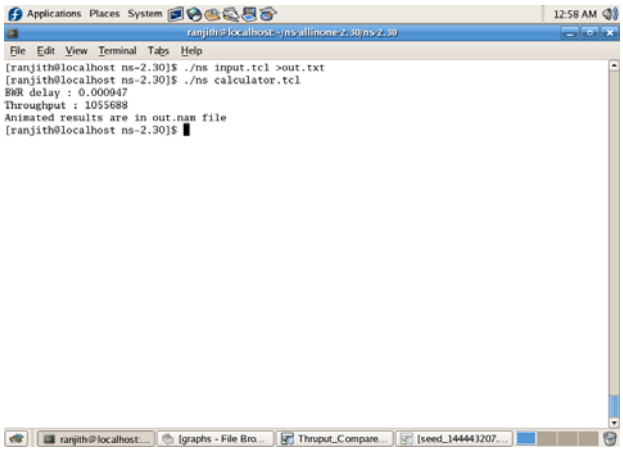


Figure 8

The calculator function takes the traced file, track the received times details and stored in the intermediate file by name recvtimes-f. The figure 10 shows the screenshot of Received time details.

A. Comparison of access delay (For single instance)

- Input given: Seed value = 144443207.
- No. Of SSs=6, 12, 18……..…60.
- WOC: Without circularity
- WC: With Circularity

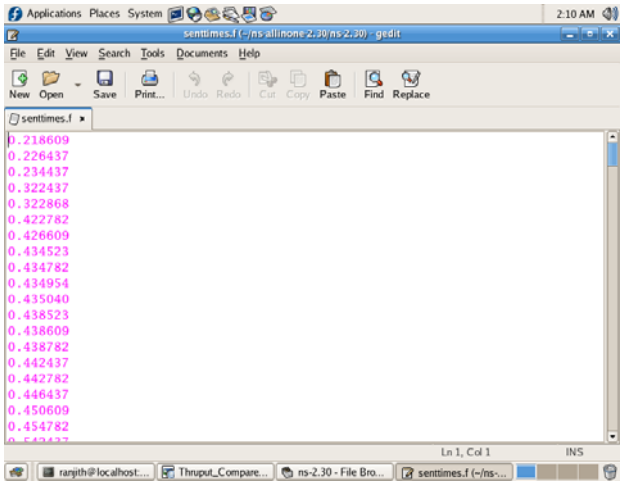


Figure 9

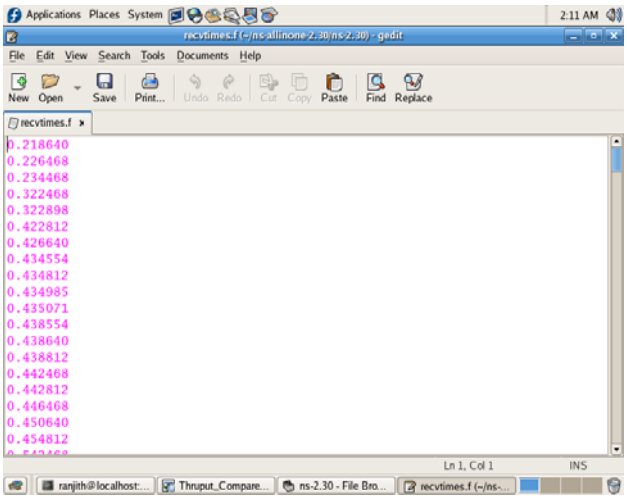


Figure 10

The outputs for the following sequence of subscriber stations (SSs) with and without circularity are observed and as shown in table.1

Table.1

SSs	WOC Access delay In seconds	SSs	WC Access delay In seconds
6	0.000929	6	0.000930
12	0.001000	12	0.000998
18	0.001016	18	0.001010
24	0.001019	24	0.001015
30	0.001023	30	0.001018
36	0.001028	36	0.001021
42	0.001035	42	0.001027
48	0.001060	48	0.001056
53	0.001066	53	0.001098
54	0.001070	54	0.001098
60	0.001104	60	0.001098

Figure 11 shows the graph of Access delay across various size of network with and without circularity.

We can see that, access delay is reduced in the case of enhanced network where we used the concept of circularity. As we see in the graph, the delay is gradually increased for both existing and new plans. But, we find there is better performance with the new scheme.

X-axis: No. Of subscriber stations Y-axis: Delay

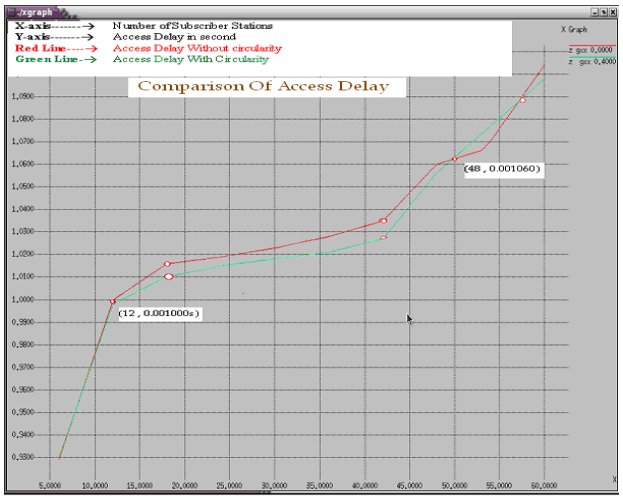


Figure 11

We have marked the two points in the graph where the two lines are crossed each other. At first point, the two lines start diverging from each other. It will cross each other once again at the second time where the network size increased.

B. Comparison of Throughput (For single instance)

- Input given: Seed value = 144443207.
- No. Of SSs = 6, 12…60.
- WOC: Without circularity
- WC: With Circularity
- TPT: Through put

The outputs for the following sequence of subscriber stations (SSs) with and without circularity are observed. The comparison of throughput is as shown in table 2.

Table.2

SSs	WOC TPT in Bits Per Second	SSs	WC TPT in Bits Per Second
6	8611410	6	8602150
12	8000000	12	8016032
18	7874015	18	7920792
24	7850834	24	7881773
30	7820136	30	7858546
36	7782101	36	7835455
42	7729468	42	7789678
48	7547169	48	7575757
53	7504690	53	7575757
54	7476635	54	7575757
60	7246376	60	7285974

The figure 12 shows the graph of comparison of the Throughput of the network by taking different number of SSs with and without circularity.

X axis: No. of subscriber stations Y axis: Throughput

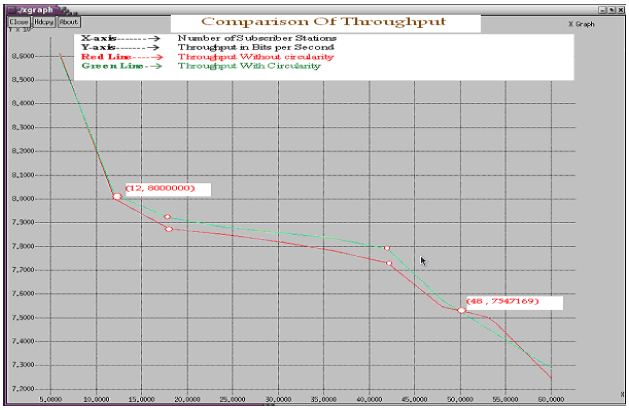


Figure 12

We can see increased Throughput in the case of enhanced network where we used the concept of circularity. In the above graph we can see that two times the two lines are crossed each other. We can see decreased throughput for the network where we included the concept of circularity. At the first co-incident point, the network size is less and the second time, the network size is increased.

C. Comparison of access delay (For 10 trials output)

When the results were run for 10 different seed value and the average is taken for each subscriber stations (SSs), for both with and without circularity we obtain the following graphs as shown in figure 13.

X-axis: No. of subscriber stations & Y-axis: Access Delay

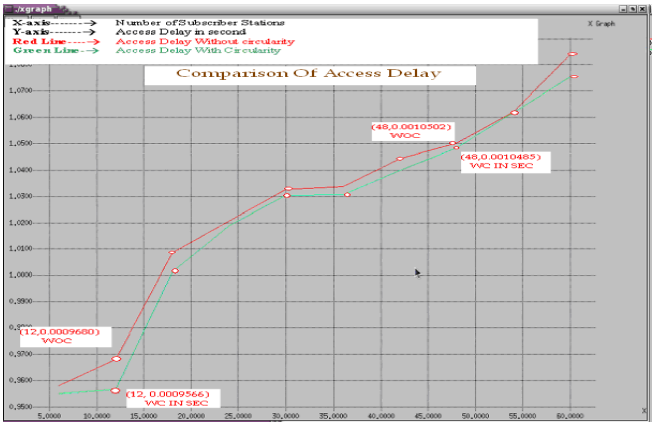


Figure 13

We can see that delay is decreased in the case of circularity implemented network.

D. Comparison of Throughput (For 10 trials output)

Figure 14 shows the comparison of throughput for both with and without circularity of 10 trials.

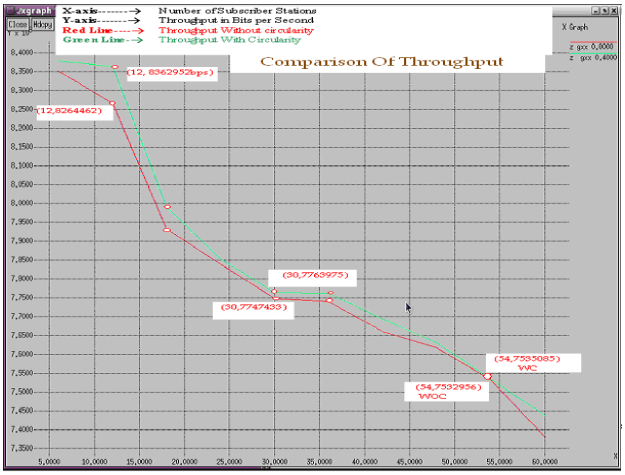


Figure 14

In figure 14 we have calculated the throughput from the previous access delays, which have been taken for ten trials. We can see the better performance in the case of circularity.

VII. CONCLUSION AND FUTURE ENHANCEMENT

The project involves the design of MAC IEEE 802.16. Here, new concept of circularity has been included in the existing contention resolution algorithm for MAC 802.16. In the existing MAC layer, there is certain amount of wastage in the available bandwidth. A concept of circularity is introduced whereby selected BWR slots are dropped in a controlled way in our new enhanced MAC protocol. We have seen the results of minimized uplink access delay, improved throughput and there by reduction in contention slots per frame. Hence circularity principle can be used to enhance the performance of MAC 802.16. We tried to reduce the collisions and hence have shown that concept of circularity can lead to higher throughput on an average of 12% more than the normal. For current design we have assumed the circularity on all SSs. This scheme can be modified in some particular SSs wherein we apply circularity concept to establish it is also more efficient than this scheme.

REFERENCES

[1] IEEE 802.16-2001, IEEE Standard for Local and Metropolitan Area Networks-Part 16: Air Interface for Fixed Broadband Wireless Access Systems, April 8, 2002.

[2] Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands; IEEE Computer Society and IEEE Microwave Theory and Techniques Society, 2005.

[3] Bhandari, B. N., Ratnam, K., Raja Kumar, and Maskad, X. L., Uplink Performance of IEEE 802.16 Medium Access Control (MAC) Layer Protocol, ICPWC 2005.

[4] Jeffrey G. Andrews, Arunabha Ghosh, and Rias Mohammed, Fundamentals of WiMAX Understanding Broadband Wireless Networking, Prentice Hall Inc.

[5] Ahmed Doha, and Hossam Hassanein, Access Delay Analysis in Reservation Multiple Access Protocols for Broadband Local and Cellular Network, Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks (LCN'04), 2004.

[6] IEEE Standard for Local and Metropolitan area networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems, IEEE Computer Society and IEEE Microwave Theory and Techniques Society, 2004.

[7] Intel Corp., IEEE 802.16 and WiMAX: Broadband Wireless Access for Everyone.*

[8] Mohammad Z. Ahmad, Damla Turgut, and R. Bhakthavathsalam, Circularity based Medium Access Control in Mobile Ad hoc Networks, 5th International Conference on AD HOC Networks and Wireless (ADHOCNOW'06), Aug.17-19, 2006, Ottawa, Canada.

[9] Marc Greis: Marc Greis 'Tutorial for the UCB/LBNL/VINT Network Simulator, http://www.isi.edu/nsnam/ns/tutorial/

[10] Seamless and Secure Mobility Tool Suite, Advanced Network Technologies Division, National Institute of Standards and Technologies.

[11] R.Bhakthavathsalam and Khurram J. Mohammed, Modification of Bandwidth Request Mechanism in MAC 802.16 for Improved Performance, ICWN'10, July 12-15, Las Vegas, USA.

[12] Bhakthavathsalam R, Prasanna S J, Enhancements to IEEE 802.11 MAC Protocol by Using the Principle of Circularity in MANETs for Pervasive Computing Environment, International Conference on Mobile, Ubiquitous & Pervasive Computing; 2006.

AUTHORS PROFILE

Dr. R.Bhakthavathsalam is presently working as a Senior Scientific Officer in SERC, IISc, Bangalore. His areas of interests are Pervasive Computing and Communication, Electromagnetics with a special reference to exterior differential forms. Author was a fellow of Jawaharlal Nehru Centre for Advanced Scientific Research (1993-1995). He is member of ACM & CSI.

[2] Dr. R.Shashikumar is presently working as a Professor in E & C Department, SJCIT, Chikballapur, Karnataka, India. He is having ten years of teaching and six years of Industry experience. His areas of interest include ASIC, FPGA, Network Security and Cryptography.

[3] Mr. V. Kiran is M.Tech student in the Department of Electronics and Communication, SJCIT, Chikballapur, Karnataka, India. His areas of interests include Pervasive computing, Network Security, Cryptography and Sensor Networks. He has an admirable level of interest in WiMAX networks.

[4] Mr. Y. R. Manjunath M.E, MISTE is working as a Senior Lecturer in the Department of Electronics and Communication, SJCIT, Chikballapur, Karnataka, India. He is having seven years of teaching experience. His areas of interest are Wireless Communication, Very Large System Integration, Digital Signal Processing and Embedded System.

An Intelligent Software Workflow Process Design for Location Management on Mobile Devices

N. Mallikharjuna Rao[1]
Associate Professor, Dept. of MCA
Annamacharya PG College of Computer Studies
Rajampet, Andhra Pradesh, India.
Email: drmallik2009@gmail.com

P.Seetharam[2]
System Engineer, Dept. of MCA
Annamacharya Institute of Technology & Sciences,
Rajampet, Andhra Pradesh, India
Email: seetharam.p@gmail.com

***Abstract*- Advances in the technologies of networking, wireless communication and trimness of computers lead to the rapid development in mobile communication infrastructure, and have drastically changed information processing on mobile devices. Users carrying portable devices can freely move around, while still connected to the network. This provides flexibility in accessing information anywhere at any time. For improving more flexibility on mobile devices, the new challenges in designing software systems for mobile networks include location and mobility management, channel allocation, power saving and security. In this paper, we are proposing intelligent software tool for software design on mobile devices to fulfill the new challenges on mobile location and mobility management. In this study, the proposed Business Process Redesign (BPR) concept aims at an extension of the capabilities of an existing, widely used process modeling tool in industry with 'Intelligent' capabilities to suggest favorable alternatives to an existing software workflow design for improving flexibilities on mobile devices.**

Key words: Wireless system, mobile, BPR, software design, intelligent design, Fuzzy database

I. INTRODUCTION

Technology improvements authorize the building of information systems which can be used at any place and at any time through mobile phones and wireless devices through networks. Mobile-wireless systems can create more benefits for organizations: e.g., productivity improvement processes and procedures flexibility, customer services improvement and information correctness for decision makers, which together stress competitive strategy, lower operation costs and improved business and subscribers processes.

A. Mobile-Wireless Systems

Schiller [2] [12] [13] [14] describes two mobility extents: one is user mobility, which allows creating the connection to the system from different geographical sites. A Mobile-wireless network is a radio network distributed over land areas called cells, each served by at least one fixed-location transceiver known as a cell site or base station. When joined together these cells provide radio coverage over a wide geographic area. This enables a large number of portable transceivers (e.g., mobile phones) to communicate with each other and with fixed transceivers and telephones anywhere in the network, via base stations, even if some of the transceivers are moving through more than one cell during transmission and second one is device mobility, which enables mobility of both user and devices, such as mobile phones.

B. Types of Mobile-Wireless System Applications

Mobile-wireless systems are classified in to two types [3]: Horizontal application and vertical applications.

1) Horizontal application

This is a flexible to an extensive range of users and organizations for retrieving the data from the devices, e.g. E-mail, browsers and file transfer applications.

2) Vertical Application

Vertical applications are precise to a type of users and organization. For example: financial applications, such as money transfer, stock exchange and information enquiry; marketing and advertising applications according to the actual user positions, i.e., pushing coupons to stores and information about sales nearby; emergency applications to check real-time information from government and medical databases and utility companies applications used by technicians and meter readers. In view of quick developments and usage with mobile devices are required to improve the software design approach for vertical applications for better outcome.

C. Problems in Mobile-Wireless Systems

Mobile-Wireless systems face some exceptional problems originating from the mobile devices. First, these devices have small memories, short battery life, and limited calculation and computation capabilities. Second, there is a wide variety of devices, possessing different characteristics, and the application must be adaptable and compatible to all of them. Third, the use of the devices is tight because of their size, tiny screens, low resolution, and small keyboards that are difficult to operate. Fourth, security problems can arise when devices are lost, due to possible illegal access to sensitive data. Fifth, location identification and profiles retrieval from database of the subscribers while the subscribers are in roaming. There is another problem we found, for probing the profiles from the Home Location Register and transfer to the Visitor Location Register while the subscribers are in roaming. It increases the transition between the HLR and VLR databases in mobile-wireless networks.

This study uses software engineering to define the mobile-wireless systems quality components and develops an approach to quantify these components, in order to enable the evaluation, comparison, and analysis of mobile-wireless system quality. Mobile-wireless systems must be measured on the basis of traditional systems e.g. easiness of maintainability, minimum complexity, lack of faults, and mean time between system failures in mobile device. In view of above, for increasing flexibility on wireless systems, we are introducing the concept of intelligent software workflow design for mobile devices. In this paper, we are proposing software workflow for only location management strategy on mobile devices.

On the other hand, fuzzy logic can help a lot for developing software for financial markets, environment control, project control and other scientific specific applications. It can help in detecting risks as early as possible in an uncertain environment. More important factor is that, it allows us to use common sense and knowledge for risk detection in software development. It provides an easy set of mathematical tools for assessing risks associated with software project management and software quality control within mobile devices for location based services. In section 2 we have discussed about Existing and its related work, section 3 we have presented proposed intelligent software workflow architecture, section 4 we have discussed about performance of the system and in section 5 we concluded the proposed system.

II. Related Work

The past fifty years of software engineering can be seen as a invariable hunt of developing software systems that are easier to design , cheaper to maintain , and more robust. Several milestones have been achieved in this effort, including data-oriented software architecture, object-oriented language, component-based software development and document-based interoperability. Among these achievements, in all the developments researchers have concentrated on two important concepts; they are data independence and process independence [4] [9] [10] [11].

Data independence can be determined as the robustness of business applications when the data structures are modified in traditional software developments. Fuzzy database will become dominant in the database market in future, generally as it achieves significant role in data independence for better results in future.

Process independence is a measure of the robustness of business applications when the process model is redesigned. The drive towards more process independence has lead to the increase in the workflow of the systems in software industry in the last few years. Very recently, major software companies have either acquired or developed workflow components for integration into their existing software platforms. Fuzzy database is having more scope for achieving many more decision making possibilities for redesigning the system with precise results.

Workflow supervision helps in achieving software flexibility by modeling business processes unambiguously and managing business processes as data that are much easier to modify than conventional program modules. Workflow management system enables reprocess of process templates, robust integration of enterprise applications, and flexible coordination of human agents and teams.

In present market, few Fuzzy Logic software tools are available to use in databases for solving complex problems faced by science and technology applications. They are presented in [5] they are: eGrabber DupChecker, Fuzzy System Component (Math Tools), Fuzzy Dupes 2007 5.6 Kroll Software-Development, Fuzzy Dupes Parallel Edition 32/64-Bit 6.0.3 (Kroll Software-Development), Sunshine Cards 1 (Free Spirits). In this paper, we took eGrabber DupChecker scenario for eliminating duplicate data from the database.

eGrabber DupChecker is a tool that uses advanced fuzzy logic technology to quickly identify hard-to-find duplicates in our Database. DupChecker automatically scans, identifies and groups duplicates for easy merging and de-duping.

A. Features

- *Robust Duplicate Checking*-Uses Fuzzy Logic Technology to quickly identify duplicates created through errors in data-entry, data-import and other variation such as Typo Errors, Pronunciation Errors, Abbreviation, Common nicknames, Match name with email ID, Company ending variations, Street ending variation, Formatting and punctuation errors, Field swap errors, Also recognizes same phone, regions and same locations.
- Merges all duplicates at one click on Merges notes, histories, activities, opportunities and all other details in Database.
- Retains history of all changes made to the merged record on databases
- Two plus merging - Merges multiple matches into one record

B. Benefits

- Maintains integrity of existing databases - Aggregates information lost across (notes, opportunities etc.,) duplicate contacts into one contact.
- Protects your company credibility - Avoid duplicated emails, postages or calls to customers or prospects, which would put your company creditability at stake.
- Saves time and money - Your sales team will not waste time calling duplicate contacts and you will cut down promotional cost spent for duplicate contacts.

Despite of recent significant developments in workflow technology and its prevalent acceptance in practice, current

workflow management systems still demand significant costs in system design and implementation. Furthermore, these systems are still lacking in their ability of handling exceptions and dealing with changes. In this paper, we study additional techniques based on intelligent workflow to incorporate more flexibility in conventional workflow management systems. Our research goal is to further improve software flexibility by achieving more process independences using fuzzy logic.

III. Proposed Method

Business Process Redesign (BPR) is a popular methodology for companies to boost the performance of their operations. In core, it combines an essential reformation of a business process with a wide-scale application of Information Technology (IT). However, BPR on the work flow currently more closely resemble the performance of an art than of science. Available precise methods hardly find their way to practice.

The keywords for BPR are 'Fundamental', 'Radical', 'Dramatic', 'Change' and 'Process'. A business process has to undergo fundamental changes to improve productivity and quality. Radical changes, as opposite to incremental changes, are made to create dramatic improvements. Reengineering is not about fine-tuning or marginal changes. BPR is for provoked companies like mobile service providers; that are willing to make important changes to achieve significant performance improvements in their organization.

BPR is a structured approach for analyzing and continually improving fundamental activities such as manufacturing, marketing, communications and other major elements of a company's operation. Wright and Yu (1998) defined the set of factors to be measured before actual BPR starts and developed a model for identifying the tools for Business Process Redesign. The BPR is to develop a framework for understanding Business Process Re-engineering and to explain the relationship between BPR and Total Quality Management (TQM), Time-based competition (TBC) and Information Technology (IT). BPR should enable firms to model and analyze the processes that supports products and services, high-light opportunities for both radical and incremental business improvements through the identification and removal of waste and inefficiency, and implement improvements through a combination of Information Technology and good working practices in their work places.

A. A framework for BPR modelling and analysis.

The proposed framework has been presented for location based services on mobile devices, to offer some guidelines for choosing suitable tools/techniques for BPR applications. The guidelines are based on the areas to be reengineered for dramatic improvements in the performance.

1) BPR Strategies

Decision making at strategic levels would require intelligent systems to select the appropriate strategies and methods with the objective of making decisions about business location, product portfolio, funding for project work, etc. this requires taking into account the risk involved and the costs and benefits of running the business. At strategic levels, aggregate and fuzzy data are used to make a decision for long-term developments and changes in an organization. The type of decisions requires experience and knowledge in selecting the most suitable methods.

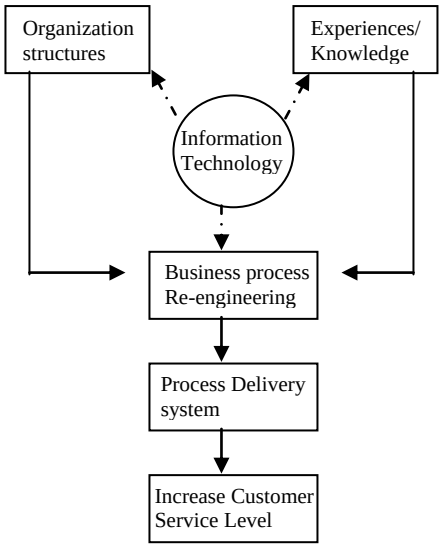


Figure 1: A Conceptual Intelligent Business model

A Conceptual Intelligent Business Model as shown in figure 1; which contains organization structures, knowledge data interact with information technology for more and instant decisions to redesign and increasing the outcome of the devices. Therefore, the selection of the tool for BPR depends upon:

1) The nature of decision areas.
2) The nature of data to be analyzed.
3) The background of users.

The decision area here is to formulate strategies for reengineering business processes. The nature of data available at the strategic level is generally not accurate at this level of decision making, and therefore models based on a system approach and conceptual framework could be used to analyze the data. Knowledge-based models are user-friendly, but have limited applications considering the areas of reengineering.

In this study, the proposed Intelligent Software Workflow process design as shown in figure 2; consists of several blocks. In this system, all the system protocols are connected to the set of attributes to software design they can automatically reflect the design process on devices and device performance. As we defined already, fuzzy logic is an intelligent system which can produce precise data values; they will give ultimate outcome for the devices. Fuzzy logic is a rule based system, with number of rules it can generate more and more combination of

results which can help to re-construct or re resign based on the values produced by the system.

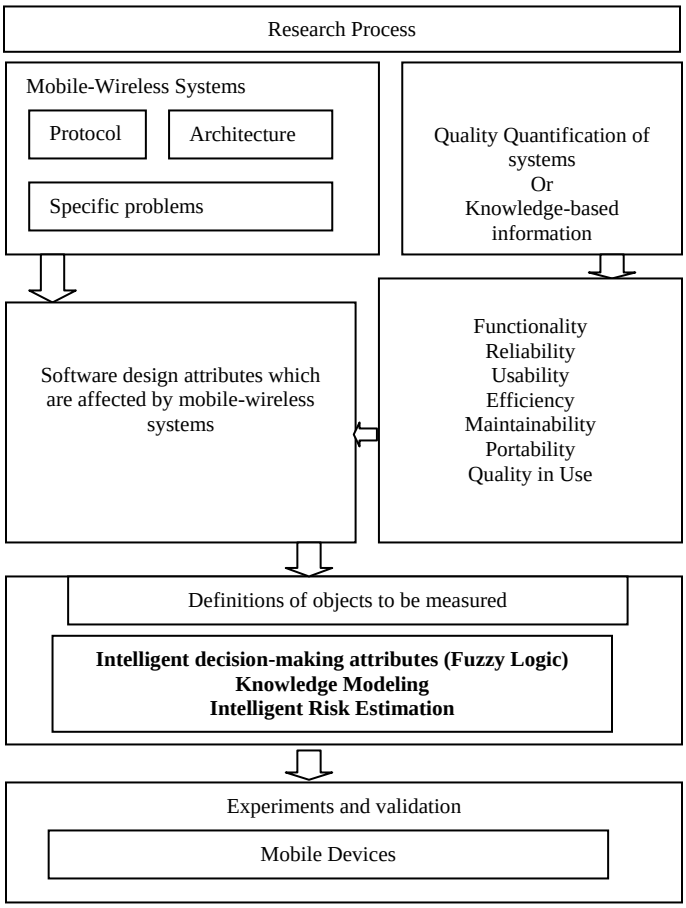


Figure 2: Intelligent Software Workflow Process Design Architecture

This can increase the functionality of the system; it includes suitability to find a call, accuracy on making the calls, interoperability and security. Increase the Reliability which includes; maturity, fault tolerance and recoverability. During mobility, network problems, hiding obstacles and hopping between antennas may disturb and interrupt communications.

In Usability, this includes the understandability, learnability, operability and attractiveness. Mobile users may not be able to focus on the systems use, so the application should not be complicated, the input must be easy to insert, intuitive and simplified by using location aware functions. Efficiency includes the time behavior and resource utilization. Time behavior in the wireless environment is important because the price of each minute of data transferring is very high, and the users will avoid expensive systems.

Maintainability includes the analyzability, changeability, stability and testability. Portability includes the adaptability, installability, co-existence and replacebility.

The following resources can be used in mobile devices to design the workflow process for location management. They are: Mobile Network Code (MNC), Location Area Code (LAC), Base Station Identity (CID) and Radio Frequency (RF) Signal– The relevant signal strength between your current tower and the adjacent one.

In this study, we took above attributes as vague input values and processed through the proposed workflow system, it makes many more decisions and knowledge based decisions and the outcome of this system produces well behaved work and timeless connection and transfer of text messages from the one terminal to another terminal without interruptions. It will increase the performance, stability and operational flexibility on mobile devices while the subscribers are in roaming.

In the next section, we have discussed few important aspects to design best and better software workflow design using fuzzy logic which enables flexibility on usage and decrease the transmission delays.

B. Knowledge Modeling

Knowledge is information combined with aptitude, framework, investigation, and suggestion. It is a high-value form of information that is ready to apply to decisions and actions. Knowledge Modeling packages combinations of data or information into a reusable format for the purpose of preserving, improving, and allocation, aggregate and processing Knowledge to simulate intelligence. In intelligent systems quality of the knowledge base and usability of the system can be preserved by filtering the domain expert is involved in the process of knowledge acquisition but also in other phases until the testing phase. The quality and usability is also dependent of the end users consideration why these users would be involved in the modeling process and until the system has been delivered.

Meta-rules are rules that are combining ground rules with relationships or other meta-rules. Thereby, the meta-rules tightly connect the other rules in the knowledge base, which becomes more consistent and more closely connected with the use of meta-rules.

Vagueness is a part of the domain knowledge and refers to the degree of plausibility of the statements. This is easy to get outcome from something in between, e.g. unlikely, probably, rather likely or possible, the rule based representation shown in the figure 3.

.

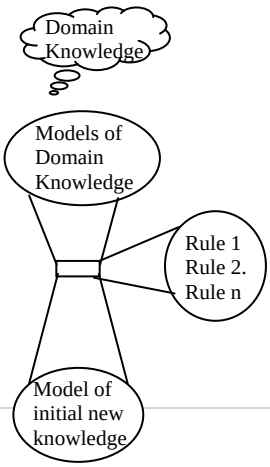


Figure 3: Knowledge modeling

Knowledge base contains a number of fuzzy linguistic variables (fuzzy sets), and a number of fuzzy inference rules. Semantics fuzzy linguistic variable (fuzzy sets) are defined by their membership functions based on software metrics. It contains a list of fuzzy inference rules about risk detection across all phases in software life cycles. Fuzzy rules are basically of "IF-THEN" structure

Fuzzy inference rules are presented in antecedent-consequences structure. Antecedents represent symptoms of software artifacts, processes or organization in terms of risks based on the 'Dimensional Analytical Model' rules use individual metrics and their combinations based on their relational ships from different dimensions.

C. Intelligent Risk Estimation

Intelligent risk estimation is the result of fuzzy inference engine as shown in figure 4, which is the central processing part of the intelligent software workflow system. In this system all the metrics have certain values and it also have the three dimensions. Three dimensions are connected to Inference Engine for making more decisions. Inference engine apply the knowledge base on this set of inputs to produce worth risk, which is called schedule low risk, standard risks and high risk. The input and output sets are stored over the time period for analysis purpose [7]. This analysis helps the mobile location management software designers to find the maximum predictable risk for the systems, even network failures or minimum number of cell towers or even low frequency. This system can produced LR (Low Risk), SR (Standard Risk) and HR (High Risk).

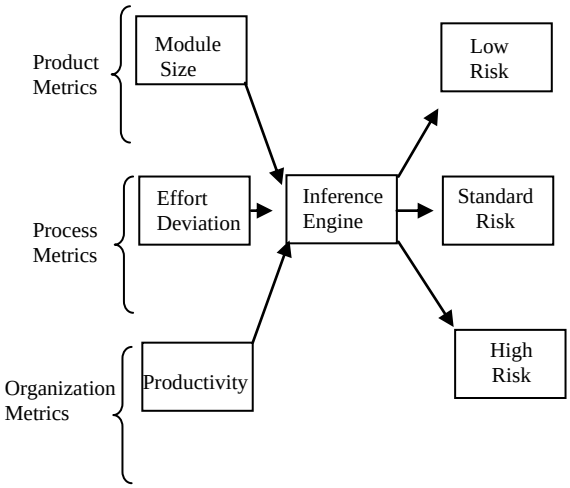


Figure 4: Fuzzy Inference Engine for Intelligent Risk Assessment

In this study, for finding risk estimation for location identification on mobile devices, measures the number of mobiles connected to networks, networks size and all other protocols are required to include into the system for finding the risk of the system. Fuzzy inference engine is evaluating the risk factor based set of inputs, Module size, effort deviation and Productivity. Inference engine take the inputs of the systems and it produces the out of low risk, standard risk and high risk factors. We are given some set of conditions rules for improving the system. They are

1) Product Metric based rule
IF Volatility index of subsystem is HIGH
AND Requirements quality is LOW
THEN Schedule Risk is VERY HIGH

2) Process Metric based rule
IF Manpower is HIGH
AND Design approaches are HIGH
THEN Product Service is HIGH

3) *Organization Metric based Rule*
IF Effort deviation is HIGH
AND Customer involvement is HIGH
THEN Risk of schedule is VERY HIGH

IV. Simulations On Proposed System

In this study, intelligent software workflow system has been developed for fuzzy query, which makes it possible to query on a conventional databases. The developed software fuzzifies the fields, which are required to be fuzzified, by connecting to the classical database and creates a supplementary database. This database includes fuzzy data and values which are related with the created membership functions. In the recent years, fuzzy query usage is increase in all the areas in the intelligent systems.

A database, called as "Subscriber profile", which includes some information about the mobile subscribers. Suppose if a query for the approximation of the necessity of the subscriber's profiles is required, in order to determine the detail of the Subscribers in normal database is difficult. For example, the SQL query for the particular assessment problem may be in the form of the following.

SELECT subscriber_name, imei#, sim#, La, mobile#, bill_payment FROM *SUBSCRIBER_PROFILE* WHERE bill_payment <=3000

Same query if we convert into fuzzy SQL will be as follows:

SELECT subscriber_name, imei#, sim#, La, mobile#, bill_payment FROM *SUBSCRIBER_PROFILE* WHERE bill_payment is HIGH or more than 3000

This means we will get the subscribers those who have their payment by more than 3000 or payment HIGH subscribers.

A. FUZZY QUERY SOFTWARE TOOL

It is intrinsically robust since it does not require precise, noise-free inputs and can be programmed to fail safely if a feedback sensor quits or is destroyed. The output control is a smooth organize function despite a wide range of input variations. The membership function is a pictographic representation of the degree of participation of each input. It associates a weighting with each of the inputs that are processed, define functional overlap between inputs, and ultimately determines an output response. Once the functions

are indirect, scaled, and combined, they are defuzzified into a crisp output which drives the system.

.

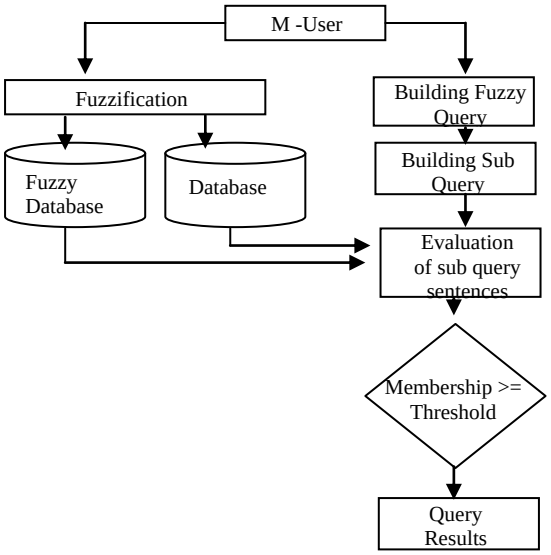


Figure 5: The Block diagram of the fuzzy tool

As it is known, a fuzzy set is a set, which contains elements having varying degrees of membership. Elements of a fuzzy set are mapped into a universe of "membership values" using a function-theoretic form. Fuzzy sets are denoted by a set symbol. As an example, let "A" denote a fuzzy set. Function μ maps the elements of fuzzy set "A" to real numbered value on the interval 0 to 1. If an element in the universe, say x, is member of the fuzzy set "A", then this mapping is given as:

$$\mu_A(x) \in [0,1]$$

$$A = (x, \mu_A(x) / x \in X)$$

$$A = \sum \mu_A(x_i) \Big/ x_i = \mu_A(x_1) + \mu_A(x_2) + ... + \mu_A(x_n) / x_n$$

Using above equation, we can calculate the membership degrees for crisp variables in a system.

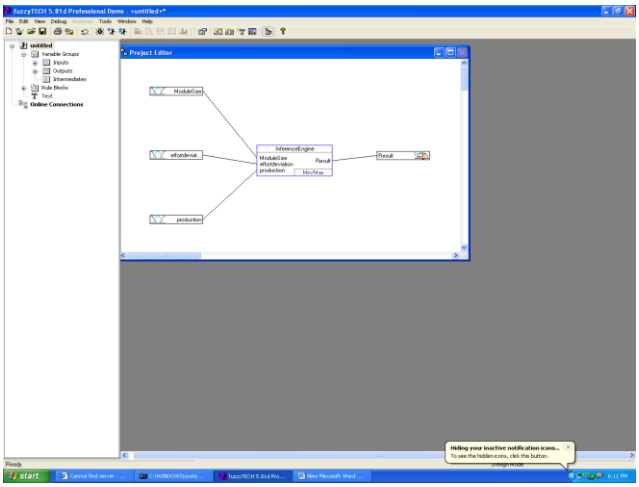


Figure 6: Fuzzy variable editor making fuzzy decisions

In this study, we have presented fuzzy representation for the risk estimation using rule based crisp values and membership degrees as shown in figure 6 and figure 7

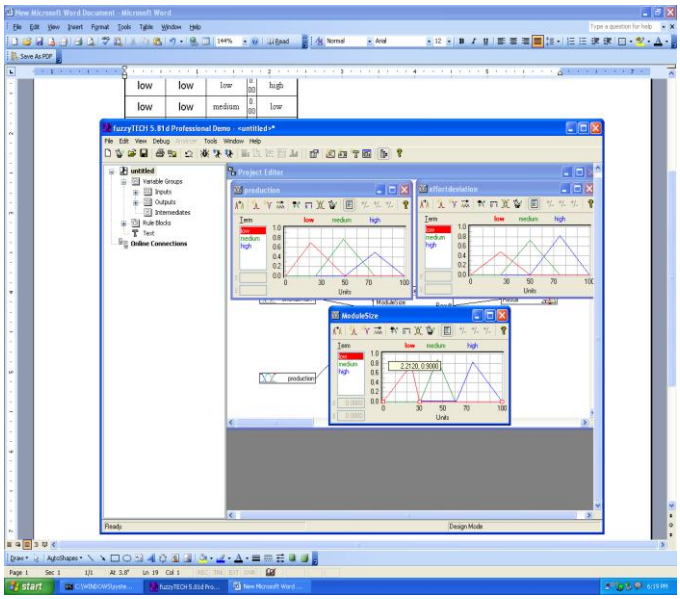


Figure 7: Membership representation

CONCLUSION

This paper introduces new concept called intelligent software design. First, it describes mobile-wireless systems and stated the new challenges faced by mobiles devices on software modeling. Second, this study gives better solution for location and mobility management software designs to improve the flexibility in transferring and processing the information from one location to another location. This Intelligent software design tool can be expanded to new kinds of mobile-wireless systems, emerging because of the rapid development of the technology and the wireless networks.

REFERENCES

[1] http://www.stw.nl/projecten/E/eit/eit6446.htm

[2] Schiller, J. Mobile Communications. Addison-Wesley

[3] Stafford, T.F., & Gillenson, M.L. Mobile Commerce: What it is and what it could be. Communications of the ACM, 46(12), 33-44

[4] Daniel D Zeng, J.Leon Zhao," Achieving Software Flexibility via Intelligent workflow Techniquies", Proceedings of 35th Hawaii International conference on system sciences-2002

[5] http://www.fileguru.com/apps/fuzzy_logic_design

[6] Ruti Gafni, "Framework for quality metrics in mobile-wireless information systems" Interdisciplinary Journal of Information, Knowledge and management, volume 3, 2008.

[7] Xiaoqing Liu , Goutam Kane and Monu Bambroo, "An Intelligent early warning system for software quality improvement and project management", IEEE, Proceedings of the 15th International conferences on Tools with Artificial Intelligence (ICTAI'03)

[8] Shu Wang, Jungwon Min and Byung K. Yi, "Location Based Services for Mobiles: Technologies and Standards",IEEE International Conference on Communication (ICC) 2008, Beijing, China.

[9] Ye Feng, Liu Xia, "Workflow-based office Automation Decision-making Support System", 2009 Asia-Pacific Conference on information Processing

[10] Qian Qi, Yinhui Chen, "Workflow-Based Process Management of Networks Management Testing", IEEE , 2009 Third International Symposium on Intelligent Information Technology Applications Workshops.

[11] John Noll, "A Peer-to-peer Architecture for Workflow in Virtual Enterprises", IEEE, Proceedings of the Fifth Inernatinal Conference on Quality Software (QSIC'05).

[12] Zhiyong Weng and Thomas Tran, " An Intelligent Agent-Based Frame work for Mobile Business", IEEE, Sixth International Conference on the management of Mobile Business (ICMB 2007).

[13] Perakath C. Banjamin, charless Marshall and Richard J.Mayer , "A workflow analysis and design Environment (WADE), Poceedings of the 1995 Winter Simulation Conference.

[14] www.computer .org/intelligent/ieee intelligent

[15] Yolanda Gil , Varun Ratnakar and Jihie Kim, WINGS: Intelligent Workflow-based Design of Computational Experiments", IEEE Publication.

[16] Artem Chebotko, Shiyong and Seunghan, "Secure Abstraction Views of Scientific Workflow Provenance Querying", IEEE Transactions on services computing 2010.

AUTHORS PROFILE

N.Mallikharjuna Rao is presently working as Associate Professor in the department of Master of Computer Applications at Annamacharya PG college of Computer Studies, Rajampet and having more than 12 years of Experience in Teaching UG and PG courses. He received his B.Sc Computer Science from Andhra University in 1995, Master of Computer Applications (MCA) from Acharya Nagarjuna University in 1998, Master of Philosophy in Computer science from Madurai Kamarj Univeristy, Tamlinadu, in India and Master of Technology in Computer Science and Engineering from Allahabad University, India. He is a life Member in ISTE and Member in IEEE, IACSIT. He is a research scholar in Acharya Nagarjuna University under the esteemed guidance of Dr.M.M.Naidu, Principal, SV.Univeristy, Tirupathi, and AP.

Mr. P.Seetharam is working as Systems Engineer in the Department of MCA at Annamacharya Institute of Technology & Sciences, Rajampet. He has more than 8 years of Experience in the areas of Wireless and Structured Network Technology. He is a Certified Systems Engineer from Microsoft and with a master degree from Information Technology.

Pattern Discovery using Fuzzy FP-growth Algorithm from Gene Expression Data

Sabita Barik[1], Debahuti Mishra[2], Shruti Mishra[3], Sandeep Ku. Satapathy[4], Amiya Ku. Rath[5] and Milu Acharya[6]
[2,3,4,6] Institute of Technical Education and Research, Siksha O Anusandhan University, Bhubaneswar, Odisha, INDIA
[1]Department of Computer Science and Engineering, Gandhi Engineering College, Bhubaneswar, Odisha, INDIA
[5]Department of Computer Science and Engineering, College of Engineering Bhubaneswar, Odisha, INDIA
sabita.barik10@gmail.com ,debahuti@iter.ac.in,shruti_m2129@yahoo.co.in, sandeepkumar04@gmail.com,
amiyamaiya@rediffmail.com and milu_acharya@yahoo.com

***Abstract*- The goal of microarray experiments is to identify genes that are differentially transcribed with respect to different biological conditions of cell cultures and samples. Hence, method of data analysis needs to be carefully evaluated such as clustering, classification, prediction etc. In this paper, we have proposed an efficient frequent pattern based clustering to find the gene which forms frequent patterns showing similar phenotypes leading to specific symptoms for specific disease. In past, most of the approaches for finding frequent patterns were based on Apriori algorithm, which generates and tests candidate itemsets (gene sets) level by level. This processing causes iterative database (dataset) scans and high computational costs. Apriori algorithm also suffers from mapping the support and confidence framework to a crisp boundary. Our hybridized Fuzzy FP-growth approach not only outperforms the Apriori with respect to computational costs, but also it builds a tight tree structure to keep the membership values of fuzzy region to overcome the sharp boundary problem and it also takes care of scalability issues as the number of genes and condition increases.**

Keywords: Gene Expression Data; Association Rule mining; Apriori Algorithm, Frequent Pattern Mining, FP-growth Algorithm

I. INTRODUCTION

Gene expression data is arranged in a data matrix, where each gene corresponds to one row and each condition to one column. Each element of this matrix represents the expression level of a gene under a specific condition and it is represented by real value [2]. Main objective of the gene expression are.: first, grouping of genes according to their expressions under the multiple conditions and second, grouping of conditions based on the expression data.

It helps the molecular biologist in many aspects, like gathering information about different cell states, functioning genes, identifying gene that reflects biological process of interest etc. The importance of gene expression data is to identify genes whose expression levels that reflects biological process of interest (Such as development of cancer). The analysis can provide clues and guesses for the functioning of genes.

Frequent patterns are pattern (such as item sets, subsequences or substructures) that appear in a data set frequently. In this paper we are interested in using association rule mining for discovering the gene expression patterns those exhibit similar expression levels and by discovering such patterns it can be helpful for medical diagnosis. One of the oldest Apriori algorithms is based upon the Apriori property that for an item set to be frequent; certainly any subset must be frequent. The Apriori algorithm employs an iterative approach known as a level wise search, where k-items are used to explore (*k+1*) item sets. Apriori algorithm suffers by scanning the database while finding the *k*-item sets in each and every step. Due to which the processing overhead is drastically increased. FP-growth algorithm which is enhanced version of apriori algorithm gives better performance while scanning the data base.

For finding the large set of data items we need association rule mining. Association rule mining finds interesting association and correlation relationship among a large set of data items. Frequent itemset mining leads to the discovery of associations and correlation relationship among genes and conditions in large data bases or data sets with massive amounts of data. The discovery of interesting correlation relationship among gene and conditions can help in medical diagnosis, gene regulatory network etc. FP-growth algorithm is that takes less time to search the each level of the tree. The FP-growth method transforms the problem of finding long frequent patterns to searching for shorter ones recursively and then concatenating the suffix. It uses the least frequent items as a suffix, offering good selectivity. The method substantially reduces search cost.

In this paper, we have used FP-growth mining algorithm hybridized with Fuzzy logic to find interesting association and correlation relationship among a large set of data items. Fuzzy logic is used to find out the support and confidence using the membership function μ = [0, 1]. Fuzzy sets can generally be viewed as an extension of the classical crisp sets. They have been first introduced by Lofti A. Zadeh in 1965 [14]. Fuzzy sets are generalized sets which allow for a graded membership of their elements. Usually the real unit interval [0;1] is chosen as the membership degree structure. Crisp sets are discriminating between members and nonmembers of a set by assigning 0 or 1 to each object of the universal set. Fuzzy sets generalize this function by assigning values that fall in a specified range, typically 0 to 1, to the elements. This evolved out of the attempt to build a mathematical model which can display the vague colloquial language. Fuzzy sets have proofed to be useful in many areas where the colloquial language is of influence. The support and confidence framework can be generalized using fuzzy membership values for not considering the crisp nature of clustering.

A. Proposed Model

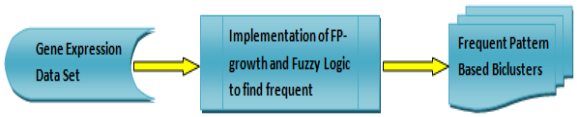


Figure 1: Our Proposed Model

Our proposed work is to find the frequent patterns from gene expression data using FP-growth algorithm which is the enhanced version of Apriori. FP-growth algorithm constructs the conditional frequent pattern (FP) tree and performs the mining on this tree. FP-tree is extended prefix tree structure, storing crucial and quantitative information about frequent sets. FP-growth method transforms the problem of finding long frequent patterns to search for shorter once recursively and then concentrating the suffix. Then, to hybridize the fuzzy logic to map the support and confidence measures to the membership value to $\mu = [0, 1]$. Finally, we validate our hybridized model by comparing our model with existing apriori models by considering various parameters. Our model (See figure 1) outperforms the apriori model on the basis of run time for finding number of patterns and also the scalability issues have been found to be improved significantly considering both the attributes and objects as they increases.

B. Paper Layout

This paper is arranged in the following manner, section I gives the introduction as well as our proposed model is outlined, section II deals with related work on frequent pattern mining. In section III the preliminary information about gene expression data, apriori algorithm, FP-growth approach and algorithms are described. Section VI describes our proposed algorithm. Section V gives the analysis of our work and shows its significance over the Apriori algorithm. Finally, section VI gives the conclusion and future directions of our work.

II. Related Work

Association rule mining finds interesting association and correlation relationship between genes [1]. A key step in the analysis of gene expression data is to find association and correlation relationship between gene expression data. DNA microarrays allow the measurement of expression levels for a large number of genes, perhaps all genes of an organism within a number of different experimental samples. It is very much important to extract biologically meaning the information from this huge amount of expression data to know the current state of the cell because most cellular processes are regulated by changes in gene expression. Gene expression data refers to a distinct class of clustering algorithms that performs simultaneous row-column clustering [2]. An imposed version of depth first search, the depth fast implementation of Apriori as devised in [8] is presented. Given a database of (e.g. supermarket) transactions, the DF algorithm builds a so called trie that contains all frequent item sets i.e. is all item sets that are contained in at least minimum suppport transactions with minimum suppport a given threshold value. There is a one to one correspondence between the paths and the frequent item sets. The new version, so called Depth Fast search, differs from Depth fast in that its data structure representing the data base is borrowed from the FP-growth algorithm. Frequent item sets play an essential role in many data mining tasks that try to find interesting patterns from databases, such as association rules, correlations, subsequences, episodes, classifiers, cluster and many more of which the mining association rule is one of the most popular problems[4].

The original motivation for searching association rules came from the need to analyze so called supermarket transaction data that is to examine customers' behavior in terms of the purchased products. Such rules can be useful for decisions concerning product pricing, promotions, store layout many others. With the rapid progress of bioinformatics of post genomic era, more and more bio information needs to be analyzed [6]. Many researchers have been also focused on using association rule mining method to construct the gene regulatory network. The existing approach of mining frequent pattern in microarray data set are item are enumeration and sample enumeration. Association rules, used widely in the area of market basket analysis, can be applied to the analysis of expression data as well [7]. Association rules can reveal biologically relevant associations between different genes or between environmental effects and gene expression. Item in gene expression data can include genes are highly expressed or repressed, as well as relevant facts describing the cellular environment of the genes e.g. the diagnosis of the tumor sample from which a profile was obtained.

III. Preliminaries

A. Gene Expression Data

One of the reasons to carry out a microarray experiment is to monitor the expression level of genes at a genome scale. Patterns could be derived from analyzing the change in expression of the genes, and new insights could be gained into the underlying biology. The processed data, after the normalization procedure, can then be represented in the form of a matrix, often called gene expression matrix (See table 1). Each row in the matrix corresponds to a particular gene and each column could either correspond to an experimental condition or a specific time point at which expression of the genes have been measured. The expression levels for a gene across different experimental conditions are cumulatively called the gene expression profile, and the expression levels for all genes under an experimental condition are cumulatively called the sample expression profile. Once we have obtained the gene expression matrix, additional levels of annotation can be added either to the gene or to the sample. For example, the function of the genes can be provided, or the additional details on the biology of the sample may be provided, such as disease state or normal state.

Table 1 Gene Expression Data Set

	C_1	C_2	C_3	C_4
$Gene_1$	10	80	40	20
$Gene_2$	100	200	400	200
$Gene_3$	30	240	40	60
$Gene_4$	20	160	60	80

B. Association Rule mining for frequent pattern

Association rule mining [12] finds interesting association and correlation relationship among a large set of data items.

The rules are considered interesting if they satisfy both a minimum support threshold and minimum confidence threshold. In general, association rule mining can be viewed as a two-step process:

- Find all the frequent item sets: by definition, each of these item sets will occur at least as frequently as a predetermined minimum support count, minimum support.
- Generate strong association rules from the frequent item sets; by definition these rules must satisfy minimum support and minimum confidence.

An association rule is an implication of the form $A => B$, where A sub set of I, B sub set of I , $A \cap B = \Phi$. The rule $A => b$ holds in the transactions set D with support s, where s is the percentage of transaction in D that contain $A \cup B$. this is taken to be probability, $P(A \cup B)$

Support $(A => B) = P(A \cup B)$ Equation (1)

The rule $A=>B$ has confidence C in the transaction set D, where C is the percentage of transaction in D containing A that also contain B. This is taken to be conditional probability $P(B/A)$.

Confidence $(A=>B) = P(B/A)$ Equation (2)

C. Apriori Algorithm

Apriori algorithm [5] is proposed by Rakesh Agrawal and Ramakrishnan Srikant from IBM Almaden Research Center. It is regarded as one of the breakthroughs in the progress of frequent pattern mining algorithms. Apriori employs an iterative approach known as level wise search, where k –item sets are used to explore k +1-itemsets. First, the set of frequent 1-itemsets is found. This is denoted as 1 L . 1 L is used to find 2 L, the frequent 2-itemsets, which is used to find 3 L, and so on, until no more frequent k –item sets can be found. The finding of each k L requires one full scan of the database. Throughout the level-wise generation of frequent item sets, an important anti-monotone heuristic is being used to reduce the search space. This heuristic is termed as Apriori *heuristic*.

The Apriori heuristic: *if any length k pattern is not frequent in the database, its length k* +1 *super-pattern can never be frequent*.

D. FP-growth Approach

It is noticed that the bottleneck of the Apriori method rests on the *candidate set generation and test*. In [11], a novel algorithm called FP-growth is proposed by Jiawei Han, Jian Pei and Yiwen Yin. This algorithm is reported to be an order of magnitude faster than the Apriori algorithm. The high efficiency of FP-growth is achieved in the following three aspects. They form the distinct features of FP-growth meanwhile. First, an extended prefix tree structure, called *frequent pattern tree* or FP-tree for short is used to compress the relevant database information. Only frequent length-1 items will have nodes in the tree, and the tree nodes are arranged in such a way that more frequently occurring nodes will have better chances of sharing than less frequently occurring nodes. Second, an FP-tree-based pattern fragmentation growth mining method FP-growth is developed. Starting from a frequent length-1 pattern (as an initial suffix pattern), FP-growth examines only its *conditional pattern base* (a "sub-database" which consists of the set of frequent items co-occurring with the suffix pattern), constructs its *conditional* FP-tree and performs mining recursively on such a tree. The pattern growth is achieved via concatenation of the suffix pattern with the new ones generated from a conditional FP-tree. Since the frequent pattern in any transaction is always encoded in the corresponding path of the frequent pattern trees, pattern growth ensures the completeness of the result.

In this context, FP-growth is not Apriori-like *restricted generation-and-test* but *restricted test only*. The major operations of mining are count accumulation and prefix path count adjustment, which are usually much less costly than candidate generation and pattern matching operations performed in most Apriori-like algorithms. Third, the search technique employed in mining is a *partition-based, divide-and conquers method* rather than Apriori-like *bottom-up generation of frequent patterns combinations*. This dramatically reduces the size of *conditional pattern base* generated at the subsequent level of search as well as the size of its corresponding *conditional* FP-tree. Inherently, it transforms the problem of finding long frequent patterns to looking for shorter ones and then concatenating with the suffix. The novelty of FP-growth is that to generate all frequent patterns, no matter how long they will be; only 2 database scans are needed. One is used to find out frequent 1-item sets, the other is used to construct a FP-tree. The remaining operation is recursively mine the FP-tree using FP-growth. FP-tree resides in main-memory and therefore FP-growth avoids the costly database scans.

E.Mining the FP-Tree using FP-Growth

The FP-Tree provides an efficient structure for mining, although the combinatorial problem of mining frequent patterns still has to be solved. For discovering all frequent item sets, the FP-Growth algorithm takes a look at each level of depth of the tree starting from the bottom and generating all possible item sets that include nodes in that specific level. After having mined the frequent patterns for every level, they are stored in the complete set of frequent patterns. FP-Growth takes place at each of these levels. To find all the item sets involving a level of depth, the tree is first checked for the number of paths it has. If it is a single path tree, all possible combinations of the items in it will be generated and added to the frequent item sets if they meet the minimum support. If the tree contains more than one path, the conditional pattern base for the specific depth is constructed. Looking at depth *a* in the FP-Tree of figure 2, the conditional pattern base will consist of the following item sets: <*b, e* :1> , <*b, d* :1> and <*d, e* :1 > . The item set is obtained by simply following each path of *a* upwards.

Table 2: Conditional pattern base

Item	Conditional pattern
a	**{<b,e:1>,<d,e:1>}**
e	**{<b,2>, <b,d:1> <d:1>**
d	**{<b:3>}**
b	**ϕ**

F. FP-Tree Construction

Constructing an FP-Tree from a fuzzy database is rather straightforward. The only difference to the original algorithm is that it is not enough to count the occurrences of an item, but the value of membership has to be considered as well. This membership value is then simply added to the overall count of the node. The database of fuzzy values in Table 3 is used for FP-Tree construction. Generating the FP-Tree from this database will lead to a tree containing the sums of the fuzzy values in the nodes. (See figure 2)

Table 3: Fuzzy Database

A	B	C	D	D
2	0	0.6	0	**1**
0	0.3	1	0	**0**
0.1	1	1	1	**0**
0	0.2	0	0.4	**1**
0	0	0.2	0	**0.2**
0	0	0.1	1	**0.1**
1	**1**	**0**	**0**	**0.1**

It is now easy to calculate the support of a path in the tree because it is simply the minimum of the path that is controlled. That tree can be used for conducting the FP-Growth algorithm.

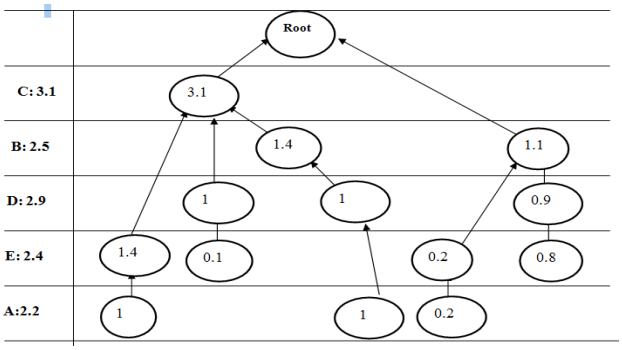


Figure 2: Fuzzy FP-Tree

F. Hybridizing FP-growth algorithm with Fuzzy Logic

Traditional quantitative association rule mining methods partition continuous domains into crisp intervals. When dividing an attribute into intervals covering certain ranges of values, the *sharp boundary problem* arises. Elements near the boundaries of a crisp set (interval) may be either ignored or overemphasized [13]. Fuzzy association rules overcome this problem by defining fuzzy partitions over the continuous domains. Moreover, fuzzy logic is proved to be a superior technology to enhance the interpretability of these intervals. Hence, fuzzy association rules are also expressions of the form $X \rightarrow Y$, but in this case, X and Y are sets of fuzzy attribute-value pairs and fuzzy confidence and support measure the significance of the rule.

FP-Growth algorithm does not produce the candidate item-sets, but it uses growth models to mine frequent pattern [13]. It compresses the database into a frequent pattern tree, but still retains itemsets associated information. Then it will divide such a compressed database into a group of conditional databases, each one of which is related to a frequent item-set and mined. FP-growth algorithm is made up for these deficiencies. FP-growth algorithm does not produce frequent item-sets, replaced by an FP-tree. At the same time, FP-growth algorithm need transaction database only twice, which makes in the proceedings as soon as possible to close out database, which occupies a large number of storage, can be achieved.

IV. Our Proposed Algorithm

In this paper, we are interested in finding the frequent patterns of genes and conditions which leads to discover the symptoms of diseases. Though, FP-growth algorithm shows it's potential over the Apriori algorithm to find the frequent patterns, we have used this FP-growth algorithm along with the support and confidence is being measured using the membership function of the fuzzy logic to find the frequent patterns which are not only using the crisp nature but also uses the fuzziness to measure the support and confidence to infer the association rules which ill dive us direction to discover the diseased symptoms. Therefore, we have hybridized the FP-growth algorithm with fuzzy logic. Our proposed algorithm is given step wise below.

Fuzzy FP-Growth Algorithm

Input: A gene expression dataset consisting of G no. of genes and C no. of conditions or samples, a set of membership functions, and a predefined minimum support threshold $\mathfrak{t}$.
OUTPUT: A Fuzzy FP-growth tree.

STEP 1: Transform the gene expression value g_{ij} of each gene g_j in the i^{th} row into a fuzzy set f_{ij} represented as $(f_{ij1}/R_{j1} + f_{ij2}/R_{j2} + \ldots + f_{ijn}/R_{jhj})$ using the given membership functions, where h_j is the number of fuzzy regions for I_j, Rjl is the l-th fuzzy region of I_j, $1 \le l \le h_j$, and f_{ijl} is v_{ij}'s fuzzy membership value in region R_{jl}.

STEP 2: Calculate the scalar cardinality $count_{jl}$ of each fuzzy region R_{jl} in all the genes as:

$$\text{Count } j_l = \sum_{I=1}^{n} f_{ijl} \qquad \text{Equation (3)}$$

STEP 3: Find max- count_j = Max $(count_{j1})$ for j = 1 to m, where h_j is the number of fuzzy regions for gene g_j and c is the number of conditions. Let max-R_j be the region with max-$count_j$ for gene g_j. It will then be used to represent the fuzzy characteristic of conditions c_j in the later mining process.

STEP 4: Check whether the value max-$count_j$ of a fuzzy region max-Rj, j = 1 to C, is larger than or equal to the predefined minimum count $G^*\mathfrak{t}$. If a fuzzy region max-R_j satisfies the above condition, put the fuzzy region with its count in $L1$. That is:

$$L_1 = \{max\text{-}R_j \mid max\text{-}count_j \ge G^* \mathfrak{t},\ 1 \le j \le C\} \qquad \text{Equation (4)}$$

STEP 5: While executing the above steps, also find the occurrence number $o(max\text{-}R_j)$ of each fuzzy region in the gene expression dataset.

STEP 6: Build the Header Table by keeping the fuzzy regions in L_1 in the descending order of their occurrence numbers.

STEP 7: Remove the fuzzy regions of the items not existing in L_1 from the transactions of the transformed database. Sort the remaining fuzzy regions in each transaction according to the order of the fuzzy regions in the Header Table.

STEP 8: Initially set the root node of the Fuzzy FP-growth tree as *root*.

STEP 9: Insert the genes in the transformed dataset into the Fuzzy FP-growth tree tuple by tuple. The following two cases may exist.

- If a fuzzy region *max*-R_j in the currently processed *i*-th gene has appeared at the corresponding path of the Fuzzy FP-growth tree, add the membership value f_{ijl} of the region *max*-R_j in the transaction to the node with *max*-R_j. Besides, calculate the membership values of the super-gene sets of *max*-R_j in the path by the intersection operator and add the values to the corresponding elements of the array in the node.
- Otherwise, add a new node with *max*-R_j to the end of the corresponding path and set the membership value f_{ijl} of the region *max*-R_j in the currently processed *i*-th gene as the value in the node. Besides, calculate the membership values of the super-gene sets of *max*-R_j in the path by the intersection operator and set the values to the corresponding elements of the array in the node. At last, insert a link from the node of *max*-R_j in the last branch to the current node. If there is no such a branch, insert a link from the entry of *max*-R_j in the Header Table to the current node.

After this step, the final Fuzzy FP-growth tree is built. In this step, a corresponding path is a path in the tree which corresponds to the fuzzy regions to be processed in objects (genes) according to the order of fuzzy regions appearing in the Header Table.

V. Result Analysis

In this paper, a hybridized tree structure called the Fuzzy FP-growth tree has been designed. The tree can keep related mining information such that the database scans can be greatly reduced. A tree construction algorithm is also proposed to build the tree from a gene expression dataset. The construction process is similar to the FP-tree-like processing for building a tight tree structure except each node in the tree has to keep the membership value of the fuzzy region in the node as well as the membership values of its super-gene sets in the path.

We have tested both the approaches in Intel Dual Core machine with 2GB HDD. The OS used is Microsoft XP and all programs are written in MATLAB 8.0. We have observed the similar trends on runtime versus number of frequent patterns found in both Apriori and our proposed Fuzzy FP-growth approaches, the figure 3 shows the running time is significantly less as compared to Apriori approach. We have tested the scalability of both the algorithms on the number of objects as well as number of columns, the figure 4 and figure 5. Our pattern based approach performs substantially better than Apriori based approach.

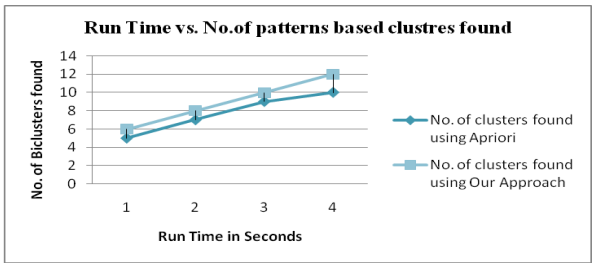


Figure 3: Run time VS No. of frequent pattern based clusters discovered

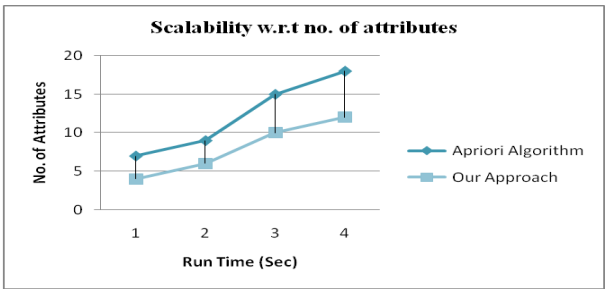


Figure 4: Scalability w.r.t No. of attributes

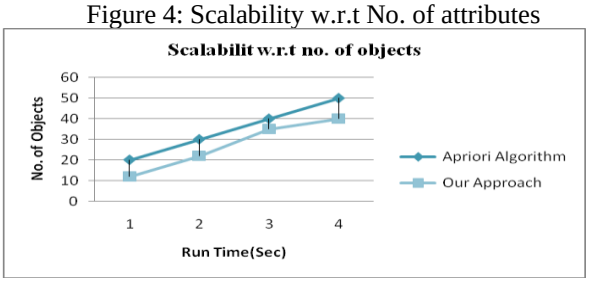


Figure 5: Scalability w.r.t No. of objects

VI. Conclusion

An advantage of FP-Growth is that the constructing FP-Tree has great performance of compression, and its process of mining can reduce the cost of rescanning data. In addition, it applies conditional FP-Tree on avoiding generating candidate item and testing examining process. About its disadvantage, mining process needs extra processing time and space to store that it continuously generates large conditional bases and conditional FP-Trees. In this paper, we combine the concept of fuzzy weight, fuzzy partition methods in data mining, and use FP-Growth to propose FP-Growth tree mining algorithm with all association rules. The heterogeneity and noisy nature of the data along with its high dimensionality decided us to use of a fuzzy association rule mining algorithm for the analysis.

A number of interesting associations have been found in this first approach. A deep study and empirical evaluation of the rules is however needed to confirm such associations. Our future work will be oriented towards modelling this algorithm

to find biclusters from gene expression dataset. Our work can be extended to find the overlapping frequent patterns using Rough Set method and also vertical data format method can be used instead of FP-Tree to find the frequent patterns from gene expression dataset.

REFERENCES

[1] M.Anandhavalli Gauthaman, "Analysis of DNA Microarray Data using Association Rules: A selective study ", *Ward Academy of Science, Engineering & Technology*, Volume-42, 2008, pp.12 to 15.

[2] Sara C. Madeira and Arlindo L. Oliveira, "Biclustering Algorithms for Biological Data Analysis: A Survey ", *IEEE Transaction on Computational on Biology & Bioinformatics*, Volume-1.No.1 January – March 2004, PP.24 – 25.

[3] C. Gyorodi, R. Gyorodi "Minimum association rule in large Database*", Proc. of Oradea EMES'02':45/50*, Oradea, Romania, 2002, pp.12 – 16.

[4] Y.chengand G.M church,"Biclustering of Gene Expression data",*Proc.Eight'l Conf.Intelligent* systems for molecular Biology(ISMB '00),pp.93-103,2000.

[5] Wim pijis and Walter A. Kostens+,"How to find frequent patterns",*Econometric Institute Report EI* 2005-24 june 1,2005

[6] Bart Gothals,"survey on frequent pattern mining",*HIIT basic Research unit Dept of Computerscience University of Helsinki p.o.box* 26,FIN-00014 *Helsinki finland.*

[7] Miao Wang, Xuequnshag, Zhanhuki, Li: "Research on microarray dataset mining", *VLDB2010 ph workshop*, September 13, 2010, Singapore

[8] M. Anandhavalli ,M K Ghose, K Gautaman," Association Rule mining in Genomics", Member, *IACSIT, IAENG International Journal of computer theory and Engineering*

[9] J.han, J.Pai, Y. Yin,"Mining Frequent patterns without candidate generation", *Proc .of Oradea EMES*'02 :45-50. *Oradea, Romania*,2002

[10] Pijls W., Bioch J.C. (1999)," Mining Frequent Itemsets in Memory-Resident Databases" *In: Postma E., Gyssens M. (Eds.) Proceedings of the Eleventh Belgium-Netherlands***,** *Conference on Artificial Intelligence* (BNAIC1999), pages 75–82.

[11] Jiwei Han, Jian Pei, Yiwen Yen and Runying Mao: "Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach, Data Mining and Knowledge Discovery, 8, 53–87, 2004, 2004 Kluwer Academic Publishers

[12] Liqiang Geng and Howard J. Hamilton. Interestingness measures for data mining: A survey. *ACM Computing Surveys*, 38(3):9, (2006)

[13] F. Javier Lopez, Marta Cuadros, Armando Blanco and Angel Concha:"Unveiling Fuzzy Associations Between Breast Cancer Prognostic Factors and Gene Expression Data", 20th International Workshop on Database and Expert Systems Application, pp.338-342,(2009).

[14] D. Dubois, E. H¨ullermeier, and H. Prade, "A systematic approach to the assessment of fuzzy association rules," D*ata Mining and Knowledge Discovery*, Vol. 13, pp. 167-192, 2006.

[15] CHRISTIAN BORGELT, AN IMPLEMENTATION OF THE FP-GROWTH ALGORITHM, OSDM '05 PROCEEDINGS OF THE 1ST INTERNATIONAL WORKSHOP ON OPEN SOURCE DATA MINING: FREQUENT PATTERN MINING IMPLEMENTATIONS, (2005).

[16] R. Das,D.K. Bhattacharyya and J.K. Kalita, Clustering Gene Expression Data Using An Effective Dissimilarity Measure, *International Journal of Computational Bioscience*, Vol. 1, No. 1, 2010,pp 55-68,(2010)

AUTHORS PROFILE

Shruti Mishra is a scholar of M.Tech(CSE) at Kaustuv Institute for Self Domain, Biju Pattanaik University, Bhubaneswar, Odisha, INDIA. Her research includes Data mining, Soft Computing Techniques etc.

Debahuti Mishra is an Assistant Professor and research scholar in the department of Computer Sc. & Engg, Institute of Technical Education & Research (ITER) under Siksha O Anusandhan University, Bhubaneswar, Odisha, INDIA. She received her Masters degree from KIIT University, Bhubaneswar. Her research areas include Datamining, Bio-informatics, Software Engineering, Soft computing. Many publications are there to her credit in many International and National level journal and proceedings.She is member of OITS, IAENG and AICSIT. She is an author of a book Aotumata Theory and Computation by Sun India Publication (2008).

Shruti Mishra is a scholar of M.Tech(CSE) Institute of Technical Education & Research (ITER) under Siksha O Anusandhan University, Bhubaneswar, Odisha, INDIA. Her research areas include Data mining, Parallel Algorithms etc.

Sandeep Kumar Satapathy is Asst.Prof. in the department of Computer Sc. & Engg, Institute of Technical Education & Research (ITER) under Siksha `O` Anusandhan University, Bhubaneswar . He has received his Masters degree from Siksha O Anusandhan University, Bhubaneswar, Odisha, INDIA. His research areas include Web Mining, Data mining etc.

Dr.Amiya Kumar Rath obtained Ph.D in Computer Science in the year 2005 from Utkal University for the work in the field of Embedded system. Presently working with College of Engineering Bhubaneswar (CEB) as Professor of Computer Science & Engg. Cum Director (A&R) and is actively engaged in conducting Academic, Research and development programs in the field of Computer Science and IT Engg. Contributed more than 30 research level papers to many national and International journals. and conferences Besides this, published 4 books by reputed publishers. Having research interests include Embedded System, Adhoc Network, Sensor Network ,Power Minimization, Biclustering, Evolutionary Computation and Data Mining.

Dr. Milu Acharya obtained her Ph.D at Utkal University. She is a Professor in Department of Mathematics at Institute of Technical Education and Research (ITER), Bhubaneswar. She has contributed more than 20 research level papers to many national and International journals and conferences Besides this, published 3 books by reputed publishers. Her research interests include Biclustering, Data Mining , Evaluation of Integrals of analytic Functions , Numerical Analysis , Complex Analysis , Simulation and Decision Theory.

Identification and Evaluation of Functional Dependency Analysis using Rough sets for Knowledge Discovery

Y.V.Sreevani[1], Prof. T. Venkat Narayana Rao[2]
Department of Computer Science and Engineering,
Hyderabad Institute of Technology and Management,
Hyderabad, A P, INDIA
s_vanikumar@yahoo.co.in[1], tvnrbobby@yahoo.com[2]

***Abstract*— The process of data acquisition gained momentum due to the efficient representation of storage/retrieving systems. Due to the commercial and application value of these stored data, Database Management has become essential for the reasons like consistency and atomicity in giving birth to DBMS. The existing database management systems cannot provide the needed information when the data is not consistent. So knowledge discovery in databases and data mining has become popular for the above reasons. The non-trivial future expansion process can be classified as Knowledge Discovery. Knowledge Discovery process can be attempted by clustering tools. One of the upcoming tools for knowledge representation and knowledge acquisition process is based on the concept of Rough Sets. This paper explores inconsistencies in the existing databases by finding the functional dependencies extracting the required information or knowledge based on rough sets. It also discusses attribute reduction through core and reducts which helps in avoiding superfluous data. Here a method is suggested to solve this problem of data inconsistency based medical domain with a analysis.**

Keywords- Roughset;knowledge base; data mining; functional dependency; core knowledge.

I. INTRODUCTION

The process of acquiring features hidden in the data is the major objective of Data Mining. Organizing these features for utilizing in planning for better customer satisfaction and promoting the business is the focus of Knowledge Representation. For discovering knowledge in data bases [6][7], in other words, reverse engineering has been attempted using the concept of Rough Sets for finding functional dependencies. Different phases of knowledge discovery process can be used for attribute selection, attribute extraction, data reduction, decision rule generation and pattern extraction. The fundamental concepts have been explored here for getting the core knowledge in Binary Data bases. Rough Sets are applied not only for Knowledge Representation but they are also being applied to Pattern Classification, Decision Making, Switching Circuits, and Data Compression etc[1][9]. It is proposed to find out the degree of dependency using Rough Sets introduced by Pawlak [1]which is used for characterizing the given instance, extracting information. This helps to know all Functional Dependencies existing in the vast databases. Rough set theory provides a collection of methods for extracting previously unknown data dependencies or rules from relational databases or decision tables. Rough set approach does not need any preliminary or additional information about data like probability in statistics, grade of membership in the fuzzy set theory. It proves to be efficient because it has got tools and algorithms which are sufficient for finding hidden patterns in data. It allows in reducing original data, i.e. to find minimal sets of data with the same knowledge as in the original data. The first, pioneering paper on rough sets, written by Zdzisław Pawlak, was published by International Journal of Computer and Information Sciences in 1982.

II. THE ROUGHSETS REPRESENTATIONS

The roughest method is basically associated with the classification and analysis of imprecise, uncertain or incomplete information or knowledge expressed in terms of data acquired from the experience. The domain is a finite set of objects. The domain of interest can be classified into two disjoint sets. The classification is used to represent our knowledge about the domain, i.e. the knowledge is understood here as an ability to characterize all classes of the classification, for example, in terms of features of objects belonging to the domain. Objects belonging to the same category are not distinguishable, which means that their membership status with respect to an arbitrary subset of the domain may not always be clearly definable. This fact leads to the definition of a set in terms of lower and upper approximations. The lower approximation is a description of the domain objects which are known with full certainty which undoubtedly belongs to the subset of interest, whereas the upper approximation is a description of the objects which would possibly belong to the subset. Any subset defined through its lower and upper approximations is called a rough set. The idea of rough set was proposed by Pawlak (1982) as a new mathematical tool to deal with vague concepts. Comer, Grzymala-Busse, Iwinski, Nieminen, Novotny, Pawlak, Obtulowicz, and Pomykala have studied algebraic properties of rough sets. Different algebraic semantics have been developed by P. Pagliani, I. Duntsch, M. K. Chakraborty, M. Banerjee and A. Mani; these have been extended to more generalize rough sets by D. Cattaneo and A. Mani, in

particular. Rough sets can be used to represent ambiguity, vagueness and general uncertainty.

A. Knowledge Base

Let us consider a finite set U≠Ø (the universe) of objects under question, and R is a family of equivalence relations over U. Any subset Q⊆U of the universe will be called a concept or a category in U and any family of concepts in U will be referred to as abstract knowledge (or in short knowledge) about U. A family of classifications over U will be called a knowledge base K over U. To this end we can understand knowledge base as a relational system K = (U, R), where U≠Ø is a finite set called the universe, and R is a family of equivalence relations over U. If E is an equivalence relation over U, then by E/R we mean the family of all equivalence classes of R (or classification of U) referred to as categories or concepts of R and $[Q]_R$ denotes a category in R containing an element q ∈ U. and If P ⊆ R and P ≠ Ø, then ∩ P (intersection of all equivalence relations belonging to P) is also an equivalence relation, and will be denoted by IND(P), and will be called an indiscernibility relation over P.

Therefore $[Q]_{IND(P)} = \cap [Q]_R$.

B. The Concept of Rough Sets

Let there be a relational system K = (U, **R**), where **U**≠Ø is a finite set called the universe, and **R** is a family of equivalence relations over **U**. Let Q⊆**U** and R be an equivalence relation [1]. We will say that Q is R-definable [12][1], if Q can be expressed as the union of some R-basic categories, otherwise Q is R-undefinable. The R-definable sets are called as R-exact sets some categories (subsets of objects) cannot be expressed exactly by employing available knowledge. Hence we arrive at the idea of approximation of a set by other sets. Let Q⊆**U** and equivalence relation R∈ IND(K) we associate two subsets i.e. $R_{LQ} = U\{Y \in \mathbf{U}/R : Y \subseteq Q\}$ and

$R^{UQ} = U\{Y \in \mathbf{U}/R : Y \cap Q \neq Ø\}$ called the R^{UQ} –UPPER and R_{LQ} -LOWER approximation of Q respectively[1][4].

From the above we shall also get the following denotations i.e.

$POS_R(Q) = R_{LQ}$, R-positive region of Q.

$NEG_R(Q) = \mathbf{U} - R^{UQ}$, R-negative region of Q.

$BN_R(Q) = R_{LQ} - R^{UQ}$,R-borderline region of Q.

The positive region $POS_R(Q)$ or the lower approximation of Q is the collection of those objects which can be classified with full certainty as members of the set Q, using Knowledge R.

In addition to the above we can define following terms-R-positive region of Q,$POS_R(Q) = R_{LQ}$.

Let X ⊆**U** , Where X is a subset of objects chosen from U and P and Q be the equivalence relations over U, then R-positive region of Q is $POS_P(Q) = \cup x \in u/_Q \; P_{LX}$

The P-positive region of Q is the set of all objects of the universe U which can be properly classified to classes of **U**/Q employing knowledge expressed by the classification **U**/P.

In the discovery of knowledge from huge databases we have to find the degree of dependency. This is used for characterizing the given instance, extracting information and which helps to know all the functional dependencies. Intuitively, a set of attributes Q depends totally on a set of attributes P, denoted P →Q, if the values of attributes from P uniquely determine the values of attributes from Q. In other words, Q depends totally on P, if there exists a functional dependency between values of P and Q. $POS_P(Q) = \cup x \in u/_Q \; P_{LX}$ called a positive region of the partition U/Q with respect to P, is the set of all elements of U that can be uniquely classified to blocks of the partition U/Q, by means of P. The degree of dependency between P and Q where P,Q ⊂ **R** is defined as follows.

If P and Q be a set of the equivalence relations over U,

Then the set of attributes of Q depends in a degree k ($0 \leq k \leq 1$), from P denoted by P → Q ,

If $k = \gamma_P(Q) = \text{Card } POS_P(Q) / \text{Card } \mathbf{U}$.

Where card denotes cardinality of the Set and the symbol γ is used to specify POS that is positive region.

If k=1, we will say that Q totally depends from P.

If O<k<1, we say that Q partially depends from P.

If k=0, we say that Q is totally independent from P.

If k = 1 we say that Q depends totally on P, and if k < 1, we say that Q depends partially (to degree k) on P.

If k = 0 then the positive region of the partition U/Q with respect to P is empty.

The coefficient k expresses the ratio of all elements of the universe, which can be properly classified to blocks of the partition U/Q, employing attributes P and will be called the degree of the dependency. Q is totally (partially) dependent on P, if all (some) elements of the universe U can be uniquely classified to blocks of the partition U/Q, employing P. If the positive region is more then ,there exists a larger dependency between P and Q. This can be used to find the dependency between attribute sets in databases.

The above described ideas can also be interpreted as an ability to classify objects. more clearly, if k=1, then all elements of the knowledge base can be classified to elementary categories of U/Q by using knowledge P. If k≠1, only those elements of the universe which belong to the positive region can be classified to categories of knowledge Q, employing knowledge P. In particular if k=0, none of the elements of the universe can be classified using P and to elementary categories of knowledge Q.

More presicely, from the definition of dependency follows, that if, then the positive region of partition U/Q induced by Q

covers k*100 percent of all objects in the knowledge base. On the other hand, only those objects belonging to positive region of the partition can be uniquely classified. This means that k*100 percent of objects can be classified into block of partition U/Q employing P. If we restrict the set of objects in the knowledge base $POS_P(Q)$,we would obtain the knowledge base in which $P \rightarrow Q$ is a total dependency.

C. Indiscernibility

The notion of indiscernibility is fundamental to rough set theory. Informally, two objects are indiscernible if one object cannot be distinguished from the other on the basis of a given set of attributes. Hence, indiscernibility is a function of the set of attributes under consideration. An indiscernibility relation partitions the set of facts related to a set of objects into a number of equivalence classes . An equivalence class of a particular object is simply the collection of those objects that are indiscernible to the object in question[8] [13]. It is often possible that some of the attributes or some of the attribute values are superfluous. This enables us to discard functionally redundant information. A reduct is defined as a minimal set of attributes that preserves the meaning of indiscernibility relation [9][10] computed on the basis of the full set of attributes. Preserving the indiscernibility preserves the equivalence classes and hence it provide us the ability to form approximations. In practical terms, reducts help us to construct smaller and simpler models, and provide us an idea on the decision-making process [6],[7]. Typically, a decision table may have many reducts. However, there are extended theories to rough sets where some of the requirements are lifted. Such extensions can handle missing values and deal with hierarchies among attribute values. . In the following, for the sake of simplicity, it will be assumed that none of the attribute values are missing in data table so as to make it easy to find the dependencies.

D. Reduct and core Pertaining to Condition Attributes

Reduct and core of condition attributes helps in removing of superfluous partitions (equivalence relations) or/and superfluous basic categories in the knowledge base in such a way that the set of elementary categories in the knowledge base is preserved. this procedure enables us to eliminate all unnecessary knowledge from the knowledge base and preserving only that part of the knowledge which is really useful[13][14].

This concept can be formulated by the following example as follows.

Let $F=\{X_1 \ldots X_N\}$ is a family of sets choosen from **U** such that $X_i \subseteq$ **U.**

We say that X_i is dispensable in F,if $\cap(F-\{X_i\}) = \cap F$.

The family F is independent if all of its components are indispensible in F; otherwise F is dependent.

The family $H \subseteq F$ is a reduct of F, if H is independent and $\cap H = \cap F$.

The family of all indispensable sets in F will be called as the core of F, denoted by CORE(F).

From the above theory available in Rough Sets proposed by Pawlak.Z(1995) [1] the following definition can be derived where CORE(F)= ∩ RED(F) and RED(F) is the family of all reducts of F.

For example Consider family **R** = {P,Q,R} of equivalence relations having the following equivalence classes :
U/P = {{x_1,x_3 , x_4 , x_5, x_6, x_7 }, {x_2, x_8}}
U/Q = {{x_1,x_3 , x_4 , x_5}, {x_2,x_6, x_7,x_8}}
U/R = {{x_1, x_5, x_6}, {x_2, x_7, x_8}, {x_3, x_4}}
The family **R** induces classification
U/IND (**R**) = {{x_1, x_5} {x_3, x_4}, {x_2, x_8} {x_6}, {x_7}}
Moreover, assume that the equivalence relation S is given with the equivalence classes U/S = {{x_1, x_5, x_6}, {x_3, x_4}, {x_2, x_7}, {x_8}} . The positive region of S with respect to R is the union of all equivalence classes of U/IND(**R**) which are included in some equivalence classes of U/S, i.e. the set $POS_{\mathbf{R}}(S)$ = {x_1,x_3 , x_4 , x_5, x_6, x_7}.
In order to compute the core and reducts of **R** with respect to S, we have first to find out whether the family **R** is S-dependent or not. According to definitions given in this section, we have to compute first whether P, Q and R are dispensable or not with respect to S (S-dispensable).
Removing P we get U/IND(R-{P}) = {{x1, x 5} {x3 , x4}, {x 2, x7, x 8} {x 6}}
Because, $POS_{(R-\{P\})}(S)$ = {x1, x3, x4, x5,x 6}≠$POS_R(S)$
the P is S-indispensable in **R.**

Dropping Q from **R** we get U/IND(**R**-{Q}) = {{x_1, x_5, x_6 }, {x_3, x_4}, {x_2, x_8}, {x_7}} , which yields the positive region $POS_{(\mathbf{R}-\{Q\})}(S)$ = {x_1,x_3 , x_4 , x_5, x_6, x_7 }= $POS_R(S)$
Hence Q is S-dispensable in **R**. Finally omitting R in **R** we obtain U/IND(**R**-{R}) = {{x_1, x_3, x_4, x_5}, {x_2, x_8}, {x_6, x_7}} and the positive region is : $POS_{(\mathbf{R}-\{R\})}(S) = \varnothing \neq POS_{\mathbf{R}}(S)$, which means that R is S-indispensable in **R**.
Thus the S-core of **R** is the set {P,R}, which is also the S-reduct of **R.**

III. PROPOSED METHOD

To find the dependency between any subset of attributes using rough sets we are using a decision table based on certain factors and circumstances related to the knowledge base or the domain we choose. Due to the inconsistent nature of the data [11], certain data values in the data table may be conflicting. Here, a method is suggested to solve this problem of data inconsistency based on the approach inspired by rough set theory by Pawlak.Z.(1995)[1]. Generate the powerset of condition attributes for each element in the powerset :

- Find the equivalence classes.
- Associate a decision attribute.
- Find the degree of dependency.

- Find the inconsistent objects where the attribute values of the decision attributes are different, even though the attribute values of condition attributes are same.
- Calculate the degree of dependency k. Display those objects whose degree of dependency lies between 0 and 1.Display the inconsistent objects set.
- End for
- End.

IV. CREATION OF DECISION TABLE FOR KNOWLEDGE DISCOVERY

Let there be a set X of interest and is unknown and we have only some information about it. Assuming some sets which are disjoint with X and some sets included in X so as to build good approximations to X and use them to reason out on X. In this paper we are considering an example of a group of individuals (Table 1) who are at a risk of influenza (Zdzislaw Pawlak,1995)[1].

TABLE I. Patient information table

	temp	cough	head_ ache	muscle _pain	influenza
p_1	normal	present	present	present	present
p_2	normal	present	absent	absent	present
p_3	medium	present	absent	absent	absent
p_4	medium	absent	present	present	present
p_5	medium	absent	present	present	absent
p_6	high	present	present	present	present
p_7	high	absent	present	present	present
p_8	high	absent	present	present	absent
p_9	high	absent	absent	absent	absent

F_1----temp (normal, 0) (medium, 1) (high, 2)
F_2---- cough (present, 1) (absent, 2)
F_3----head_ ache (present, 1)(absent, 2)
F_4-----muscle_ pain (present, 1) (absent, 2)
F_5----influenza (present, 1) (absent, 2)

V. DECISION RULES

A decision rule [1],[5] is defined to be a logical expression in the form .IF (condition …) then (decision…), where in the condition is a set of elementary conditions connected by "and" and the decision is a set of possible outcomes/actions connected by "or". The above mentioned decision rule can be interpreted within the rough set framework and the If- then-part of the rule lists more than one possible outcome, that can be interpreted as describing one or more cases[8] . The If-then-part of the rule lists a single action Yes (or No.), that can be interpreted for describing one or more cases that lie in either the inside (or the outside) region of the approximation [14]. A set of decision rules forms the decision algorithm. Based on the above theory we consider the physical conditions related to nine patients (Table 1 and Table II) and their corresponding characteristic attribute values are used to derive the following rules which in turn help in building the decision table more rational.

Rule 1: if (temp=normal and cough=present and head_ache=present and muscle_pain=present) then (Influenza=present) .
Rule 2:if (temp=normal and cough=present and head_ache=absent and muscle_pain=absent) then (Influenza=present) .
Rule 3:if(temp=medium and cough=present and head_ache=absent and muscle_pain=absent) then (Influenza=absent) .
Rule 4:if (temp=medium and cough=absent and head_ache=present and muscle_pain=present) then (Influenza=present) .
Rule 5:if (temp=medium and cough=absent and head_ache=present and muscle_pain=present) then (Influenza=absent) .
Rule 6:if (temp=highand cough=present and head_ache=present and muscle_pain=present) then (Influenza=present) .
Rule 7: if(temp=high and cough=absent and head_ache= present and muscle_pain= present) then (Influenza=present).
Rule 8: if(temp= high and cough= absent and head_ache= present and muscle_pain= present) then (Influenza= absent) .
Rule 9:if(temp= high and cough= absent and head_ache= absent and muscle_pain= absent) then (Influenza= absent).

Using the above rules we can construct a decision table (Table I.) for nine different patients having different characteristics who are at a risk of influenza. The columns are labeled by factors or circumstances that reflect the physical condition of the patient in terms of set of condition attributes and decision attributes. The rows are labeled by objects where in each row represents a piece of information about the corresponding to each patient. Once the relation/table is created it is possible to find all the functional dependencies (Table III.) which would be useful for decision support systems as well as knowledge building/rule generation.

TABLE II. Decision Table

U	*Condition attributes*				*Decision attribute*
	F_1	F_2	F_3	F_4	F_5
p_1	0	1	1	1	1
p_2	0	1	2	2	1
p_3	1	1	2	2	2
p_4	1	2	1	1	1
p_5	1	2	1	1	2
p_6	2	1	1	1	1
p_7	2	2	1	1	1
p_8	2	2	1	1	2
p_9	2	2	2	2	2

The power set generated for the above condition attributes are:
$\{F_1\}_,\{F_2\}_,\{F_3\}_,\{F_4\}$
$\{F_1,F_2\},\{F_1,F_3\},\{F_1,F_4\},\{F_2,F_3\},\{F_2,F_4\},\{F_3,F_4\}$
$\{F_1,F_2,F_3\},\{F_1,F_3,F_4\},\{F_2,F_3,F_4\}$,$\{F_1,F_2,F_3,F_4\}$.

Using the power set we can generate various attribute sets for which functional dependencies are to be identified i.e. from the above table.

The equivalence classes for each element for the powerset are generated as below.

$U/F_{1=}$ {{P_1,P_2},{P_3,P_4,P_5},{P_6,P_7,P_8,P_9}}

U/F_2={{ P_1,P_2, P_3,P_6},{ P_4,P_5,P_7,P_8,P_9}}

U/F_3= {{P_1, P_4,P_5, P_6 ,P_7,P_8}, { P_2, P_3,P_9}}

$U/F_{4=}$ {{ P_1, P_4,P_5, P_6 ,P_7,P_8}, { P_2, P_3,P_9}}etc.

The equivalence classes for decision attribute are:

$U/F_{5=}$ {{P_1,P_2, P_4, P_6 ,P_7}, { P_3, P_5,P_8,P_9}}.

By applying the theory of rough sets , The equivalence classes generated between every element of powerset and the decision attribute F_5 are :

$F_1 \rightarrow F_5$

F_1= {{P_1, P_2}, {P_3,P_4,P_5}, {P_6,P_7,P_8,P_9}}

F_5= {{P_1 ,P_2, P_4, P_6, P_7}, {P_3,P_5,P_8,P_9}}.

$POS_{F1}(F5) = \{P_1, P_2\}$, $k = \gamma_{F1}(F5) = 2/9$.

$F_2 \rightarrow F_5$

F_2={{ P_1,P_2, P_3,P_6}, { P_4,P_5,P_7,P_8,P_9}}.

F_5= {{P_1, P_2, P_4, P_6 , P_7}, { P_3, P_5,P_8,P_9}}.

$POS_{F2}(F5) = \{0\}$, $k = \gamma_{F2}(F5) = 0/9$.

$F_3 \rightarrow F_5$

F_3= {{ P_1, P_4, P_5, P_6 , P_7, P_8}, { P_2, P_3, P_9}}

F_5= {{P_1, P_2, P_4, P_6, P_7}, { P_3, P_5, P_8, P_9}}.

$POS_{F3}(F5) = \{0\}$, $k = \gamma_{F3}(F5) = 0/9$.

$F_4 \rightarrow F_5$, F_4= { P_1, P_4,P_5, P_6 ,P_7,P_8}, { P_2, P_3,P_9}

F_5= {P_1,P_2, P_4, P_6 ,P_7}, { P_3, P_5,P_8,P_9} , $POS_{F4}(F5) = \{0\}$.

$k = \gamma_{F4}(F5) = 0/9$.

$F_1F_2 \rightarrow$ F5

F_1F_2= {{P_1,P_2}, {P_3}, {P_4,P_5}, {P_6}, {P_7,P_8,P_9}}

F_5= {{P_1,P_2, P_4, P_6 ,P_7}, { P_3, P_5,P_8,P_9}}.

$POS_{F1F2}(F5) = {}=\{P_1, P_2, P_6, P_3\}$. $k = \gamma_{F1F2}(F5) = 4/9$.

$F_1F_3 \rightarrow$ F5 ,

F_1F= {{P_1}, {P_2}, {P_3}, {P_4,P_5}, {P_6,P_7,P_8}, {P_9}}

F_5= {P_1, P_2, P_4, P_6 ,P_7}, { P_3, P_5,P_8,P_9}}.

$POS_{F1F3}(F5) = \{P_1, P_2, P_3, P_9\}$. $k = \gamma_{F1F3}(F5) = 4/9$.

$F_1F_4 \rightarrow$ F5

F_1F_4 = {{P_1}, {P_2},{P_3}, {P_4,P_5}, {P_6,P_7,P_8}, {P_9}}

F_5= {{P_1,P_2, P_4, P_6 ,P_7}, { P_3, P_5,P_8,P_9}}.

$POS_{F1F4}(F5) = \{P_1, P_2, P_3, P_9\}$. $k = \gamma_{F1F4}(F5) = 4/9$.

$F_2F_3 \rightarrow$ F5

F_2F_3= {{P_1, P_6}, { P_2, P_3}, {P_4,P_5,P_7,P_8}, {P_9}}

F_5= {{P_1, P_2, P_4, P_6 , P_7}, { P_3, P_5,P_8,P_9}}.

$POS_{F2F3}(F5) = \{P_1, P_6, P_9\}$. $k = \gamma_{F2F3}(F5) = 3/9$.

$F_2F_4 \rightarrow$ F5

F_2F_4= {{P_1,P_6}, { P_2, P_3}, {P_4,P_5,P_7,P_8}, {P_9}}.

F_5= {{P_1,P_2, P_4, P_6 ,P_7}, { P_3, P_5,P_8,P_9}}.

$POS_{F2F4}(F5) = \{P_1, P_6, P_9\}$. $k = \gamma_{F2F4}(F5) = 3/9$.

$F_3F_4 \rightarrow$ F5

F_3F_4= {{P1,P_4,P_5,P_6,P_7,P8}, {P_2,P_3,P_9}}.

F_5= {{P_1,P_2, P_4, P_6 ,P_7}, { P_3, P_5,P_8,P_9}}.

$POS_{F3F4}(F5) = \{0\}$.

$k = \gamma_{F2F4}(F5) = 0/9$.

$F_1F_2F_3 \rightarrow F_5$

$F_1F_2F_3$= {{P_1}, {P_2},{ P_3}, {P_4,P_5}, {P_6}, {P_7,P_8}, {P_9}}.

F_5= {{P_1, P_2, P_4, P_6 , P_7}, { P_3, P_5,P_8,P_9}}.

$POS_{F1F2F3}(F5) = \{P_1, P_2, P_6, P_3, P_9\}$. $k = \gamma_{F1F2F3}(F5) = 5/9$.

$F_2F_3F_4 \rightarrow F_5$

$F_2F_3F_4$= {{P_1, P_6}, {P_2, P_3}, {P_4, P_5, P_7, P_8}, {P_9}}.

F_5= {{P_1, P_2, P_4, P_6, P_7}, { P_3, P_5, P_8, P_9}}.

$POS_{F2F3F4}(F5) = \{P_1, P_6, P_9\}$. $k = \gamma_{F2F3F4}(F5) = 3/9$.

$F_1F_2F_4 \rightarrow F_5$

$F_1F_2F_4$= {{P_1}, {P_2}, P_3}, {P_4,P_5}, {P_6}, {P_7,P_8}, {P_9}}.

F_5= {{P_1, P_2, P_4, P_6 ,P_7}, { P_3, P_5, P_8, P_9}}.

$POS_{F1F2F4}(F5) = \{P_1, P_2, P_6, P_3, P_9\}$. $k = \gamma_{F1F2F4}(F5) = 5/9$.

$F_1F_3F_4 \rightarrow F_5$

$F_1F_3F_4$= {{P_1},{P_2},{ P_3}, {P_4,P_5},{P_6,P_7,P_8},{P_9}}

F_5= {{P_1,P_2, P_4, P_6 ,P_7}, { P_3, P_5, P_8, P_9}}.

$POS_{F1F3F4}(F5) = \{P_1, P_2, P_3, P_9\}$. $k = \gamma_{F1F3F4}(F5) = 4/9$.

$F_1F_2F_3F_4 \rightarrow F_5$

$F_1F_2F_3F_4$= {{p_1}, {p_2}, {p_3}, {p_4,p_5}, {p_6},{ p_7,p_8}, {p_9}}.

F_5= {{P_1,P_2,P_4,P_6,P_7}, {P_3,P_5,P_8,P_9}}.

$POS_{F1F2F3F4}(F5) = \{P_1, P_2, P_3, P_6, P_9\}$. $k = \gamma_{F1F2F3F4}(F5) = 5/9$.

TABLE III. Dependency table

Power setelements(ps)	$k = \gamma_{ps}(F_5)$
F_1	2/9
F_2	0/9
F_3	0/9
F_4	0/9
F_1F_2	4/9
F_1F_3	4/9
F_1F_4	4/9
F_2F_3	3/9
F_2F_4	3/9
F_3F_4	0/9
$F_1F_2F_3$	5/9
$F_1F_3F_4$	4/9
$F_2F_3F_4$	3/9
$F_1F_2F_4$	5/9
$F_1F_2F_3F_4$	5/9

VI. ANALYSIS BASED ON REDUCT AND CORE

By the above procedure we can extract Core of condition attributes which are explicitly necessary for deriving knowledge and coming to some conclusions related to the extraction of knowledge[2],[4]. We need to pursue a method which would give information of whether a particular characteristic attribute is necessary or not, based on which it can be established whether a patient has influenza or not. Analysis over the decision table is performed in this paper by identifying those core attributes whose removal would results in further inconsistency in the decision table which was consistent other wise. In the above decision table [2][3] (Table II.) by dropping F_1 rules P_2 and P_3 turns out to be inconsistent and positive region of the algorithm changes. Therefore, F_1 forms the core of the attribute set in the decision table. Similarly by dropping F_2 results in making P_6 and P_8 inconsistent and thus change in positive region of the algorithm[12]. The above procedure is repeatedly applied and

checked for the inconsistency in the decision table and also to extract the core knowledge from the input domain.

U	*Condition attributes*			*Decision attribute*	
	F_1	F_2	F_3	F_4	F_5
p_1	0	1	1	1	1
p_2	0	1	2	2	1
p_3	1	1	2	2	2
p_4	1	2	1	1	1
p_5	1	2	1	1	2
p_6	2	1	1	1	1
p_7	2	2	1	1	1
p_8	2	2	1	1	2
p_9	2	2	2	2	2

Figure 1. Attribute reduction

When we remove attribute F_1 rules 2 and 3 gets violated and the data corresponding to objects P_2 and P_3 turn into inconsistent as shown in Figure 1. But the removal of condition attribute F_4 still preserves the consistency of data and does not form the core of the condition attributes. The basic idea behind extraction of core knowledge is to retrieve knowledge of a characteristic attribute by observing its behavior, and this behavior is used to generate the algorithm and can be further used to simulate the actions in the future [4].

To find the reducts drop, take attributes as they appear in the power set and check whether any superfluous partitions (equivalence relations) or/and superfluous basic categories in the knowledge base [13] that are existing so that the set of elementary categories in the knowledge base is preserved. This procedure enables us to eliminate all unnecessary knowledge from the knowledge base, preserving only that part of the knowledge which is really useful.

For example drop F_4 and F_3 from F_1,F_2,F_3,F_4 and check the changes in the positive region.

This is done as follows

$$\text{Card}(\text{ Pos }_{\{F1,F2,F3-\{F4\}\}})(F_5)= \{P_1,P_2,P_6,P_3,P_9\}=5$$
$$k= \gamma_P\ (Q) = 5/9$$
$$\text{Card}(\text{ Pos }_{\{F1,F2,F4\}-\{F3\}}\ (F_5)= \{P_1,P_2,P_6,P_3,P_9\}=5$$
$$k= \gamma_P\ (Q) = 5/9.$$
$$\text{POS }_{\{\{F1,F2,F3\ -\ \{F4\}\}}(F_5) = \text{POS }_{\{\{F1,F2,F4\}-\{F3\}\ \}}(F_5)= \text{POS }_{F1,F2,F3,F4}\ (F_5)$$

From above we can notice that even with the removal of attributes F_4 and F_3 there is no change in the positive region.Therefore the reducts are {F1, F2, F3} and {F1, F2,F4} and the core attributes are { F1,F2}.

VII. CONCLUSION

Rough set theory provides a collection of methods for extracting previously unknown data dependencies or rules from relational databases or decision tables. As established above it can be said that roughsets relates to entities databases, data mining, machine learning, and approximate reasoning etc. This paper enables us to examine and to eliminate all unnecessary knowledge from the knowledge base by preserving only that part of the knowledge which is really useful. This paper gives some insight into roughsets which can be used to know data dependencies and extraction of knowledge. The ideas envisaged and depicted here are useful in the domain which deal huge collection of databases to analysis and take rational decisions in the areas such as banking, stock markets, medical diagnosis etc.

REFERENCES

[1] Zdzislaw Pawlak (1995)"Rough Sets: Theoretical Aspects Of Reasoning about Data", Institute of computer science, Noowowlejska 15/19, 00-665 Warsaw, Poland.

[2] H.Sug,"Applying rough sets to maintain data consistency for high degree relations", NCM'2008, Vol.22, 2008, pp.244-247.

[3] Duntsch, and G.Gediga,"Algebraic aspects of attribute dependencies in information systems", Fundamenta Informaticae, Vol.29,1997,pp.119-133.

[4] I. I.Duntsch, and G.Gediga,"Statistical evalution of rough set dependency analysis,"Internation journal of human computer studies,Vol.46,1997.

[5] T.Y.Lin, and H.Cao,"Seraching decision rules in very large databases using rough set theory,"Lecture notes in artificial intelligence Ziarco and Yao eds., 2000,pp.346-353.

[6] Arun K.Pujari,"Data Mining Techniques" Universities Press (India) Limited.

[7] Silbreschatz, Korth & Sudarshan (1997) "Database System Concepts"3/c McGraw Hill Companies, Inc.

[8] Dr.C.Raghavendra Rao, "Functional Dependencies and their role on Order Optimization.

[9] R.Stowinski, ed, Intelligent decision support: Handbook of Applications and advances of the rough set theory, Kulwer Academic publishers, 1992.

[10] A.Ohrn,Discernibility and rough sets in medicine: tools and Applications phd thesis,Department of computer and information science, Norwegian University of Science & Technology, 1999.

[11] J.G.Bazan, M.S.Szczuka, and j.Wroblewski,"A new Version of rough set exploration system,"Lecture notes in artificial intelligence,Vol. 2475, 2002,pp.397_404.

[12] N.Ttow, D.R.Morse, and D.M.Roberts,"Rough set approximation as formal concept,"journal of advanced computational intelligent Informatics, Vol.10, No.5, 2006, pp.606-611.

[13] Merzana kryszkiewicz,Piotr lasek**"**Fast Discovery of Minimal Sets of Attributes Functionally Determining a Decision Attribute" RSEISP '07**:** Proceedings of the international conference on Rough Sets and Intelligent Systems Paradigms .

[14] Hu.K, Sui.Y, Lu.Y, Wang.J, and Shi.C, "Concept approximation in concept lattice,*Proceedings of 5th Pacific-Asia Conference on Knowledge Discovery and Data Mining"**,PAKDD'01*, 167-173, 2001.

AUTHORS PROFILE

Graduated in AM.I.E.T.E. from I.E.T.E, New Delhi, India, in 1997 and M.Tech in Computer science from Osmania University,Hyderabad, A.P.,India in 2003. Currently working in Hyderabad Institute of Technology and Management as Associate professor in CSE department (HITAM) R.R.Dist, A.P, and India. She has 8 years of experience. Her research interests include Data mining, Distributed Systems, Information Retrieval systems.

Professor T.Venkat Narayana Rao, received B.E in Computer Technology and Engineering from Nagpur University, Nagpur, India, M.B.A (Systems) and M.Tech in Computer Science from Jawaharlal Nehru Technological University, Hyderabad, A.P., India and a Research Scholar in JNTU. He has 20 years of vast experience in Computer Science and Engineering areas pertaining to academics and industry related I.T issues. He is presently Professor and Head, Department of Computer Science and Engineering, Hyderabad Institute of Technology and Management (HITAM), Gowdavally, R.R.Dist., A.P, INDIA. He is nominated as an Editor and Reviewer to 15 International journals relating to Computer Science and Information Technology. He is currently working on research areas which include Digital Image Processing, Digital Watermarking, Data Mining, Network Security and other Emerging areas of Information Technology . He can be reached at tvnrbobby@yahoo.com

A Generalized Two Axes Modeling, Order Reduction and Numerical Analysis of Squirrel Cage Induction Machine for Stability Studies

Sudhir Kumar and P. K. Ghosh
Faculty of Engineering and Technology
Mody Institute of Technology and Science
Lakshmangarh, District Sikar
Rajasthan 332311, India
skumar_icfai@yahoo.co.in, pkghosh_ece@yahoo.co.in

S. Mukherjee
Department of Electrical Engineering
Indian Institute of Technology
Roorkee, Uttrakhand, India
mukherjee.shaktidev@gmail.com

***Abstract*— A substantial amount of power system load is made of large number of three phase induction machine. The transient phenomena of these machines play an important role in the behavior of the overall system. Thus, modeling of induction machine is an integral part of some power system transient studies. The analysis takes a detailed form only when its modeling becomes perfect to a greater accuracy. When the stator eddy current path is taken into account, the uniform air-gap theory in phase model analysis becomes inefficient to take care of the transients. This drawback necessitates the introduction of analysis of the machine in d-q axis frame. A widely accepted induction machine model for stability studies is the fifth-order model which considers the electrical transients in both rotor and stator windings and the mechanical transients. In practice, some flux-transient can be ignored due to the quasi-stationary nature of the variables concerned. This philosophy leads to the formation of reduced order model. Model Order Reduction (MOR) encompasses a set of techniques whose goal is to generate reduced order models with lower complexity while ensuring that the I/O response and other characteristics of the original model (such as passivity) are maintained. This paper takes the above matter as a main point of research. The authors use the philosophy of the speed-build up of induction machine to find the speed versus time profile for various load conditions and supply voltage disturbances using numerical methods due to Runge-Kutta, Trapezoidal and Euler's in Matlab platform. The established fact of lesser computation time in reduced order model has been verified and improvement in accuracy is observed.**

Keywords- Induction machine, Model order reduction, Stability, Transient analysis.

I. INTRODUCTION

The mathematical description of many physical systems when obtained using theoretical considerations often results in large-order models. To be more precise, in the time domain or state space representation, the modeling procedure leads to a high order state space model and a higher order transfer function model in frequency domain representation. It is often desirable for analysis, simulation and system design to represent such models by equivalent lower order state variable or transfer function models which reflect the dominant characteristics of the system under consideration. Model order reduction is important and common theme within the simulation, control and optimization of complex physical processes. Some of the reason for using reduced order models of higher order system could be (i) to have a better understanding of the system (ii) to reduce computational complexity (iii) to reduce hardware complexity and (iv) to make feasible controller design. The Model order reduction techniques for both type of reduction have been proposed by several researchers. A large number of methods are available in the literature for order reduction of linear continues systems in time domain as well as in frequency domain [1-7]. The extension of single input single output (SISO) methods to reduce multi input multi output (MIMO) systems has been also carried out in [8-11]. Numerous methods of order reduction are also available in the literature [12-17], which are based on the minimization of the integral square error criterion. Some of methods are based on stability criterion. Each of these methods has both advantages and disadvantages when tried on a particular system. In spite of several methods available, no approach always gives the best results for all the systems.

In this paper, two axes modeling (d-q model) using park's transformation theory in state space representation has been developed and reduced order model has also been derived and simulation results are obtained using various numerical methods such as Runge-Kutta, Trapezoidal and Euler method under no load and load conditions and supply voltage disturbances.

Further, it is well known fact that the three phase Induction machines are the work-force in the generator-mode and motor-mode applications in the power systems. As induction motors are becoming day by day a main component of A.C drive systems, a detailed study of the different aspects of its modeling draws the attention of the researchers [18-31]. The important approach of modeling of induction machine is the Park's transformation theory, which converts the machine from phase model to axis model. Two separate coils are assumed to lie on each axis, both for stator and rotor of the induction machine. The voltage balance equations for each coil are formulated and then the torque-balance equation of the machine is established. As transformer emf is a component of voltage balance equation, the machines equation involves the

time rate of changes of flux linkage '(PΨ)', where p stands for the time derivative and 'Ψ' is flux-linkage associated with each coil. The number of P -terms in complete model indicates the order of the model.

This work investigates the different aspects of reduced order modeling of the three-phase induction motor as a power system component. In power system simulation, especially in the transient stability studies, it is desirable to represent induction motor loads with a reduced-order model in order to decrease computational efforts and time. The order reduction is achieved by setting the derivative of stator flux linkage to zero in the stator differential equation in which all equations are referred to the synchronously revolving reference frame. This method of reducing the order of the induction machine is designated as the theory of neglecting stator flux transients.

II. Basic Assumption And Model Of Single Cage Three Phase Induction Machine

The theory of neglecting stator flux transients applied to induction motor, the axis equation modeled in synchronously rotating reference frame requires two assumptions, and these are:

(i) Terminal voltage cannot continually change in magnitude at a high rate for any length of time and it must remain at synchronous frequency
(ii) Oscillating torque component caused by stator flux linkage transients can be safely ignored.

The methods of deriving reduced order equations are described below.

For reducing the order of the model of single cage type of Induction machine, stator flux transients have been neglected. The simplest method of deriving linearized reduced order models of induction machine is the identification of the Eigen- values/poles in the frequency domain and to neglect the faster transient. Therefore, for required process of model order reduction, the time domain equations are to be transformed into Laplace domain. Avoiding the process, one can also reduced the model order, if he assumes rotor resistance of larger value and this involves higher starting torque. The rotor copper loss will increase and efficiency will be poor. Thus the rotor flux transient cannot be neglected.

The transient performance of electrical machines has received increasing attention in recent years primarily due to the application of digital computers to large and complicated models of electromechanical systems [18-29]. Computer simulation of induction machine is performed to determine the machine behavior under various abnormal operating conditions such as start up and sudden change in supplies voltage and load torque, etc.

The detailed induction machine model is non-linear fifth order model, which is referred to as the "full order model". This model takes into account the electrical transients of both the stator and rotor windings, requires a relatively large amount of computational time. Consequently, the third-order model is achieved through neglecting the electrical transients of stator winding, is used to predict the transient behavior of induction motors.

III. Formulation of Full- order Model of Single Cage Induction MachinE

The reduced order model is derived from the full order model of induction machine. The mathematical model is expressed in synchronously rotating d-q references frame using per unit system. The machine flux linkages are selected as state variables and the power invariant transformation is used to convert the phase variable to their equivalent d-q variable.

A. Full- Order Model (5th-Order Model)

The induction machine equations are expressed in synchronously rotating frame of reference:

$$P\psi_{sd} = -a_1\psi_{sd} + \psi_{sq} + a_2\,\psi_{rd} + V_{sd} \quad (1)$$

$$P\psi_{sq} = -\psi_{sd} - a_1\psi_{sq} + a_2\,\psi_{rq} + V_{sq} \quad (2)$$

$$P\psi_{rd} = a_3\,\psi_{sq} + a_4\,\psi_{rq} - (1-w_r)\,\psi_{rd} \quad (3)$$

$$P\psi_{rq} = a_3\,\psi_{sd} + (1-w_r)\,\psi_{rq} - a_4\,\psi_{rd} \quad (4)$$

$$Pw_r = (L_m/2HL_{sr})\,[(\psi_{sq}\psi_{rd} - \psi_{sd}\psi_{rq}) - (L_{sr}\,T_L/L_m)] \quad (5)$$

where,

$$a_1 = R_s\,L_{rr}/L_{sr},\ a_2 = R_sL_m/L_{sr},\ a_3 = R_r\,L_m/L_{sr},$$

$$a_4 = R_r\,L_{ss}/L_{sr} \quad (6)$$

The quantities ψ_{sd} and ψ_{sq} are the flux linkage of the stator winding along direct and quadrature axes respectively. ψ_{rd} and ψ_{rq} refer to those for rotor winding. R_r and R_s are the winding resistance of rotor and stator respectively. L_{ss} and L_{rr} represent the self-inductance of stator and rotor winding respectively, while L_{sr} denotes the mutual inductance between them. The quantities L_s and L_r are the apparent three phase rotor and stator self-inductance respectively and L_m is the magnetizing inductance in the winding. The quantities V_{sd} and V_{sq} refer to the stator terminal voltage along d and q axes respectively. H is the inertia constant in second. We have the relations

$$L_{ss} = L_s + L_m \quad (7)$$

$$L_{rr} = L_r + L_m \quad (8)$$

$$L_{sr} = L_{rr}\,L_{ss} - L_m^2 \quad (9)$$

The phase voltages V_a, V_b and V_c may be expressed in matrix form as

$$\begin{bmatrix} V_a \\ V_b \\ V_c \end{bmatrix} = V_m \begin{bmatrix} \sin\theta \\ \sin(\theta - 2\pi/3) \\ \sin[\theta + 2\pi/3] \end{bmatrix} \quad (10)$$

The d-q voltage are related to the phase voltage by the expression

$$\begin{bmatrix} V_{sd} \\ V_{sq} \end{bmatrix} =$$

$$\sqrt{\frac{2}{3}} \begin{bmatrix} \sin\theta & \sin(\theta - 2\pi/3) & \sin(\theta + 2\pi/3) \\ \cos\theta & \cos(\theta - 2\pi/3) & \cos(\theta + 2\pi/3) \end{bmatrix} \begin{bmatrix} V_a \\ V_b \\ V_c \end{bmatrix} \tag{11}$$

In equation (10) and (11), θ = wt + α is the phase angle, and α is the switching time.

The Stator d-q currents expressed in term of flux linkage can be written as

$$\begin{bmatrix} i_{sd} \\ i_{sq} \end{bmatrix} = \begin{bmatrix} (L_{rr}\psi_{sd} - L_m\psi_{rd})/L_{sr} \\ (L_{rr}\psi_{sq} - L_m\psi_{rq})/L_{sr} \end{bmatrix} \tag{12}$$

Representing by p and q the active and reactive powers, we may have the matrix equation:

$$\begin{bmatrix} P \\ Q \end{bmatrix} = \begin{bmatrix} V_{sd} & V_{sq} \\ V_{sq} & -V_{sd} \end{bmatrix} \begin{bmatrix} i_{sd} \\ i_{sq} \end{bmatrix} \tag{13}$$

Simulation results are obtained using various numerical methods such as Runge-Kutta, Trapezoidal and Euler method under no load and load conditions and supply voltage disturbances. The flowchart for simulation algorithm is given in figure 1.

IV. FORMULATION OF REDUCED-ORDER MODEL OF SINGLE CAGE INDUCTION MOTOR

Reduced order model applied to normal three-phase induction Machine of single cage design allows to neglect the stator flux Transient. This assumption makes the d/dt = P terms reduced in comparison to full order model hence the resulting matrixes of flux-linkage equation will be simpler and computational time will be less.

Neglecting Stator Transient

The fifth-order model is reduced to the third order model by setting the time rate of change of the stator flux linkage to zero. Solving these flux linkages after setting their derivatives to zero in (1) and (2) yields

$$\begin{bmatrix} \psi_{sd} \\ \psi_{sq} \end{bmatrix} =$$

$$\frac{1}{a_1^2 + 1} \left\{ \begin{bmatrix} a_1 a_2 & a_2 \\ -a_2 & a_1 a_2 \end{bmatrix} \begin{bmatrix} \psi_{rd} \\ \psi_{rq} \end{bmatrix} + \begin{bmatrix} a_1 & 1 \\ -1 & a_1 \end{bmatrix} \begin{bmatrix} V_{sd} \\ V_{sq} \end{bmatrix} \right\} \tag{14}$$

Substituting above in (3)-(5) results in the third order model

$$\frac{d\psi_{rd}}{dt} = (b_1 - \varpi_r)\psi_{rq} + b_2\psi_{rd} + b_3 V_{sq} + b_4 V_{sd} \tag{15}$$

$$\frac{d\psi_{rq}}{dt} = b_2 - \psi_{rq} - (b_1 - \varpi_r)\psi_{rd} - b_4 V_{sq} - b_3 V_{sd} \tag{16}$$

$$\frac{d\psi_r}{dt} = \frac{L_m}{2HL_{sr}(a_1^2 + 1)} \left\{ \begin{aligned} &-a_2(\psi_{dr}^2 + \psi_{qr}^2) - \frac{L_{sr}^2(a_1^2 + 1)}{L_m} T_1(a_1\upsilon_{sd} + \upsilon_{sq})\psi_{qr} \\ &+ (a_1\upsilon_{sq} + \upsilon_{sd})\psi_{dr} \end{aligned} \right\} \tag{17}$$

where

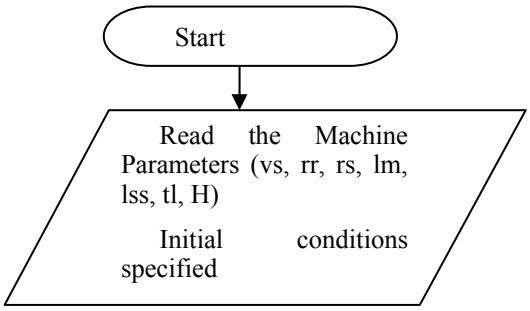
$$b_1 = 1 + \frac{a_2 a_3}{a_1^2 + 1},\ b_2 = a_4 + \frac{a_1 a_2 a_3}{a_1^2 + 1}$$

$$b_3 = \frac{a_3}{a_1^2 + 1},\quad b_4 = a_4 + \frac{a_1 a_3}{a_1^2 + 1}.$$

V. RESULTS AND DISCUSSIONS

The simulation results and performance characteristics of full order model and reduced order of single cage induction machine for stator current, stator flux linkage, rotor speed etc. have been obtained as shown in figures 2(a-c) and 3(a-c) under no load and load torque disturbances and supply voltages changes. The machine parameters for the single cage induction machine under analysis are given in table I.

Reduced order model characteristics have been compared with their full order model. The characteristics obtained for full order model and reduced order model are almost equal. The computational time obtained for full order model and reduced order model using three numerical methods are also compared in table II. The algorithm presented in this paper introduces a considerable advantage in computational speed as indicted by comparative results given in the table II.



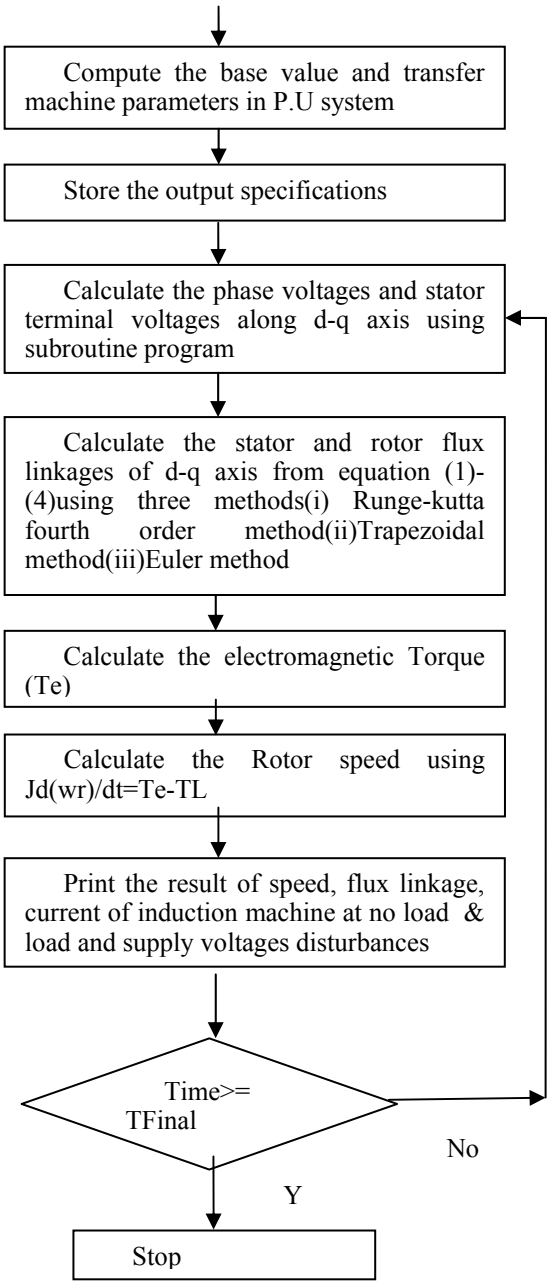


Figure 1: Flow chart of simulation algorithm

TABLE I. SPECIFICATION AND MACHINE PARAMETERS OF SINGLE CAGE INDUCTION MACHINE

2 HP, 1200 RPM, 6 Pole, 60 Hz, THREE PHASE MACHINE	
Machine parameters	**P.U Value**
Supply Voltage	200
Stator Resistance (R_s)	0.1742
Rotor Resistance (R_r)	0.0637
Self Inductance of stator winding (L_{ss})	1.754
Mutual Inductance between Stator & Rotor (L_{sr})	2.475
Magnetizing Inductance (L_m)	1.65
Inertia Constant (H)	0.0331 sec
Mutual Inductance between stator and rotor(L_{sr})	2.475

TABLE II. COMPARISON OF COMPUTATIONAL TIME OBTAINED FOR FULL ORDER AND REDUCED ORDER MODELS IN THREE NUMERICAL METHODS USING MATLAB

Runge-Kutta (RK4) Solver	
Model	**Computation Time(ms)**
Full order model	491
Reduced order model	282
Trapezoidal Solver	
Model	**Computation Time(ms)**
Full order Model	545
Reduced order model	312
Euler Solver	
Model	**Computation Time(ms)**
Full order Model	402
Reduced order model	235

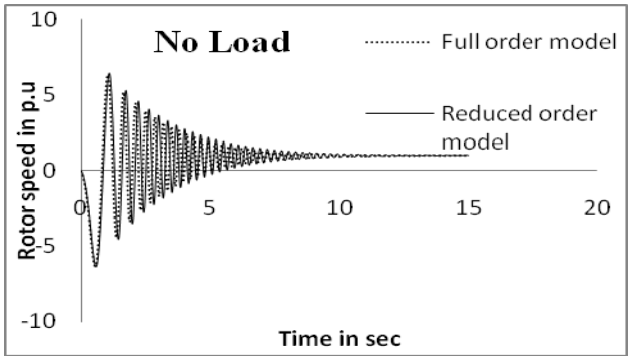


Figure: 2 (a)

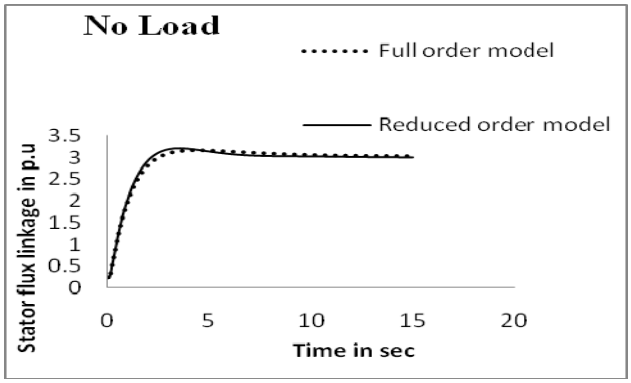


Figure: 2 (b)

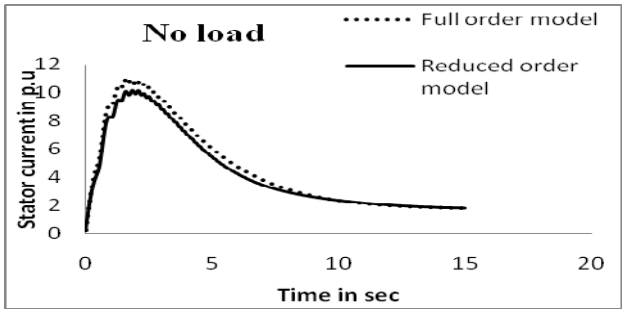


Figure: 2(c)

Fig2 (a-c): Response of Full order Model & Reduced order model under no load condition and at rated supply voltage

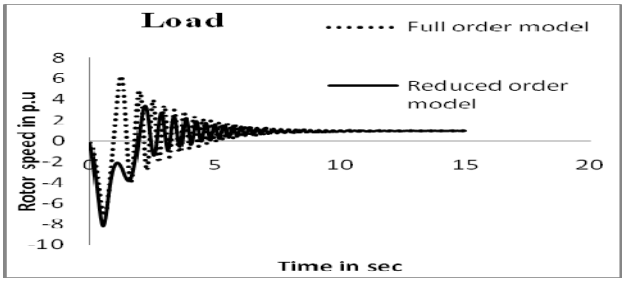


Figure: 3(a)

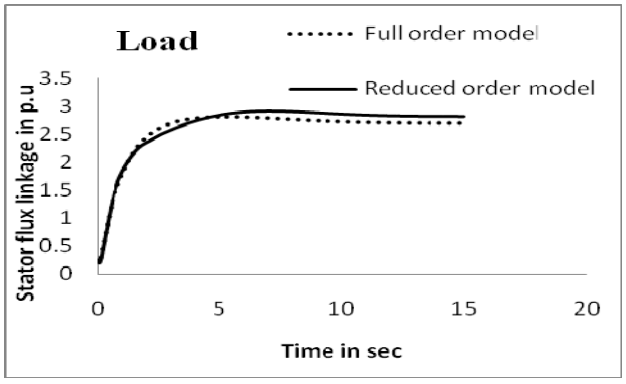


Figure: 3 (b)

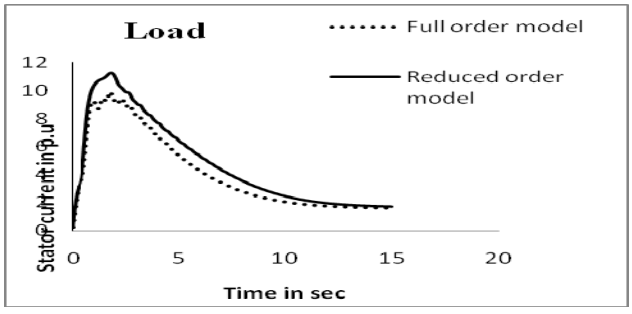


Figure: 3(c)

Figure 3(a-c): Response of Full Order Model & Reduced order Model under load condition and supply voltage change (10%)

VI. CONCLUSIONS

The reduced order model of single cage induction machine was validated by digital simulation and compared with the response of full order model using three numerical methods such as Runge-Kutta method, Trapezoidal method and Euler's method. The reduced order model is applied to a 2 HP, 1200 rpm, 6 pole, 60 Hz single cage induction machine and its performance characteristics is compared with the full order model. The performance characteristic of reduced order model is approximately same as full order model of single cage induction machine. The simulation time taken for full order and reduced order models are calculated using three numerical methods. The results show the saving of considerable amount of time in reduced order model as compared to full order model.

The numerical techniques, namely, Runge-Kutta method Trapezoidal method and Euler's method for solving the non-linear differential equation using MATLab are used to integrate the differential equation and integration size is taken to be 0.002 second. The quantities selected for simulation are stator current, stator flux-linkages, rotor speed under no load and load conditions and supply voltage disturbances.

In this paper, the method of solution has been decided based on the physical fact of the speed build-up of the induction machine. This physical insight is a must because the speed build-up of induction machine is equivalent to flux-linkages build-up iterated with electromagnetic torque neglecting the stator flux transients in the full order model. This process governs the algorithm for solution of flux-linkages and speed. The reduced order model demonstrates the improvement of the processing time considerably.

ACKNOWLEDGMENT

Sudhir Kumar sincerely acknowledges the valuable discussions made with Prof. (Dr.) A. B. Chattopadhyay, Birla Institute of Technology and Science, Pilani (Rajasthan) Dubai campus, and Professor Vivek Tiwari, National Institute of Technology, Jamshedpur (Jharkhand) India.

REFERENCES

[1] R. Genesio and M. Milanese, "A note on the derivation and use of reduced order models", *IEEE Trans. Automat. Control*, Vol. AC-21,No. 1, pp. 118-122, February 1976.

[2] M. S. Mahmoud and M. G. Singh, *Large Scale Systems Modelling*, Pergamon Press, International Series on Systems and Control 1st ed.,Vol. 3, 1981.

[3] M. Jamshidi, *Large Scale Systems Modelling and Control Series*, New York, Amsterdam, Oxford, North Holland, Vol. 9, 1983.

[4] S. K. Nagar and S. K. Singh, "An algorithmic approach for system decomposition and balanced realized model reduction", *Journal of Franklin Inst.*, Vol. 341, pp. 615-630, 2004.

[5] V. Singh, D. Chandra and H. Kar, "Improved Routh Pade approximants:

A computer aided approach", *IEEE Trans. Automat. Control*, Vol. 49, No.2, pp 292-296, February 2004.

[6] S. Mukherjee, Satakshi and R.C.Mittal, "Model order reduction using response-matching technique", Journal *of Franklin Inst.*, Vol. 342 , pp. 503-519, 2005.

[7] B. Salimbahrami, and B. Lohmann, "Order reduction of large scale second-order systems using Krylov subspace methods", *Linear Algebra Appl.*, Vol. 415, pp. 385-405, 2006.

[8] S. Mukherjee and R.N. Mishra, "Reduced order modelling of linear multivariable systems using an error minimization technique", *Journal of Franklin Inst.*, Vol. 325, No. 2 , pp. 235-245, 1988.

[9] S. S. Lamba, R. Gorez and B. Bandyopadhyay, "New reduction technique by step error minimization for multivariable systems", *Int. J.Systems Sci.*, Vol. 19, No. 6, pp. 999-1009, 1988.

[10] R. Prasad and J. Pal, "Use of continued fraction expansion for stable reduction of linear multivariable systems", *Journal of Institution of* Engineers, India, IE(I) Journal – EL, Vol. 72, pp. 43-47, June 1991.

[11] R. Prasad, A. K. Mittal and S. P. Sharma, "A mixed method for the reduction of multi-variable systems", *Journal of Institution of Engineers,India, IE(I) Journal – EL*, Vol. 85, pp. 177-181, March 2005.

[12] T.C. Chen, C.Y. Chang and K.W. Han, "Reduction of transfer functions by the stability equation method", Journal *of Franklin Inst.,* Vol. 308, No. 4, pp. 389-404, 1979.

[13] T.C. Chen, C.Y. Chang and K.W. Han, "Model reduction using the stability equation method and the continued fraction method", *Int. J. Control*, Vol. 32, No. 1, pp. 81-94, 1980.

[14] C.P. Therapos, "Stability equation method to reduce the order of fast oscillating systems", *Electronic Letters*, Vol. 19, No.5, pp.183-184, March 1983.

[15] T.C. Chen, C.Y. Chang and K.W. Han, "Stable reduced order Pade approximants using stability equation method", *Electronic Letters*, Vol. 16, No. 9, pp. 345-346, 1980.

[16] J. Pal, "Improved Pade approximants using stability equation method", *Electronic Letters*, Vol. 19, No.11, pp.426-427, May 1983.

[17] R. Parthasarathy and K.N. Jayasimha, "System reduction using stability equation method and modified Cauer continued fraction", *Proceedings IEEE*, Vol. 70, No. 10, pp.1234-1236, October 1982.

[18] G. G. Richard and O.T.Tan "Simplified model for induction machine transients under Balanced & unbalanced condition", *IEEE Trans. on industrial Application*, Jan /Feb 1981.

[19] F.D.Radrigas and O.Waaynaxuk, "A method of representing isolated operation of Group of induction motor loads" *, IEEE Trans. on power systems*, Vol. PWRS-2, No.3, pp. 568-575, August, 1987.

[20] Stag & EI-Abiad, *Computer method in power system analysis*, 1968. International student edition.

[21] G.G Richard, "Reduced order model for an induction motor group during Bus-Transfer", *IEEE Trans. on power system*, Vol.1, No. 2 pp. 494-498, May 1989.

[22] T.L. Savanna and P.C. Krause, "Accuracy of a Reduced order model of induction machine in dynamo-stability studies", *IEEE Trans. on AC power app. & system*, Vol. PAS-98, No. 4, pp. 1192-1197, July/Aug 1979.

[23] S.Ertem andY.Baghzouz, "Simulation of induction machinery for power system studies", *IEEE Trans. on Energy conversion*, Vol- 4, No. 1, March 1989.

[24] D. Wasynvzuk, Y.Diao &P.C. Krause, "Theory and comparison of reduced - order model of induction machines", *IEEE Trans. power apparatus & system*, Vol. PAS 104, No.3, pp. 598-606, March 1985.

[25] D.G Shultz & J.L. Melsa, *State function and linear control system*, Mc-Graw Hill ,1957

[26] B. Adkins and R.Harley, *The general Theory of Alternating Current Machine: Application to practical problems. London*, Chapman and Hall, 1975

[27] P.C Krause and C.H Thomas, "Simulation of Electrical Machinery", *IEEE Trans. Power Appratus and Systems*, Vol..PAS-86, pp.1038-1053, Nov 1965.

[28] G.G. Richards and O.T Tan, "Decomposed, Reduced order Model for Double cage induction Machine", *IEEE Trans. on Energy conversion*, Vol. EC-1. No.3, pp. 87-93, Sep. 1986.

[29] N.A. Kh N.A. Khalil, O.T Tan, I.U Baran. "Reduced order model for double cage induction machines", IEEE Trans. power apparatus and System, Vol. PAS-101, No.9, pp. 3135.39, Sept. 1982.

[30] G. Parmar, S. Mukherjee and R. Prasad, "System reduction using factor division algorithm and eigen spectrum analysis", *Applied Mathematical Modelling*, Elsevier, Vol. 31, pp 2542-2552, 2007.

[31] G. Parmar, S. Mukherjee and R. Prasad, "Reduced order modelling of linear MIMO systems using Genetic Algorithm", *Int. Journal of Simulation Modelling*, Vol. 6, No. 3, pp. 173-184, 2007.

AUTHORS PROFILE

Sudhir Kumar was born in Patna, India in 1974. He received his B. Tech in Electrical and Electronics Engineering from Bangalore University in 1996 and M. Tech in Electrical Power Systems Engineering from National Institute of Technology, Jamshedpur, Jharkhand, India in 2000. Presently, he is working as Assistant Professor in Faculty of Engineering and Technology, Mody, Institute of Technology and Science (Deemed University) at Laxmangarh, Sikar (Rajasthan), India. He is Associate Member of Institution of Engineers, India (AMIE) and Life member of ISTE. His areas of interests includes Power systems, Network analysis, Control Systems, and Model Order Reduction of large scale systems.

Dr. P. K. Ghosh was born in Kolkata, India in 1964. He received his B.Sc (Hons in Physics), B.Tech and M.Tech degrees in 1986, 1989, and 1991, respectively from Calcutta University. He earned Ph.D.(Tech) degree in Radio Physics and Electronics in 1997 from the same University. He served various institutions, namely, National Institute of Science and Technology (Orissa), St. Xavier's College (Kolkata), Murshidabad College of Engineering and Technology (West Bengal), R. D. Engineering College (Uttar Pradesh) and Kalyani Government Engineering College (West Bengal) before he joins Mody Institute of Technology and Science (Rajasthan). He is life member of Indian Society for Technical Education (ISTE), New Delhi. His research interests are in the areas of reduced order modeling, VLSI circuits and devices and wireless communications.

Dr. Shaktidev Mukherjee was born in Patna, India, in 1948. He received B.Sc. (Engg.), Electrical from Patna University in 1968 and M.E., Ph.D. from the University of Roorkee in 1977 and 1989, respectively. After working in industries till 1973, he joined teaching and taught in different academic institutions. He is the former Vice-Chancellor of Mody Institute of Technology and Science, (Deemed University), Lakshmangarh, Sikar (Rajasthan), India. Presently, he is Professor in the Department of Electrical Engineering at Indian Institute of Technology, Roorkee ,India. He is a life member of Systems Society of India (LMSSI). His research interests are in the areas of Model Order Reduction and Process Instrumentation and Control.

Single Input Multiple Output (SIMO) Wireless Link with Turbo Coding

M.M.Kamruzzaman*
Information &Communication Engineering
Southwest Jiaotong University,
Chengdu, Sichuan 610031, china
E-mail: kamruzzaman@qq.com

Dr. Mir Mohammad Azad**
Computer Science and Engineering
Shanto Mariam University of Creative
Technology, Dhaka, Bangladesh
E-mail: azadrayhan@gmail.com

***Abstract*— Performance of a wireless link is evaluated with turbo coding in the presence of Rayleigh fading with single transmitting antenna and multiple receiving antenna. QAM modulator is considered with maximum likelihood decoding. Performance results show that there is a significant improvement in required signal to noise ratio (SNR) to achieve a given BER. It is found that the system attains a coding gain of 14.5 dB and 13 dB corresponding to two receiving antenna and four receiving antennas respectively over the corresponding uncoded system. Further, there is an improvement in SNR of 6.5 dB for four receiving antennas over two receiving antennas for the turbo coded system.**

Keywords- Diversity; multipath fading; Rayleigh fading; Turbo coding; Maximal Ratio Combining (MRC); maximum-likelihood detector; multiple antennas.

I. INTRODUCTION

The increasing demand for high data rates in wireless communications due to emerging new technologies makes wireless communications an exciting and challenging field[1-2]. Turbo code is the most exciting and potentially important development in the coding theory which is capable of achieving near Shannon capacity performance [3-5]. Diversity technique can be applied to minimize the effect of fading in a wireless channel. Space time block coding is found to be an effective means to combat fading in a multipath propagation environment [6-7]. In [6-7], it is found that use of two transmit antenna and multiple receive antennas and space time block coding provides remarkable performance improvement. The use of space-time block coding with concatenating turbo product codes is also recently reported .Instead of using STBC, the performance of a wireless link may be improved by applying turbo coding in a SIMO configuration. In this paper, performance analysis of a wireless link with turbo coding in conjunction with multiple receiving antenna diversity is presented for a rayleigh fading channel with single input multiple output (SIMO) configuration. Performance result is presented in terms of bit error rate for multiple receiving antenna diversity schemes. Computer simulation is carried out to evaluate the performance result.

II. SYSTEM MODEL

The block diagram of the system under consideration is shown in Fig.1 and Fig. 2. The data is turbo encoded with frame size 1024, with the following generator matrix with code rate ½ and

$$G(D) = \left[1, \frac{1+D+D^3}{1+D^2+D^3}, \frac{1+D+D^2+D^3}{1+D^2+D^3}\right]$$

modulated by a 64 QAM modulator before transmission. At the receiving end, signal is received by multiple receiving antennas. The detail block diagram of the combiner is shown in Fig.3.

If the symbol transmitted by transmit antenna is s1, then the signal received by the receiving antennas can be written as:

$$r_1 = h_1 s1 + \eta_1$$

$$r_2 = h_2 s1 + \eta_2$$

$$\vdots$$

$$r_n = h_n s1 + \eta_n$$

Where $h_{1\ldots}h_n$ and $\eta_{1\ldots..}$ η_n represent the channel coefficients and additive white noise for IF link.

The combiner combines received signals which are then sent to the maximum likelihood detector. If we consider two receiving antennas, then received signal r_0 and r_1 are fed to combiner and then combiner generates the following signal [6]:

$$\tilde{s}_1 = h_1^* r_1 + h_2 r_2^* \quad (1)$$

for more then two receive antenna then r_0, r_1, r_2, r_3 ---------- r_n

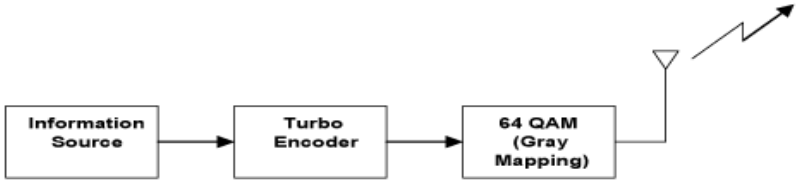


Figure 1. Block diagram of transmitter with turbo encoder and one transmit antenna

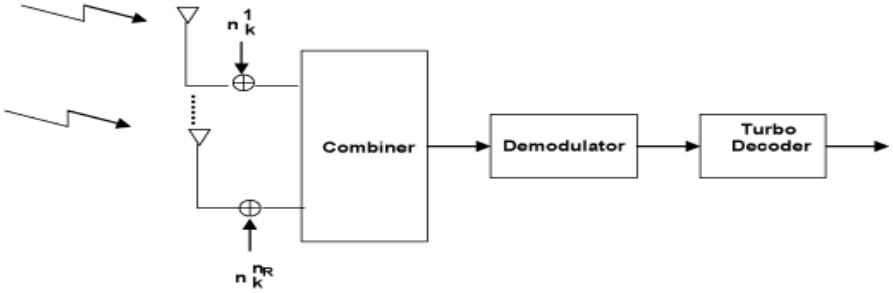


Figure 2. Block diagram of receiver with multiple receiving antennas and turbo decoder.

Fig. 1. : Block diagram of a wireless link with turbo coding and multiple receiving antennas
(a) Transmitter (b) Receiver

received signal are send to combiner and combiner generate the following signals [6]:

$$\tilde{s}_1 = h_1^* r_1 + h_2 r_2^* + h_3^* r_3 + h_4 r_4^* \text{ ------ } h_{n-1}^* r_{n-1} + h_n r_n^* \quad (2)$$

III. MAXIMUM LIKELIHOOD DECISION RULE

Maximum likelihood decoding of combined signal $\tilde{s}_1$ can be achieved using the decision metric [7]

$$\sum_{i=2,4,6....}^{n} \left(\left| r_{i-1} - h_{i-1} s_1 \right|^2 + \left| r_i - h_i s_1^* \right|^2 \right) \quad (3)$$

Over all possible values of s1.

We expand the above equation and delete the terms that are independent of the code words and observe that the above minimization is equivalent to minimizing [7]

The above can be rewrite:

$$- \sum_{i=2,4,6.....}^{n} \left[r_{i-1} h_{i-1}^* s_1^* + \left(r_i \right)^* h_i s_1 \right] + \left| s_1 \right|^2 \sum_{i=1}^{n} \left| h_i \right|^2 \quad (4)$$

Thus the minimization of **(1)** and in turn is equivalent to minimizing the decision metric [7]

$$\left| \left[\sum_{i=2,4,6.....}^{n} \left(r_{i-1} h_{i-1}^* + \left(r_i \right)^* h_i \right) \right] - s_1 \right|^2 + \left(-1 + \sum_{i=1}^{n} \left| h_i \right|^2 \right) \left| s_1 \right|^2 \quad (5)$$

for detecting s1.

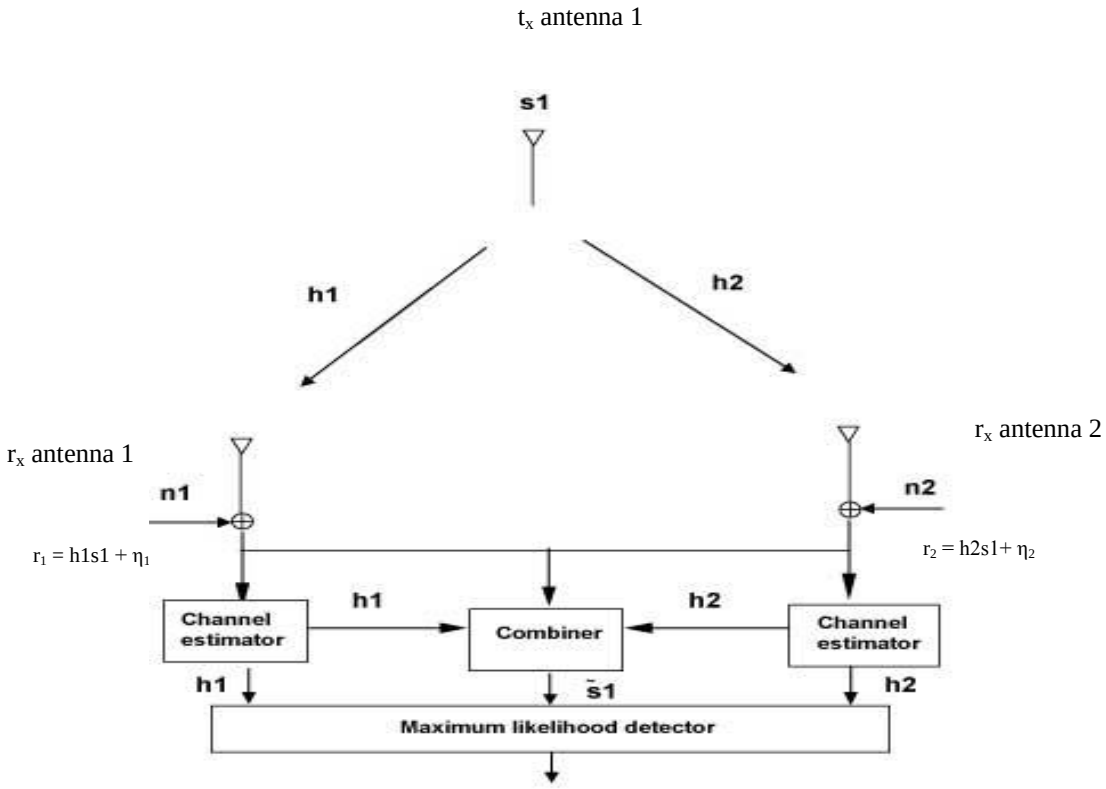


Figure 3. Receive diversity with one transmit antenna and two receive antenna

Output of the combiner is demodulated using a coherent demodulator. Decoding of the turbo coded bits are carried out following the MAP decoding algorithm.

IV. RESULT AND DISCUSSION

Computer simulation is carried out to find the bit error rate (BER) performance of a wireless transmission system with turbo coding with multiple receiving antennas. The results are evaluated for several combinations of t_x and r_x antennas with and without Turbo coding. The plots of BER as a function of SNR (E_s/N_o) are shown in fig.3. It is found that there is a significant improvement in required SNR for achieving a given BER with two receiving antenna compared to four receiving antennas of a turbo coded system. The coding gain is found to be 6.5 dB at BER 10^{-6} of a turbo coded system for two receiving antenna compared to four receiving antennas. There is also significant improvement with turbo code compared to the uncoded system. For two receiving antennas, the coding gain is found to be 14.5 dB and for four receiving antennas, the coding gain is found to be 13 dB at BER 10^{-6} over the uncoded system with same numbers of receiving antenna. Thus the effectiveness of turbo coding is more significant for higher number of receiving antenna.

ACKNOWLEDGMENT

The authors would like to thank anonymous reviewers for helpful comments to improve the presentation of the paper.

REFERENCES

[1] Haowei Bai, Honeywell Aerospace; Mohammed Atiquzzaman, University of Oklahoma "Error Modeling Schemes for Fading channels in wireless Communications: A Survey" IEEE Communications Surveys & Tutorials, 2003

[2] B.Sklar,"Rayleigh Fading Channels in Mobile Digital Communication Systems, part I: Characterization, "IEEE Commun. Mag., vol. 35, no. 9, sept. 1997,pp.136-46.

[3] G. Berrou, A.Glavieuc, and P.Thitmajshima, "Near Shannon limit error-correcting coding: Turbo codes," in Proc. 1993, Int. Conf. Com., Geneva, Seitzerland, May 1993, pp.1064-1070.

[4] S.Benedetto, D.Divsalar, G.Montorsi, and F.Pollara, "Soft output decoding algorithms in iterative decoding of Turbo codes", TDA progress report 42-124, February 15, 1996.

[5] Mustafa Eroz, A.Roger Hammons Jr., "On the design of prunable interleavers for Turbo codes," in Proc. IEEE VTC'99, Houston, TX, May15-19, 1999.

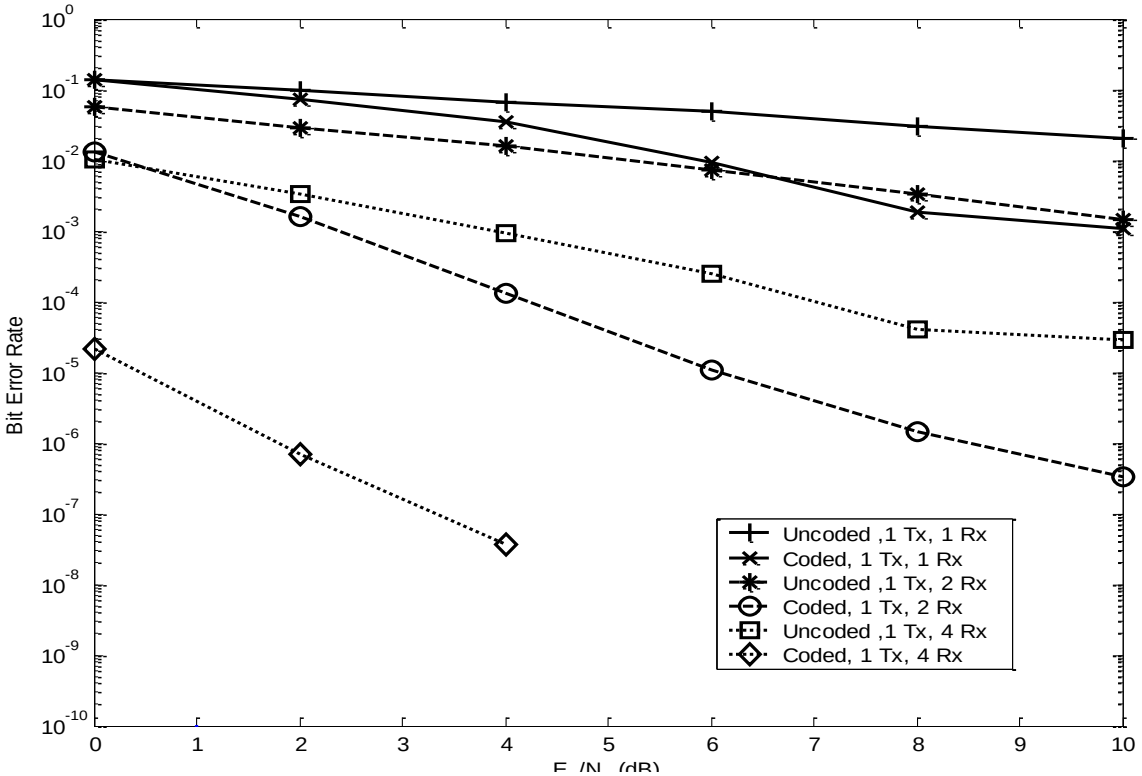


Figure 4. Plots of BER versus E_s/N_o for a wireless system with and without turbo codifferent number of receiving antennas.

[6] S.M. Alamouti, "A simple transmit diversity scheme for wireless communications", IEEE J. Select.Areas Commun., vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[7] Vahid Tarokh, Hamid Jafarkhani, and A. Robert Calderbank, "Space–time block coding for wireless communications: performance results", IEEE J. Select. Areas Commun., vol. 17, pp. 451–460, Mar. 1999.

[8] Yinggang Du Kam Tai Chan "Enhanced space-time block coded systems by concatenating turbo product codes" Communications Letters, IEEE, Volume: 8, Issue: 6 On page(s): 388 – 390 June 2004

[9] Vahid Tarokh and Hamid Jafarkhani, "A Differential Detection Scheme forTransmit Diversity," IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 18, NO. 7, JULY 2000

[10] Zhuo Che_n□, Branka Vucetic□, Jinhong Yuan□_, and Ka Leong Lo, "Analysis of Transmit Antenna Selection/Maximal-Ratio Combining in Rayleigh Fading," IEEE Channels Issue Date : 29 June-4 July 2003 On page(s): 94 [Information Theory, 2003. Proceedings. IEEE International Symposium on]

[11] Rodrigues, M.R.D. ; Chatzgeorgiou, I. ; Wassell, I.J. ; Carrasco, R.,"On the Performance of Turbo Codes in Quasi-Static Fading Channels," IEEE Issue Date : 4-9 Sept. 2005 On page(s): 622 – 626 [Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on]

[12] Yi Hong, Jinhong Yuan, Zhuo Chen and Branka Vucetic, "Space–Time Turbo Trellis Codes for Two,Three, and Four Transmit Antennas" IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, VOL. 53, NO. 2, MARCH 2004

[13] Hicham Bouzekri and Scott L. Miller, "An Upper Bound on Turbo Codes Performance Over Quasi-Static Fading Channels," IEEE COMMUNICATIONS LETTERS, VOL. 7, NO. 7, JULY 2003

[14] Vivek Gulati and Krishna R. Narayanan, "Concatenated Codes for Fading Channels Based on Recursive Space–Time Trellis Codes ," IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, VOL. 2, NO. 1, JANUARY 2003

[15] Vahid Tarokh, Nambi Seshadri and A. R. Calderbank, "Space–Time Codes for High Data Rate Wireless Communication:Performance Criterion and Code Construction," IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 44, NO. 2, MARCH 1998

[16] Gerhard Bauch and Naofal Al-Dhahir, "Reduced-Complexity Space–Time Turbo-Equalization for Frequency-Selective MIMO Channels, " IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, VOL. 1, NO. 4, OCTOBER 2002

[17] Hyundong Shin and Jae Hong Lee, "Channel Reliability Estimation for Turbo Decoding in Rayleigh Fading Channels With Imperfect Channel Estimates," IEEE COMMUNICATIONS LETTERS, VOL. 6, NO. 11, NOVEMBER 2002

[18] Caijun Zhong and Shi Jin, " Capacity Bounds for MIMO Nakagami-Fading Channels," IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 57, NO. 9, SEPTEMBER 2009

[19] Che Lin, Vasanthan Raghavan, and Venugopal V. Veeravalli, "To Code or Not to Code Across Time: Space-Time Coding with Feedback," IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 26, NO. 8, OCTOBER 2008

[20] Naveen Mysore and Jan Bajcsy, "BER Bounding Techniques for Selected Turbo Coded MIMO Systems in Rayleigh Fading," IEEE Publication Year: 2010 , Page(s): 1 – 6 [Information Sciences and Systems(CISS), 2010 44th Annual Conference on]

[21] Khoirul Anwar and Tadashi Matsumoto, "MIMO Spatial Turbo Coding with Iterative Equalization," IEEE Issue Date : 23-24 Feb. 2010 On page(s): 428 - 433 [Smart Antennas (WSA), 2010 International ITG Workshop on]

AUTHORS PROFILE

* **M. M. Kamruzzaman**, Received B.E. degree from Bangalore University, India and M.S. degree from United International University, Bangladesh in Computer Science and Engineering and now Studying PhD in the department of Information & communication Engineering at Southwest Jiaotong University, Chengdu, Sichuan 610031, china. Before doing PhD he was working as faculty of Presidency University, Dhaka, Bangladesh. His areas of interest include modern coding theory, Turbo coding, Space Time Coding and wireless communications

**** Dr. Mir Mohammad Azad**, Received Bachelor of Computer Application (BCA) Degree from Bangalore University, Karnataka, India, Masters of Computer Application (MCA) Degree from Bharath Institute of Higher Education and Research Deemed University (Presently Bharath University), Tamilnadu India. Received PhD in Computer Science from Golden State University. Presently Working as an Assistant Professor and Departmental Head of Computer Science and Engineering, Shanto Mariam University of Creative Technology, Dhaka, Bangladesh. His areas of interest include Networking and Wireless communications.
.

Reliable Multicast Transport Protocol: RMTP

Pradip M. Jawandhiya
Research Scholar, Jawaharlal Darda Inst.of Enggneering & Technology,
MIDC,Lohara, Yavatmal – 445001(M.S.), INDIA
+91 9822726385
pmjawandhiya@rediffmail.com

Dr. M. S. Ali
Principal, Prof. Ram Meghe Inst.of Technology & Management,
Bypass highway road, Badnera,
Dist. Amravati (M.S.), INDIA
+91 9370155150
softalis@hotmail.com

Sakina F.Husain & M.R.Parate
Faculty,Jawaharlal Darda Institute of Engineering & Technology,
MIDC,Lohara, Yavatmal – 445001(M.S.), INDIA
+91 9890140576
Skap_88_e@yahoo.com
Mayur_parate@yahoo.in

Prof. J.S.Deshpande
Pro Vice-Chancellor Sant Gadge Baba Amravati University, Amravati, INDIA
+91 9422881986

***Abstract*— This paper presents the design, implementation, and performance of a reliable multicast transport protocol (RMTP). RMTP is based on a hierarchical structure in which receivers are grouped into local regions or domains and in each domain there is a special receiver called a designated receiver (DR) which is responsible for sending acknowledgments periodically to the sender, for processing acknowledgment from receivers in its domain, and for retransmitting lost packets to the corresponding receivers. Since lost packets are recovered by local retransmissions as opposed to retransmissions from the original sender, end-to-end latency is significantly reduced, and the overall throughput is improved as well. Also, since only the DR's send their acknowledgments to the sender, instead of all receivers sending their acknowledgments to the sender, a single acknowledgment is generated per local region, and this prevents *acknowledgment implosio*n. Receivers in RMTP send their acknowledgments to the DR's periodically, thereby simplifying error recovery. In addition, lost packets are recovered by selective repeat retransmissions, leading to improved throughput at the cost of minimal additional buffering at the receivers. This paper also describes the implementation of RMTP and its performance on the Internet.**

Keywords- multicast routing, MANET, acknowledgement implosion, designated receiver

I. INTRODUCTION

MULTICASTING provides an efficient way of disseminating data from a sender to a group of receivers. Instead of sending a separate copy of the data to each individual receiver, the sender just sends a single copy to all the receivers. A multicast tree is set up in the network with the sender as the root node and the receivers as the leaf nodes. Data generated by the sender goes through the multicast tree, traversing each tree edge exactly once. However, distribution of data using the multicast tree in an unreliable network does not guarantee reliable delivery, which is the prime requirement for several important applications, such as distribution of software, financial information, electronic newspapers, billing records, and medical images. Reliable multicast is also necessary in distributed interactive simulation (DIS) environment, and in collaborative applications. Therefore, reliable multicasting is an important problem which needs to be addressed. Several papers have addressed the issue of multicast routing [1], but the design of a reliable multicast transport protocol in broadband packet-switched networks has only recently received attention [2].Reliable multicast protocols are not new in the area of distributed and satellite broadcast systems [3]. However, most of these protocols apply to local area networks and do not scale well in wide area networks, mainly because the entities involved in the protocol need to exchange several control messages for coordination purposes. In addition, they do not address fundamental issues of flow control, congestion avoidance, end-to-end latency, and propagation delays which play a critical role in wide area networks. Several new distributed systems have been built for group communication recently, namely, Totem [10] and Transis [7]. Totem [10] provides reliable totally ordered multicasting of messages based on which more complex distributed applications can be built. Transis [7] builds the framework for fault tolerant distributed systems by providing mechanisms for merging components of a partitioned network that operate autonomously and later become reconnected.

Both these systems assume the existence of multiple senders and try to impose a total ordering on delivery of packets. However, the reliable multicast transport protocol in this paper has been designed to operate at a more fundamental level where the objective is to deliver packets in ordered lossless manner from a single sender to all receivers. In other words, our protocol can potentially be used by Totem to provide reliable total ordering in a wide area packet-switched network. Other transaction-based group communication semantics like atomic multicast, permanence, and serializability can also be built using our reliable multicast transport protocol. *Multicasting* is a very

broad term and different multicasting applications have, in general, different requirements. For example, a real-time multipoint-to-multipoint multimedia multicasting application, such as, nationwide video conferencing, has very different requirements from a point-to-multipoint reliable data transfer multicasting application, such as, the distribution of software. Recently, researchers have demonstrated multicasting real-time data, such as real-time audio and video, over the Internet using the multicast backbone *(MBon*e) [8]. Since most real-time applications can tolerate some data loss but cannot tolerate the delay associated with retransmissions, they either accept some loss of data or use forward error correction for minimizing such loss. Multicasting of multimedia information has been recently receiving a great deal of attention [4].However, the main objective of these multicast protocols is to guarantee quality of service by reducing end-to-end delay at the cost of reliability. In contrast, the objective of our protocol in this paper is to guarantee *reliability* achieving high throughput, maintaining low end-to-end delay. This is achieved by reducing unnecessary retransmissions by the sender. In addition, we adopt a novel technique of grouping receivers into local regions and generating a single acknowledgment per local region to avoid the acknowledgment implosion problem [12] inherent in any reliable multicasting scheme.

We also use the principle of periodic sending of state information from the receivers to the transmitter to avoid complex error-recovery procedures. Finally we use a selective repeat retransmission scheme to achieve high throughput. In this paper, we describe our detailed experience with the design and implementation of reliable multicast transport protocol (RMTP).

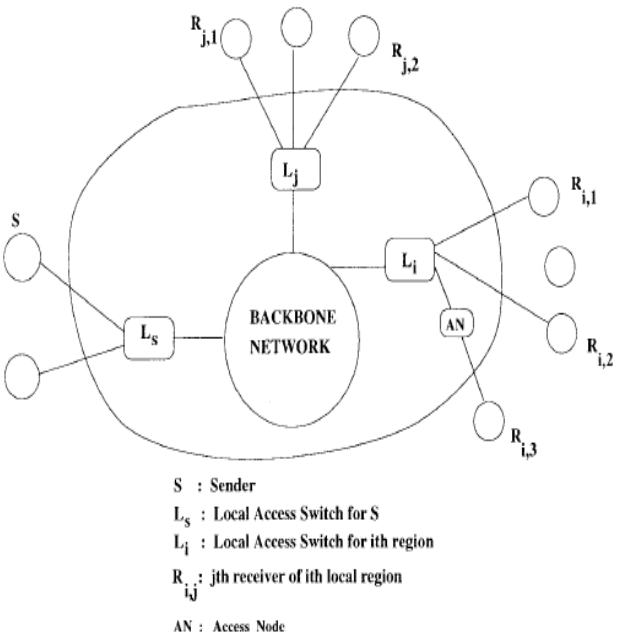


Fig 1: Model of the network

/*The original work consisted of proposing three different multicast transport protocols, comparing them using simulation, and recommending one for reliable multicasting. In fact,*/ the notion of local recovery using a designated receiver (DR) was proposed for the first time in the literature in [11]. The recommended protocol was implemented and its performance, measured on the Internet, was reported in [9]. In this paper, we have combined the ideas and results from [9] and [11] to present a comprehensive picture of our efforts in designing RMTP. RMTP is very general in the sense that it can be built on top of either virtual-circuit networks or datagram networks. The only service expected by the protocol from the underlying network is the establishment of a multicast tree from the sender to the receivers. However, resource reservation is not really necessary for the proper functioning of RMTP. The function of RMTP is to deliver packets from the sender to the receivers in sequence along the multicast tree, independent of how the tree is created and resources are allocated. For example, RMTP can be implemented over available bit rate (ABR) type service in ATM networks for reliable multicasting applications. In this paper, we have addressed the design issues for RMTP in the Internet environment. In particular, the notion of multilevel hierarchy using an internet-like advertisement mechanism is described, and issues related to row control and late-joining receivers in an ongoing multicast session are dealt with extensively. In addition, a detailed description of the implementation using MBone [19] technology in the Internet is also presented and performance measurements are included as well. Most of these ideas and results are taken from [27]. Rest of the paper is organized as follows. Section II discusses the network architecture and the assumptions made in the design of RMTP. Implementation of RMTP is presented in Section III, and its performance evaluation parameters on the Internet and its measurements are presented in Section IV followed by some conclusions.

II. RMTP

2.1 Network Architecture and Assumptions

Let the sender and receiver be connected to the backbone network through local access switches either directly or indirectly through access nodes (fig 2.1) The following are some assumptions made in the protocol design.

a) The receivers can be grouped into local regions [6] based on their proximity in the network. For example, if a hierarchical addressing scheme like E.164 (which is very similar to the current telephone numbering system) is assumed, and then receivers can be grouped into local regions based on area code in an internet protocol.

Network, receivers can be grouped into local regions by using the time-to-live (TTL) field of IP packets. More details on how the TTL field can be used are given in the next section.

b) A multicast tree rooted at the sender S and spanning all the receivers, is set up at the network layer (ATM layer in the context of ATM networks). This is referred to as the global multicast tree in several parts of the paper to distinguish it from the local multicast tree which is a part of the global multicast tree. The global multicast tree is shown by solid lines in Fig 2. Receivers in the local region served by L_i are denoted by Ri,j .Note that L_i denotes the local access switch for the ith region and is not a receiver.

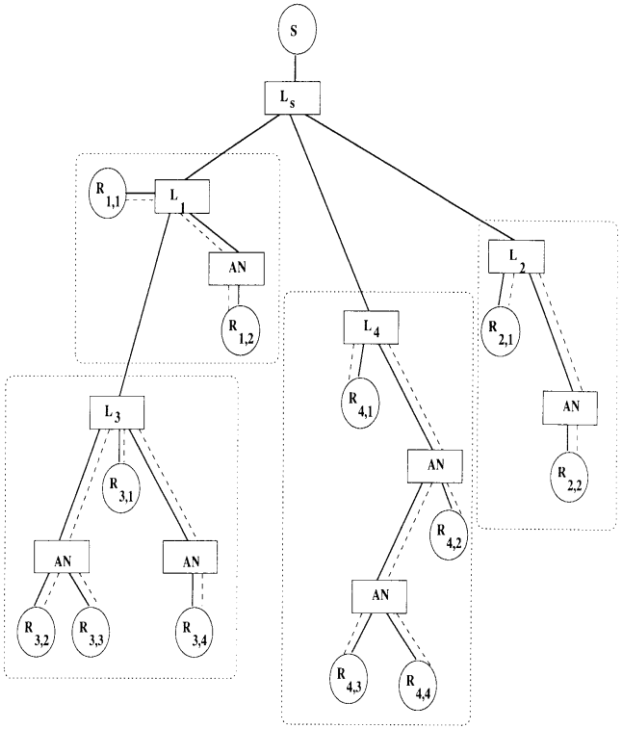


Fig 2:Global multicast tree rooted at S and local multicast trees rooted at $R_{i,j}$'s

c) RMTP is described in this paper as a protocol for point-to-multipoint reliable multicast. Multipoint-to-multipoint reliable multicast is possible if multicast trees are set up for each sender.

2.2 Design Overview

RMTP provides sequenced, lossless delivery of bulk data from one sender to a group of receivers. The sender ensures reliable delivery by selectively retransmitting lost packets in response to the retransmission request of the receivers[9]. If each receiver sends its status (ACK/NACK) all the way to the sender, it results in the throttling of the sender which is the well-known ACK-implosion problem. In addition, if some receivers are located far away from the sender and the sender retransmits lost packets to these distant receivers, the end-to-end delay is significantly increased, and throughput is considerably reduced.

RMTP has been designed to alleviate the ack-implosion problem by using a tree-based hierarchical approach. The key idea in RMTP is to group receivers into local regions and to use a DR as a representative of the local region. Although the sender multicasts every packet to all receivers using the global multicast tree, only the DR's send their own status to the sender indicating which packets they have received and which packets they have not received. The receivers in a local region send their status to the corresponding DR. Note that a DR does not consolidate status messages of the receivers in its local region., but uses these status messages to perform local retransmissions to the receivers, reducing end-to-end delay significantly. Thus the sender sees only the DR's and DR sees only the receivers in its local region. Processing of status messages is distributed among the sender and the DR's, thereby avoiding the ack-implosion problem.

In Fig. 2, receiver Ri,1 is chosen as the DR for the group of Ri,j's, in the local region served by Li. A local multicast tree, rooted at Ri,1, is defined as the portion of the global multicast tree spanning the Ri,j's in the local region served by Li. Local multicast trees are indicated by dashed lines in Fig. 2.

III. IMPLEMENTATION

3.1 RMTP Packet

The sender in RMTP divides the data to be transmitted into fixed-size data packets, with the exception of the last one. A data packet is identified by packet type DATA, while type DATA EOF identifies the last data packet.

Packet Types	
ACK	ACK packet
ACK_TXNOW	ACK - immediate transmission req.
DATA	Data packet
DATA_EOF	Last data packet
RESET	Packet to terminate a connection
RTT_MEASURE	Packet to measure round-trip time
RTT_ACK	ACK to RTT_MEASURE packet
SND_ACK_TOME	Packet for selecting an AP

Table 1: RMTP Packet Types

The sender assigns each data packet a sequence number, starting from zero [5]. A receiver periodically sends ACK packets to the sender/DR. An ACK packet contains the lower end of receive window (L) and a fixed-length bit vector of receive window size indicating which packets are received and which packets are lost. Table 4.1 lists the packet types used in RMTP. Each of their functions will be described in the following subsections

3.2 RMTP Connection

An RMTP connection is identified by a pair of endpoints: a source endpoint and a destination endpoint. The source endpoint consists of the sender's network address and port number; the destination endpoint consists of the multicast group address and a port number. Each RMTP connection has a set of associated connection parameters (see Table 1). RMTP assumes that there is a Session Manager who is responsible for providing the sender and the receiver(s) with the associated connection parameters. RMTP uses default values for any connection parameter that is not explicitly given.

Once the Session Manager has provided the sender and receivers with the session information, receivers initialize the connection control block and remain in an unconnected state; the sender meanwhile starts transmitting data. On receiving a data packet from the sender, a receiver goes from the unconnected state to the connected state. In the

connected state, receivers emit ACK's periodically, keeping the connection alive.

Connection Parameters	
W_r	receive window size in packets
W_s	send window size in packets
T_{dally}	delay after sending the last packet
T_{retx}	time interval to process retx requests
T_{rtt}	time interval to measure RTT
T_{sap}	time interval to send SND_ACK_TOME
T_{send}	time interval to send data packets
T_{ack}	time interval to send status packets
$Packet\_Size$	data packet size in octets
$Cache\_Size$	sender's in-memory data cache size
$CONG_{thresh}$	congestion avoidance threshold
$MCAST_{thresh}$	multicast retransmission threshold

Table3.2: RMTP connection parameters

RMTP is designed based on the IP-multicast philosophy in which the sender does not explicitly know who the receivers are. Receivers may join or leave a multicast session without informing the sender [6]. Therefore the goal in RMTP is to provide reliable delivery to the current members of the multicast session. Since the sender does not keep an explicit list of receivers, termination of RMTP session is timer based. After the sender transmits the last data packet, it starts a timer that expires after Tdally. (ADR also starts the timer when t has correctly received all the data packets.) When the timer expires, the sender deletes all state information associated with the connection (i.e., it deletes the connection's control block). Time interval Tdally is at least twice the lifetime of a packet n an internet. Any ACK from a receiver resets the timer to its initial value. A normal receiver deletes its connection control block and stops emitting ACK's when it has correctly received all data packets. A DR behaves like a normal receiver except that it deletes its connection control block only after the Tdally timer expires.

Since the time period between the transmission of consecutive ACK's from a receiver is much smaller than Tdally , the session Manager is not a part of RMTP transport protocol, but is used at the session layer to manage a given RMTP session.

Sender assumes that either all receivers have received every packet or something "exceptional" has happened. Possible exceptional situations include: network partition and receivers voluntarily or involuntarily leaving the multicast group [8]. RMTP assumes that the Session Manager is responsible for detecting such situations and taking necessary actions.

In addition to normal connection termination, RESET packets can be used to terminate connections. For example, when RMTP detects that the sending application has aborted before data transfer is complete, it uses RESET to inform all the receivers to close the connection.

3.3 RMTP Entities

RMTP has three main entities: 1) Sender, 2) Receiver, and 3) DR. A block diagram description of each of these entities is given in Fig. 3 we describe the major components of these entities below [6].

The Sender entity has a controller component called CONTROLLER, which decides whether the sender should be transmitting new packets(using the Tx component), retransmitting lost packets (using the RTx component), or sending messages advertising itself as an ACK Processor (AP) (using the AP A component and SEND ACK TOME message). There is another component called STATUS PROCESSOR, which processes ACK's (status) from receivers and updates relevant data structures.

Also, note that there are several timer components: TSend, TRetx, and T Sap in the Sender entity, to inform the controller about whether the Tx component, the RTx component or the AP A component should be activated. Timer T Dally is used for terminating a connection.

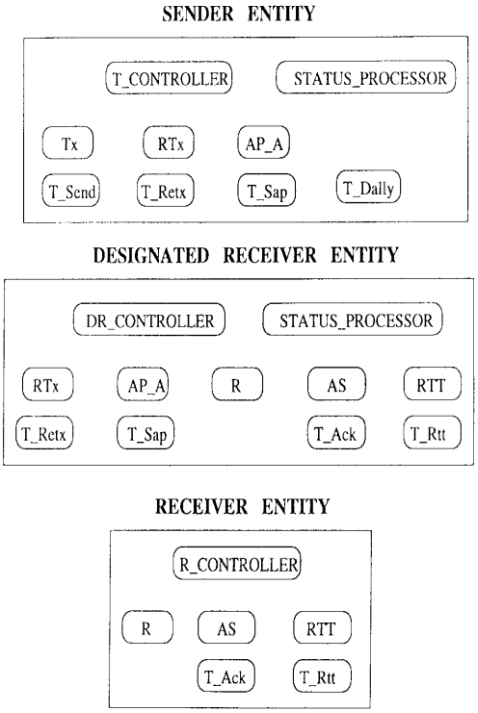


Fig 3: Block diagram of RMTP.

The Receiver entity also has a controller component called RCONTROLLER which decides whether the receiver should be delivering data to the receiving application (using the R component), sending ACK messages (using the AS component), or sending RTT measure packets (using the RTT component) to dynamically compute the round-trip time (RTT) between itself and its corresponding ACK Processor. Note that there are two timer components: 1) T Ack and 2) T Rtt to inform the controller as to whether the AS or the RTT component should be activated. The component R is not timer driven. It is activated asynchronously whenever the receiving application asks for packets.

The DR entity is, in fact, a combination of the Sender entity and the Receiver entity. Key functions

performed by the components of each entity are described next.

3.4. Transmission

RMTP sender (in particular, the Tx component of sender entity in Fig. 4) multicasts data packets at regular intervals defined by a configuration parameter Tsend. The number of packets transmitted during each interval normally depends on the space available in send window. The sender can at most transmit one full window of packets(Ws) during Tsend, thereby limiting the sender's maximum transmission rate to Ws * packet_size/Tsend . To set a multicast session's maximum data transmission rate, the Session Manager simply sets the parameters Ws, packet_size/Tsend , and Tsend accordingly. However, during network congestion, the sender is further limited by the congestion window during the same Tsend interval.

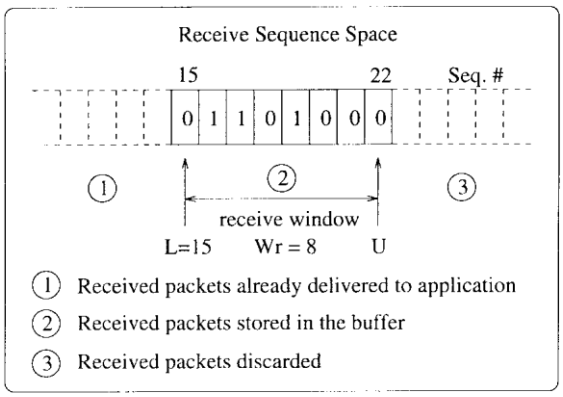


Fig 4: A receiver's receive window and related variables.

3.5 Acknowledgments

RMTP receivers (in particular, the AS component of the receiver/DR entity in Fig. 4.2) send ACK packets periodically, indicating the status of receive window. Receivers use a bit vector of Wr bits (size of receive window) to record the existence of correctly received packets stored in the buffer. As fig. 4 illustrates, each bit corresponds to one packet slot in the receive buffer. Bit 1 indicates a packet slot contains a valid data packet. For example, Fig. 4 shows a receive window of eight packets; packets 16, 17, and 19 are received correctly and stored in the buffer [6]. When a receiver sends an ACK to its AP, it includes the left edge of the receive window and the bit vector. Note that the receiver delivers packets to the application in sequence. For example, if the receiver receives packet 15 from the sender and does not receive packet 18, it can deliver packets 15–17 to the application and advance L to 18.

3.6 Late Joining Receivers

Since RMTP allows receivers to join any time during an ongoing session, a receiver joining late will need to catch up with the rest. In addition, some receivers may temporarily fall behind because of various reasons such as network congestion or even network partition. There are two features in RMTP which together provide the functionality of allowing lagging receivers to catch up with the rest:

1) Immediate transmission request and

2) Data cache in the sender and the DR's.

3.7 Immediate Transmission Request

When a receiver joins late, it receives packets being multicast by the sender at that time, and by looking at the sequence number of those packets, it can immediately find out that it has missed earlier packets.

At that instant, it uses an ACK TXNOW packet to request its AP for immediate transmission of earlier packets. An ACK TXNOW packet differs from an ACK packet only in the packet type field. When an AP receives an ACK TXNOW packet from a receiver R, it checks bit vector V and immediately transmits the missed packet(s) to R using unicast.

3.8 Data Cache

RMTP allows receivers to join an ongoing session at any time and still receive the entire data reliably. However, this flexibility does not come without a price. In order to provide this feature, the senders and the DR's in RMTP need to buffer the entire file during the session. This allows receivers to request for the retransmission of any transmitted data from the corresponding AP. A two-level caching mechanism is used in RMTP. The most recent packets of data are cached in memory, and the rest are stored in disk.

3.9 Flow Control

A simple window-based flow control mechanism is not adequate in a reliable multicast transport protocol in the Internet environment. The main reason is that in the Internet multicast model, receivers can join or leave a multicast session without informing the sender. Thus a sender does not know who the receivers are at any instant during the lifetime of a multicast session [5].

Therefore if we want to design a transport-level protocol to ensure guaranteed delivery of data packets to all the current members of a multicast session, without explicitly knowing the members, a different technique for flow control is needed. Note that if RMTP used a simple window-based flow control mechanism, then the sender would have to know if all the DR's in level 1 have received the packets before the window is advanced. However, the sender may not know how many level 1 DR's are there, because the underlying multicast tree can change and either new DR's may be added to the multicast tree dynamically or old DR's may leave! Fail.

In order to deal with this situation, the sender operates in a cycle. The sender transmits a window full of new packets in the first cycle and in the beginning of the next cycle, it updates the send window and transmits as many new packets as there is room for in its send window. The window update is done as follows. Instead of making sure that each level 1 DR has received the packets, the sender makes sure that all the DR's, that have sent status messages within a given interval of time, have successfully received the relevant packets before advancing the lower end of its send window. Note that

the advancement of send window does not mean that the sender discards the packets outside the window. The packets are still kept in a cache to respond to retransmission requests. In addition, note that the sender never transmits more than a full window of packets during a fixed interval, thereby limiting the maximum transmission rate to Ws * Packet_Size/Tsend. This scheme of flow control can thus be referred to as rate-based windowed flow control.

3.10 Congestion Avoidance

RMTP provides mechanisms to avoid flooding an already congested network with new packets, without making the situation even worse. The scheme used in RMTP for detecting congestion is described below.

RMTP uses retransmission requests from receivers as an indication of possible network congestion [5]. The sender uses a congestion window cong_win to reduce data transmission rate when experiencing congestion. During Tsend, the sender computes the number of ACK's, N, with retransmission request. If exceeds a threshold, CONGtresh, it sets cong_win to one. Since the sender always computes a usable send window as Min(avail_win, cong_win), setting to one reduces data transmission rate to at most one data packet per Tsend if avail_win is nonzero. If N does not exceed CONGthres during Tsend, the sender increases cong_win by one until cong_win reaches Ws. The procedure of setting cong_win to one and linearly increasing cong_win is referred to as slow-start and is used in TCP implementation. The sender begins with a slow-start to wait for the ACK's from far away receivers to arrive.

3.11 Choice Of Dr's And Formation Of Local Regions

RMTP assumes that there is some information about the approximate location of receivers and based on that information, either some receivers or some servers are chosen as DR's. Although specific machines are chosen to act as DR's, the choice of an AP for a given local region is done dynamically. The basic idea is outlined below.

Each DR as well as the sender periodically sends a special packet, called the SEND ACK TOME packet, in which the time-to-live (TTL) field is set to a predetermined value (say 64), using the multicast tree down to each receiver. Thus, if there are several DR's along a given path from the sender to a given receiver, the receiver will receive several SEND ACK TOME packets, one from each DR. However, since the TTL value of an IP datagram gets decremented by one at each hop of the network, the closer a DR is to a given receiver, the higher is the TTL value in the corresponding SEND ACK TOME packet. Therefore, if each receiver chooses the DR, whose SEND ACK TOME packet has the largest TTL value, it will have chosen the DR nearest to it in terms of number of hops. Effectively, a local region will be defined around each DR.

This approach gives us several benefits in terms of robustness and multiple levels of hierarchy. First of all, if the DR, selected by a set of receivers as their AP, fails, then the same set of receivers will choose the DR least upstream from the failed DR, as their new AP [6]. This is because SEND ACK TOME packets from the failed DR will no longer arrive at the receivers and the SEND ACK TOME packet from the DR least upstream from the failed DR will have the largest TTL value. This leads to the dynamic selection of AP for a given set of receivers.

3.12 Multilevel Hierarchy in RMTP

RMTP has been described earlier as a two-tier system in which the sender multicasts to all receivers and DR's; and DR's retransmit lost packets to the receivers in their respective local regions. However, the limitations of a two-level hierarchy are obvious in terms of scalability and a multilevel hierarchy is desirable. The objective of this section is to describe how a multilevel hierarchy is obtained in RMTP with the help of the DR's sending SEND ACK TOME packets.

Recall that each DR periodically sends SEND ACK TOME packets along the multicast tree, and each receiver chooses the DR whose SEND ACK TOME packet has the largest TTL value. Moreover, note that each DR is also a receiver [5]. Therefore, if each DR ignores its own SEND ACK TOME packets, it will choose the DR least upstream from itself as its DR and will send its status messages to that DR during the multicast session. Fig. 5 illustrates the idea.

Effectively, if there are n DR's along a path from the sender to a group of receivers, and these DR's are different hop counts away from the receivers in question, there will be n local regions in an n-level hierarchy, such that the DR of the n th level will send its status to the DR in level n-1, a DR of level n-1 will send its status to the DR in level n-2, and so on, until the DR in level 1 sends its status to the sender (DR at level 0). That is, a DR at the ith level acts as a receiver for the i-1th level for all i, i=n,…., 1 where the zero level refers to the global multicast tree rooted at the sender.

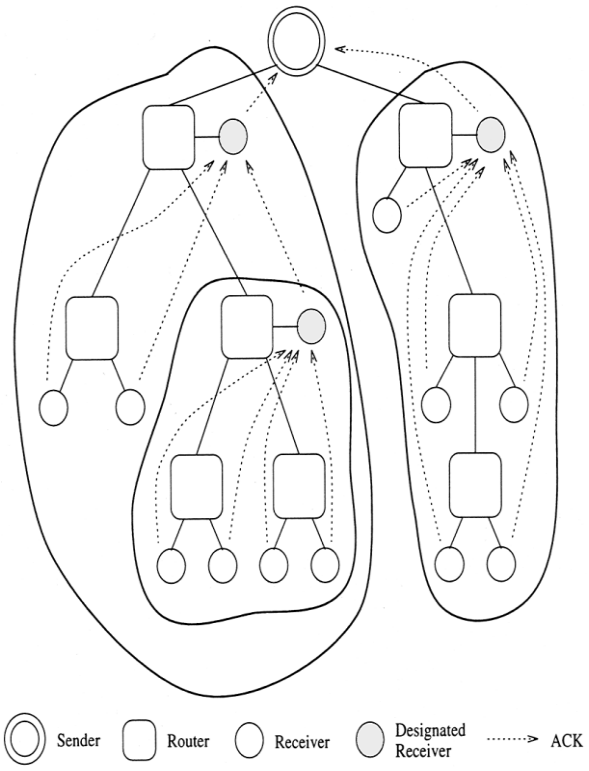


Fig 5: Multilevel Hierarchy of DR's.

IV. Performance Evaluation Routing Metrics

Routing table information used by switching software to select the best route. But how, specifically, are routing table built? What is the specific nature of the information that they content? How routing algorithms determine that one route is preferable to others?

Routing algorithms can use many different metrics to determine the best route. Sophisticated routing algorithms can base route selection on multiple metrics, combining them in single (hybrid) metric. All of the following metric has been used:

Route Acquisition Time

It is the time required to establish route(s) when requested and therefore is of good importance to on demand routing protocols.

Packet Delivery Ratio

The ratio of the data delivered to destination (i.e. throughput) to the data send out by the sources.

Average End-To-End Delay

The average time it takes for packet to reach the destination. It includes all possible delays in the source and each intermediate host, caused by routing discovery, queuing at the interface queue, transmission at the MAC layer, etc. Only successfully delivered packets are counted.

Power Consumption

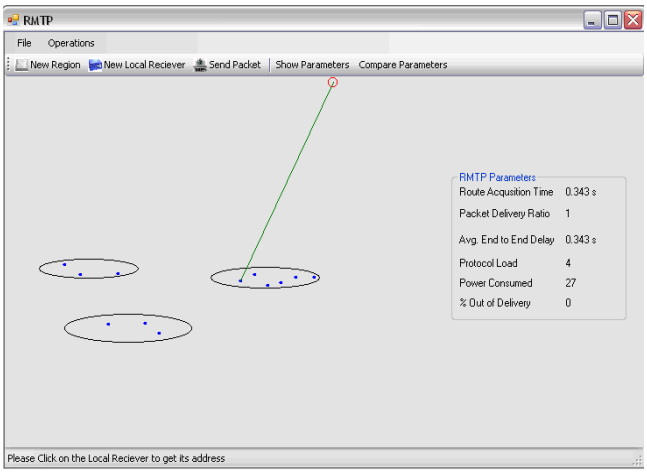
It is the total consumed energy divided by the number of delivered packets. We measure the power consumption because it is one of the precious commodities in mobile communication.

Protocol Load

The routing load per unit data successfully delivered to the destination. The routing load is measure as the number of protocol messages transmitted hop wise (i.e. the transmission on each hop is counted once). A unit data can be a byte or packet.

Percentage Out Of Order Delivery

An external measure of connectionless routing performance of particular interest to transport your protocol such as TCP, which prefer in order delivery.



V. Conclusion

We have presented in this paper the complete design and implementation of RMTP and also provided performance measurements of the actual implementation on the internet. The main contribution of the design include reducing the acknowledgement traffic. The design also include extension of two-level hierarchy to multilevel hierarchy of DR's in the Internet environment. It also includes the use of periodic status messages and the use of selective repeat retransmission mechanism to improve throughput.

The performance figure of RMTP implementation for the data transmission and calculation of parameters is also given in the paper.

References

[1] L.Aguilar, "Datagram routing for Internet multicasting," ACM Comp. Comm. Rev. vol. 14, no 2, pp. 58-63, 1984.

[2] S. Armstron, A. Freier, and K. Marzullo, "RFC-1301 Multicast transport protocol," Feb. 1992.

[3] R. Aiello, E. Pagani, and G. P. Rossi, "Design of a reliable multicast protocol," in proc. IEEE. INFOCOM '93, Mar. 1992, pp. 75-81.

[4] F. Adelstein and M. Singhal, "Real-time causal message ordering in multimedia systems," in Proc. 15th ICDCS '95, June 1995.

[5] J-M. Chang and N. F. Maxemchuk, "Reliable broadcast protocols," ACM Trans. Comp. Syst., vol. 2, no. 3, pp. 251-173, Aug. 1984.

[6] K. P. Birman and T. A. Joseph, "Reliable communication in the presence of failures," ACM Trans. Comp. Syst., vol 5, no. 1, Feb. 1987.

[7] D. Dolev and D. Malki,"The transis approach to high availability clster communication," Comm. ACM, vol. 39, no. 4, pp. 64-70, Apr. 1996.

[8] H. Eriksson, "MBONE: The multicast backbone," Comm. ACM, vol. 37, no. 8, pp. 54-60, Aug. 1994.

[9] J. C. Lin and S. Paul, "RMTP: A reliable multicast transport protocol," in Proc. IEEE INFOCOM '96, Mar. 1996, pp. 1414-1424.

[10] L. E. Moser, P. M. Melliar-Smith, D. A. Agarwal, R. K. Budhia, and C. C. Lingley-Papadopoulos, "Totem: Afault-tolerant multicast group communication system,"Comm. ACM, vol- 39, no. 4, pp. 54-63, Apr. 1996.

[11] S. Paul, K. K. Sabnani, and D. M. Kristol, "Multicast Transport protocols for high speed networks," in Proc. Int. Conf. Network Protocols, 1994, pp. 4-14.

[12] B. Rajagopalan, "Reliability and scaling issues in multicast communication," in Proc. ACM SIGCOMM '92, Sept. 1992, pp. 188-1989.

Modular neural network approach for short term flood forecasting a comparative study

Rahul P. Deshmukh
Indian Institute of Technology, Bombay
Powai, Mumbai
India
deshmukh.rahul@iitb.ac.in

A. A. Ghatol
Former Vice-Chancellor
Dr. Babasaheb Ambedkar Technological University,
Lonere, Raigad, India
vc_2005@rediffmail.com

Abstract— **The artificial neural networks (ANNs) have been applied to various hydrologic problems recently. This research demonstrates static neural approach by applying Modular feedforward neural network to rainfall-runoff modeling for the upper area of Wardha River in India. The model is developed by processing online data over time using static modular neural network modeling. Methodologies and techniques for four models are presented in this paper and a comparison of the short term runoff prediction results between them is also conducted. The prediction results of the Modular feedforward neural network with model two indicate a satisfactory performance in the three hours ahead of time prediction. The conclusions also indicate that Modular feedforward neural network with model two is more versatile than other and can be considered as an alternate and practical tool for predicting short term flood flow.**

Keywords- Artificial neural network, Forecasting, Rainfall, Runoff, Models.

I. INTRODUCTION

The main focus of this research is development of Artificial Neural Network (ANN) models for short term flood forecasting, determining the characteristics of different neural network models. Comparisons are made between the performances of different artificial neural network models of Modular feedforward neural network for optimal result.

The field engineers face the danger of very heavy flow of water through the gates to control the reservoir level by proper operation of gates to achieve the amount of water flowing over the spillway. This can be limited to maximum allowable flood and control flood downstream restricting river channel capacity so as to have safe florid levels in the river within the city limits on the downstream [21].

By keeping the water level in the dam at the optimum level in the monsoon the post monsoon replenishment can be conveniently stored between the full reservoir level and the permissible maximum water level. Flood estimation is very essential and plays a vital role in planning for flood regulation and protection measures.

The total runoff from catchment area depends upon various unknown parameters like Rainfall intensity, Duration of rainfall, Frequency of intense rainfall, Evaporation, Interception, Infiltration, Surface storage, Surface detention, Channel detention, Geological characteristics of drainage basin, Meteorological characteristics of basin, Geographical features of basin etc. Thus it is very difficult to predict runoff at the dam due to the nonlinear and unknown parameters.

In this context, the power of ANNs arises from the capability for constructing complicated indicators (non-linear models). Among several artificial intelligence methods artificial neural networks (ANN) holds a vital role and even ASCE Task Committee Reports have accepted ANNs as an efficient forecasting and modeling tool of complex hydrologic systems[22].

Neural networks are widely regarded as a potentially effective approach for handling large amounts of dynamic, non-linear and noisy data, especially in situations where the underlying physical relationships are not fully understood. Neural networks are also particularly well suited to modeling systems on a real-time basis, and this could greatly benefit operational flood forecasting systems which aim to predict the flood hydrograph for purposes of flood warning and control[16].

Artificial neural netwark are applied for flood forecasting using different models.

Multilayer perceptrons (MLPs) are feedforward neural networks trained with the standard backpropagation algorithm. They are supervised networks so they require a desired response to be trained. They learn how to transform input data into a desired response, and widely used for modeling prediction problems [2].

Backpropagation computes the sensitivity of the output with respect to each weight in the network, and modifies each weight by a value that is proportional to the sensitivity.

Radial basis functions networks have a very strong mathematical foundation rooted in regularization theory for solving ill-conditioned problems. The mapping function of a radial basis function network, is built up of Gaussians rather than sigmoids as in MLP networks [7].

A subset of historical rainfall data from the Wardha River catchment in India was used to build neural network models for real time prediction. Telematic automatic rain

gauging stations are deployed at eight identified strategic locations which transmit the real time rainfall data on hourly basis. At the dam site the ANN model is developed to predict the runoff three hours ahead of time.

In this paper, we demonstrate four different models of Modular feedforward neural network (M FF) models for real time prediction of runoff at the dam and compare the effectiveness of these methods. As the name indicates, the modular feedforward networks are special cases of MLPs, such that layers are segmented into modules. This tends to create some structure within the topology, which will foster specialization of function in each sub-module.

At a time when global climatic change would seem to be increasing the risk of historically unprecedented changes in river regimes, it would appear to be appropriate that alternative representations for flood forecasting should be considered.

II. METHODOLOGY

In this study four methods are employed for rainfall-runoff modeling using Modular feedforward neural network model.

Of the entire learning algorithm, the error backpropagation method is the most widely used. Although this algorithm has been successful in many applications, it has disadvantages such as the long training time that can be inconvenient in practical and on-line applications. This necessitates the improvement of the basic algorithm or integration with other forms of network configurations such as modular networks studied here.

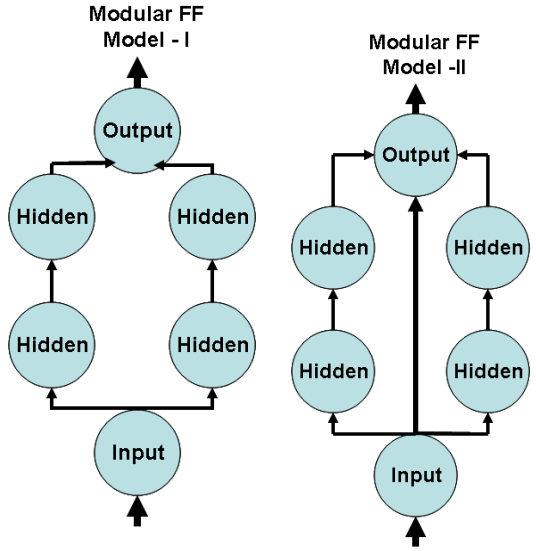


Figure 1. The Modular feedforward model I and II neural network

The modular architecture allows decomposition and assignment of tasks to several modules. Therefore, separate architectures can be developed to each solve a sub-task with the best possible architecture, and the individual modules or building blocks may be combined to form a comprehensive system. The modules decompose the problem into two or more subsystems that operate on inputs without communicating with each other. The input unites are mediated by an integrating unit that is not permitted to feed information back to the module (Jacobs, Jordan, Nowlan, and Hinton, 1991).

Four different models as shown in Figure 1 and Figure 2 are studied. We use two hidden layers, tanh activation function with 0.7 momentum and mean squared error of the cross validation set as stopping criteria which give the optimal results.

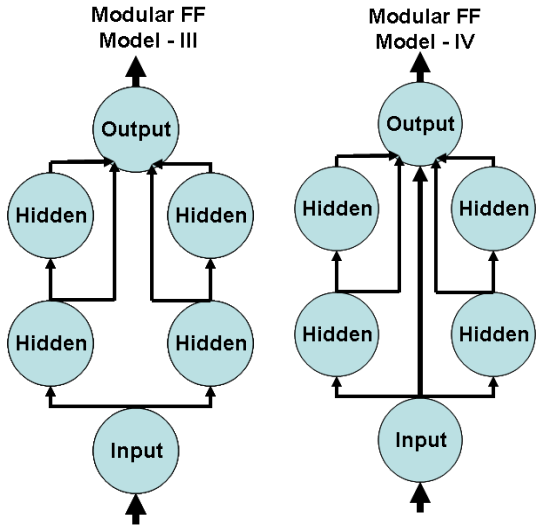


Figure 2. The Modular feedforward model III and IV neural network

Performance Measures:

The learning and generalization ability of the estimated NN model is assessed on the basis of important performance measures such as MSE (Mean Square Error), NMSE (Normalized Mean Square Error) and r (Correlation coefficient)

A. *MSE (Mean Squzre Error) :*

The formula for the mean square error is:

$$MSE = \frac{\sum_{j=0}^{P}\sum_{i=0}^{N}\left(d_{ij} - y_{ij}\right)^2}{NP} \quad \ldots \quad (1)$$

Where

P = number of output PEs,

N = number of exemplars in the data set,

y_{ij} = network output for exemplar i at PE j,

d_{ij} = desired output for exemplar i at PE j.

B. NMSE (Normalized Mean Square Error:

The normalized mean squared error is defined by the following formula:

$$NMSE = \frac{P\,N\,MSE}{\sum_{j=0}^{P} \frac{N\sum_{i=0}^{N} d_{ij}^{\,2} - \left(\sum_{i=0}^{N} d_{ij}\right)^2}{N}} \quad \ldots (2)$$

Where

P = number of output processing elements,

N = number of exemplars in the data set,

MSE = mean square error,

d_{ij} = desired output for exemplar i at processing element j.

C. r (correlation coefficient) :

The size of the mean square error (MSE) can be used to determine how well the network output fits the desired output, but it doesn't necessarily reflect whether the two sets of data move in the same direction. For instance, by simply scaling the network output, the MSE can be changed without changing the directionality of the data. The correlation coefficient (r) solves this problem. By definition, the correlation coefficient between a network output x and a desired output d is:

$$r = \frac{\frac{\sum_i \left(x_i - \bar{x}\right)\left(d_i - \bar{d}\right)}{N}}{\sqrt{\frac{\sum_i \left(d_i - \bar{d}\right)^2}{N}}\sqrt{\frac{\sum_i \left(x_i - \bar{x}\right)^2}{N}}} \quad \ldots (3)$$

where $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ and $\bar{d} = \frac{1}{N}\sum_{i=1}^{N} d_i$

The correlation coefficient is confined to the range [-1, 1]. When r = 1 there is a perfect positive linear correlation between x and d, that is, they co-vary, which means that they vary by the same amount.

III. Study Area and Data Set

The Upper Wardha catchment area lies directly in the path of depression movements which originates in the Bay of Bengal. When the low pressure area is formed in the Bay of Bengal and cyclone moves in North West directions, many times this catchment receives very heavy intense cyclonic precipitation for a day or two. Occurrence of such events have been observed in the months of August and September. Rainfall is so intense that immediately flash runoff, causing heavy flood has been very common feature in this catchment.

For such flashy type of catchment and wide variety in topography, runoff at dam is still complicated to predict. The conventional methods also display chaotic result. Thus ANN based model is built to predict the total runoff from rainfall in Upper Wardha catchment area for controlling water level of the dam.

In the initial reaches, near its origin catchment area is hilly and covered with forest. The latter portion of the river lies almost in plain with wide valleys.

The catchment area up to dam site is 4302 sq. km. At dam site the river has wide fan shaped catchment area which has large variation with respect to slope, soil and vegetation cover.

Figure 3- Location of Upper Wardha dam on Indian map

Data: Rainfall runoff data for this study is taken from the Wardha river catchment area which contains a mix of urban and rural land. The catchments is evenly distributed in eight zones based on the amount of rainfall and geographical survey. The model is developed using historical rainfall runoff data , provided by Upper Wardha Dam Division Amravati, department of irrigation Govt. of Maharashtra. Network is trained by rainfall information gathered from eight telemetric rain-gauge stations distributed evenly throughout the catchment area and runoff at the dam site.

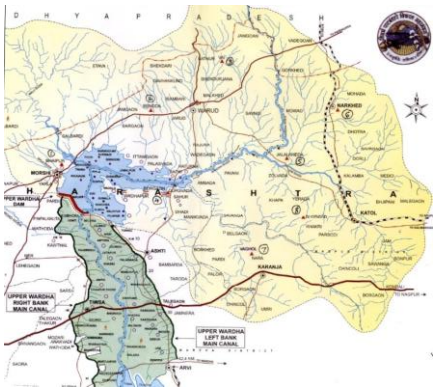


Figure 4- The Wardha river catchment

The data is received at the central control room online through this system on hourly basis. The Upper Wardha dam reservoir operations are also fully automated. The amount of inflow, amount of discharge is also recorded on hourly basis. From the inflow and discharge data the cumulative inflow is calculated. The following features are identified for the modeling the neural network .

TABLE I. THE PARAMETERS USED FOR TRAINING THE NETWORK

M onth	R G1	R G2	R G3	R G4	R G5	R G6	R G7	R G8	C IF

- Month – The month of rainfall
- Rain1 to Rain8 – Eight rain gauging stations.
- Cum Inflow – Cumulative inflow in dam

Seven years of data on hourly basis from 2001 to 2007 is used. It has been found that major rain fall (90%) occurs in the month of June to October Mostly all other months are dry hence data from five months. June to October is used to train the network

IV. RESULT

The neural network structure is employed to learn the unknown characterization of the system from the dataset presented to it. The dataset is partitioned into three categories, namely training, cross validation and test. The idea behind this is that the estimated NN model should be tested against the dataset that was never presented to it before. This is necessary to ensure the generalization. An experiment is performed at least twenty five times with different random initializations of the connection weights in order to improve generalization.

The data set is divided in to training , testing and cross validation data and the network is trained for all models of Modular feedforward neural network model for 5000 epochs. Fig 5 to Fig 8 shows the plot of actual Vs predicted values for runoff for Modular feedforward neural network models.

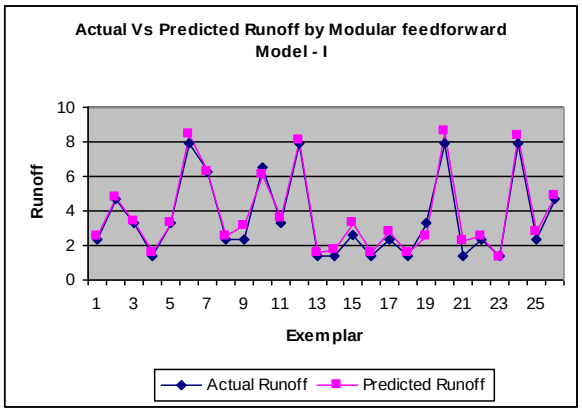


Figure 5- Actual Vs. Predicted runoff by MFF M-I

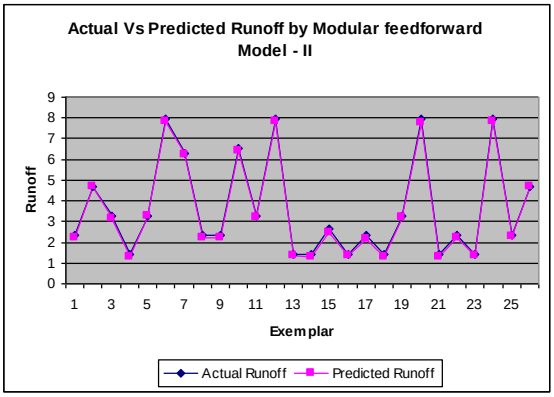


Figure 6.– Actual Vs. Predicted runoff by M FF M-II

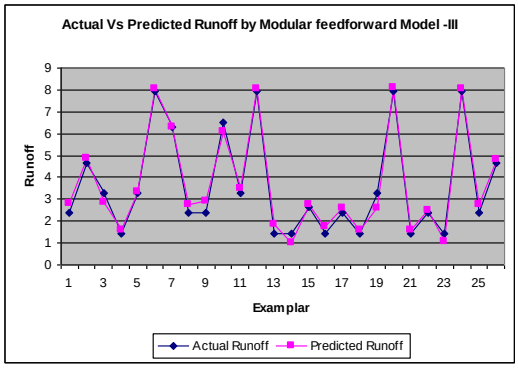


Figure 7.– Actual Vs. Predicted runoff by M FF M-III

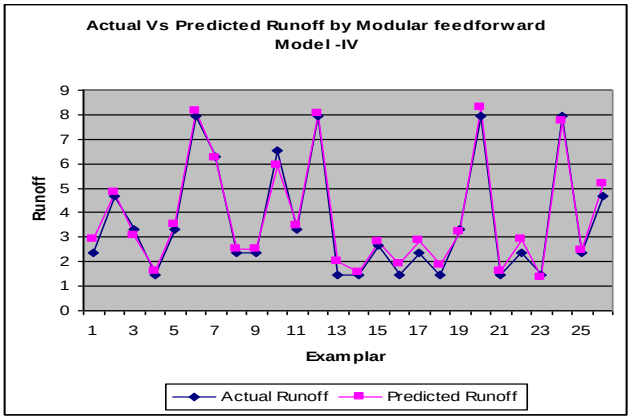


Figure 8.– Actual Vs. Predicted runoff by M FF M-IV

The error found in the actual and predicted runoff at the dam site is plotted for all four models of M FF neural networks as shown in the Figure 9 to Figure 12.

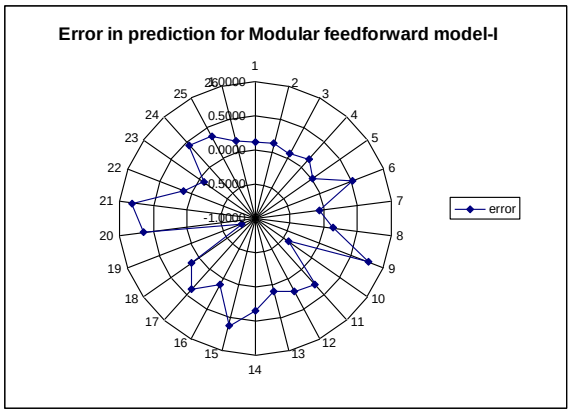


Fig 9 – Error graph of MLP Model

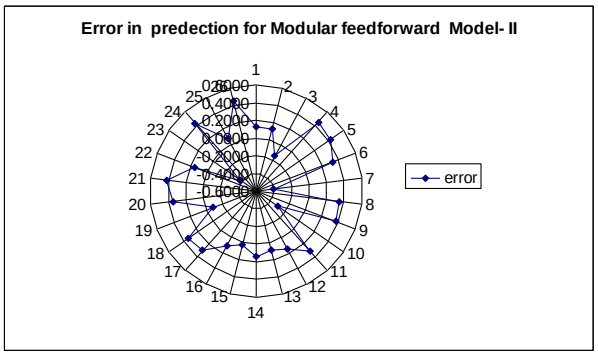


Fig 10 – Error graph of M FF Model

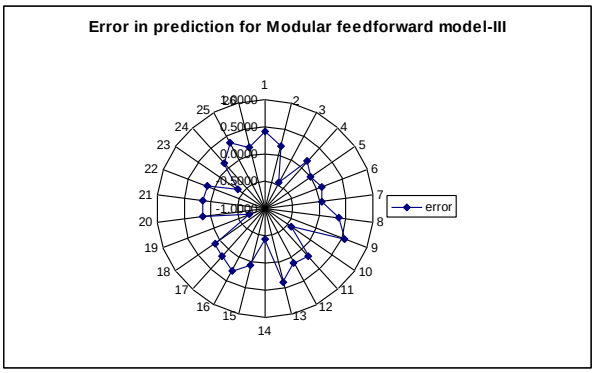


Fig 11 – Error graph of M FF Model

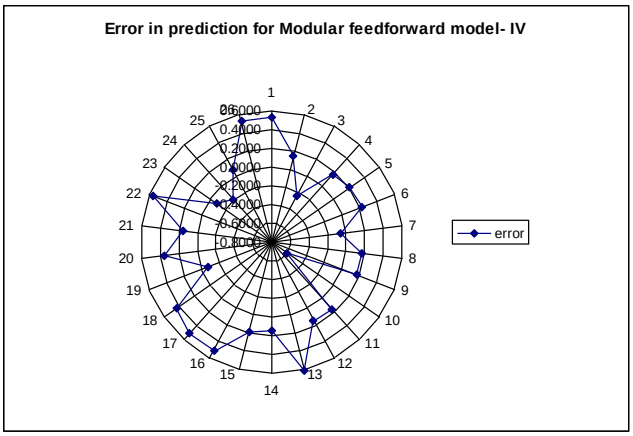


Fig 12 – Error graph of M FF Model

After training the network the performance is studied and in the Table-2 to Table-5 the parameters and the performances of four different models of Modular feedforward neural network are listed.

TABLE II. M FF M-I NETWORK PERFORMANCE

Parameter	Performance- M-I
MSE	0.1547
NMSE	0.1436
Min Abs Error	0.0652
Max Abs Error	0.8193
r	0.4619

TABLE III. M FF M-II NETWORK PERFORMANCE

Parameter	Performance-M-II
MSE	0.0872
NMSE	0.0764
Min Abs Error	0.0246
Max Abs Error	0.4601
r	0.8106

TABLE IV. MFF M-III NETWORK PERFORMANCE

Parameter	Performance-M-III
MSE	0.0968
NMSE	0.1196
Min Abs Error	0.0342
Max Abs Error	0.6904
r	0.7342

TABLE V. M FF M-IV NETWORK PERFORMANCE

Parameter	Performance-M-IV
MSE	0.1068
NMSE	0.1263
Min Abs Error	0.0519
Max Abs Error	0.6036
r	0.6184

The parameters and performance for all four models of M FF model are compared on the performance scale and are listed in the Table 6 shown below. The comparative analysis of the MSE, NMSE and r (the correlation coefficient) is done.

TABLE VI. COMPARISON OF PERFORMANCE PARAMETERS

TABLE VII.

Module / Parameter	M-I	M-II	M-III	M-IV
MSE	0.1547	0.0872	0.0968	0.1068
NMSE	0.1436	0.0764	0.1196	0.1263
Min Abs Error	0.0652	0.0246	0.0342	0.0519
Max Abs Error	0.8193	0.4601	0.6904	0.6036
r	0.4619	0.8106	.7342	0.6184

The main advantage of M FF is that in contrast to the MLP, modular feedforward networks do not have full interconnectivity between the layers. Therefore, a smaller number of weights are required for the same size network (the same number of PEs). This tends to speed the training and reduce the number of examples needed to train the network to the same degree of accuracy.

V. CONCLUSION

An ANN-based short-term runoff forecasting system is developed in this work. A comparison between four different models of Modular feedforward neural network model is made to investigate the performance of four distinct approaches. We find that Modular feedforward neural network with model-II approach is more versatile than others. Modular feedforward neural network with module-II is performing better as compare to other approaches studied as far as the overall performance is concerned for forecasting runoff for 3 hrs lead time. Other models of Modular feedforward neural network are also performing optimally. Which means that static model of Modular feedforward neural network with model-II is powerful tool for short term runoff forecasting for Wardha River basin

ACKNOWLEDGMENT

This study is supported by Upper Wardha Dam Division Amravati, department of irrigation Govt. of Maharashtra, India.

REFERENCES

[1] P. Srivastava, J. N. McVair, and T. E. Johnson, "Comparison of process-based and artificial neural network approaches for streamflow modeling in an agricultural watershed," Journal of the American Water Resources Association, vol. 42, pp. 545563, Jun 2006.

[2] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," Neural Netw., vol. 2, pp. 359-366,1989.

[3] M. C. Demirel, A. Venancio, and E. Kahya, "Flow forecast by SWAT model and ANN in Pracana basin, Portugal," Advances in Engineering Software, vol. 40, pp. 467-473, Jul 2009.

[4] A. S. Tokar and M. Markus, "Precipitation-Runoff Modeling Using Artificial Neural Networks and Conceptual Models," Journal of Hydrologic Engineering, vol. 5, pp. 156-161,2000.

[5] S. Q. Zhou, X. Liang, J. Chen, and P. Gong, "An assessment of the VIC-3L hydrological model for the Yangtze River basin based on remote sensing: a case study of the Baohe River basin," Canadian Journal of Remote Sensing, vol. 30, pp. 840-853, Oct 2004.

[6] R. J. Zhao, "The Xinanjiang Model," in Hydrological Forecasting Proceedings Oxford Symposium, lASH, Oxford, 1980 pp. 351-356.

[7] R. J. Zhao, "The Xinanjiang Model Applied in China," Journal of Hydrology, vol. 135, pp. 371-381, Ju11992.

[8] D. Zhang and Z. Wanchang, "Distributed hydrological modeling study with the dynamic water yielding mechanism and RS/GIS techniques," in Proc. of SPIE, 2006, pp. 63591Ml-12.

[9] J. E. Nash and I. V. Sutcliffe, "River flow forecasting through conceptual models," Journal ofHydrology, vol. 273, pp. 282290,1970.

[10] D. Zhang, "Study of Distributed Hydrological Model with the Dynamic Integration of Infiltration Excess and Saturated Excess Water Yielding Mechanism." vol. Doctor Nanjing: Nanjing University, 2006, p. 190.529

[11] E. Kahya and J. A. Dracup, "U.S. Streamflow Patterns in Relation to the EI Nit'lo/Southern Oscillation," Water Resour. Res., vol. 29, pp. 2491-2503 ,1993.

[12] K. J. Beven and M. J. Kirkby, "A physically based variable contributing area model of basin hydrology," Hydrologi cal Science Bulletin, vol. 43, pp. 43-69,1979.

[13] N. J. de Vos, T. H. M. Rientjes, "Constraints of artificial neural networks for rainfall-runoff modelling: trade-offs in hydrological state representation and model evaluation", Hydrology and Earth System Sciences, European Geosciences Union, 2005, 9, pp. 111-126.

[14] Holger R. Maier, Graeme C. Dandy, "Neural networks for the perdiction and forecasting of water resources variables: a review of modeling issues and applications", Environmental Modelling & Software, ELSEVIER, 2000, 15, pp. 101-124.

[15] T. Hu, P. Yuan, etc. "Applications of artificial neural network to hydrology and water resources", Advances in Water Science, NHRI, 1995, 1, pp. 76-82.

[16] Q. Ju, Z. Hao, etc. "Hydrologic simulations with artificial neural networks", Proceedings-Third International Conference on Natural Computation, ICNC, 2007, pp. 22-27.

[17] G. WANG, M. ZHOU, etc. "Improved version of BTOPMC model and its application in event-based hydrologic simulations", Journal of Geographical Sciences, Springer, 2007, 2, pp. 73-84.

[18] K. Beven, M. Kirkby, "A physically based, variable contributing area model of basin hydrology", Hydrological Sciences Bulletin, Springer, 1979, 1, pp.43-69.

[19] K. Thirumalaiah, and C.D. Makarand, Hydrological Forecasting Using Neural Networks Journal of Hydrologic Engineering. Vol. 5, pp. 180-189, 2000.

[20] G. WANG, M. ZHOU, etc. "Improved version of BTOPMC model and its application in event-based hydrologic simulations", Journal of Geographical Sciences, Springer, 2007, 2, pp. 73-84.

[21] H. Goto, Y. Hasegawa, and M. Tanaka, "Efficient Scheduling Focusing on the Duality of MPL Representatives," Proc. IEEE Symp. Computational Intelligence in Scheduling (SCIS 07), IEEE Press, Dec. 2007, pp. 57-64.

[22] ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, "Artificial neural networks in hydrology I: preliminary concepts", Journal of Hydrologic Engineering, 5(2), pp.115-123, 2000.

AUTHORS PROFILE

Rahul Deshmukh received the B.E. and M.E. degrees in Electronics Engineering from Amravati University. During 1996-2007, he stayed in Government College of Engineering, Amravati in department of Electronics and telecommunication teaching undergraduate and postgraduate students. From 2007 till now he is with Indian Institute of Technology (IIT) Bombay, Mumbai. His area of reserch are artificial intelligence and neural networks.

A. A. Ghatol received the B.E. from Nagpur university foallowed by M. Tech and P.hd. from IIT Bombay. He is best teacher award recipient of government of Maharastra state. He has worked as director of College of Engineering Poona and Vice-Chancellor, Dr. Babasaheb Ambedkar Technological University, Lonere, Raigad, India. His area of research is artificial intelligence, neural networks and semiconductors

A suitable segmentation methodology based on pixel similarities for landmine detection in IR images

Dr. G.Padmavathi[1] ,
Department of Computer Science, Avinashilingam University for Women ,Coimbatore, India
ganapathi.padmavathi@gmail.com

Dr. P. Subashini[2] ,
Department of Computer Science, Avinashilingam University for Women ,Coimbatore, India
mail.p.subashini@gmail.com

Ms. M. Krishnaveni[3]
Department of Computer Science, Avinashilingam University for Women ,Coimbatore, India
krishnaveni.rd@gmail.com

***Abstract*— Identification of masked objects especially in detection of landmines is always a difficult problem due to environmental inference. Here, segmentation phase is highly concentrated by performing an initial spatial segmentation to achieve a minimal number of segmented regions while preserving the homogeneity criteria of each region. This paper aims in evaluating similarities based segmentation methods to compose the partition of objects in Infra-Red images. The output is a set of non-overlapping homogenous regions that compose the pixels of the image. These extracted regions are used as the initial data structure in feature extraction process. Experimental results conclude that h-maxima transformation provides better results for landmine detection by taking the advantage of the threshold. The relative performance of different conventional methods and proposed method are evaluated and compared using the Global Consistency Error and Structural Content. It proves that h-maxima gives significant results that definitely facilitate the landmine classification system more effectively.**

Keywords- Segmentation, Global Consistency error, h-maxima, threshold, Landmine detection

I. INTRODUCTION

The signature of buried land mine in IR images varies significantly depending on external parameters such as weather, soil moisture, solar radiation, burial depth, and time[7]. By literature [Oscar González Merino] [11] the working of many image-based landmine detection algorithms, it is concluded that the fundamental challenges arise from the fact that the mean spectral signatures of disturbed soil areas that indicate mine presence are nearly always very similar to the signatures of mixed background pixels that naturally occur in heterogeneous scenes composed of various types of soil and vegetation [1][12]. Mine detection using infrared techniques is primarily based on exploiting temperature differences between pixels on the mines and background pixels [2]. Thus, there is always a need for robust algorithm that has the capability to analyze the pattern of distribution of the pixels to separate pixels of the mines from background pixels. Here the segmentation methods are evaluated with performance metrics [3]. Experimental results show that h-maxima transformation is more adoptable for IR target images. The paper is organized as follows: Section 2 deals with the need for pixel based segmentation for IR images. Section 3 converses the comparison of segmentation methods and the subjective assessment of the approach Section 4 explores the performance evaluation of the methods. The paper ends with observations on future work and some conclusions.

II. PRELIMINARIES OF SEGMENTATION

There are many unsupervised and supervised segmentation algorithms [6]. They only use low-level features, e.g. intensity and texture, to generate homogeneous patches from an input image. Four categories for segmentation are: histogram shape based methods, where, for example, the peaks, valleys and curvatures of the smoothed histogram[4]. Clustering based methods, where the gray level samples are clustered in two parts as background and foreground (object), or alternately are modeled as a mixture of two Gaussians. Entropy based methods results in algorithms that use the entropy of the foreground and background regions, the cross entropy between the original and binarized image, etc. Object attribute-based methods search a measure of similarity between the gray level and the binarized images, such as fuzzy shape similarity, edge coincide, etc. This paper quantifies the segmentation methods based on the similarities of the pixels including the intensity and the object structure.

III. SIMILARITIES BASED SEGMENTATION TECHNIQUES

Segmentation is a pre-process which partitioned image into unique multiple regions, where region is set of pixels. Mathematically segmentation can be defined as follows:

If I is set of all image pixels, then by applying segmentation we get different unique regions like {S1, S2, S3,…,Sn } which when combined formed 'I' . Basic formulation is as follows:

(a) $\bigcup_{i=1,n}^{n} S_i = I \quad where \quad S_i \bigcap S_j = \varnothing$

(b) Si is a connected region, i=1, 2….n.

(c) P(S i) = TRUE for i=1, 2... n.

(d) $P\left(S_i \bigcup S_j\right) = FALSE \quad for i \neq j.$

Where P (Si) is a logical predicate defined over the points in set Si.

Condition (a) indicates that segmentation must be complete, every pixel in the image must be covered by segmented regions. Segmented regions must be disjoint. Condition (b) requires that points in a region be connected in some predefined sense like 4- or 8- connected. Condition (c) deals, the properties must be satisfied by the pixels in a segmented region- e.g. $P(S_i)$ = TRUE. if all pixels in S_i have the same gray level. Last condition (d) indicates that adjacent regions S_i and S_j are different in the sense of predicate P.

Ever in image processing research there is no common solution to the segmentation problem [6]. One of the main reasons of segmentation algorithms is to precisely segment the image without under or over segmentation. Almost all image segmentation techniques proposed so far are ad hoc in nature. These below are the following approaches of image segmentation taken in this paper and demonstrated with IR images. Given below in fig 1 is the approach taken for image segmentation for IR images.

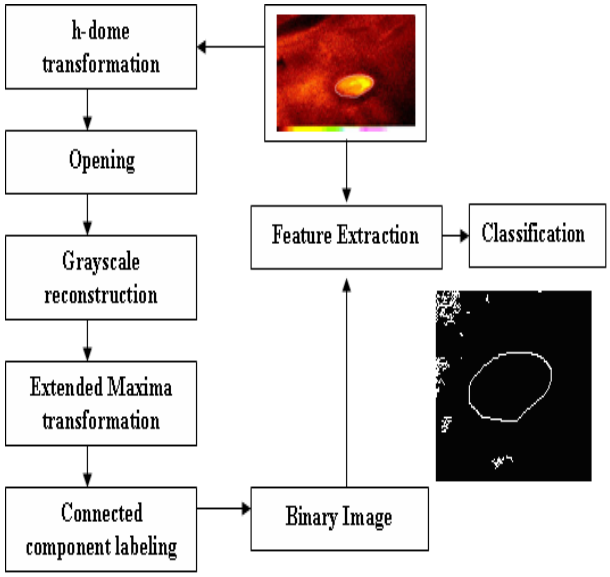


Figure 1: Schematic representation of proposed algorithm for identification of landmine

A. H-maxima algorithm

Every IR image taken has been implemented with morphological reconstruction, extended maxima transformation using thresholding[10] . The extended maxima transformation is the regional maxima computation of the corresponding *h*-maxima transformation. As a result, it produces a binary image. A connected-component labeling operation is performed, in order to evaluate the characteristics and the location of every object. As a second object reduction step, objects not located within a region of another object, are also discarded, since mine objects are not typically clustered. The region of interest (ie) the target is got by connected component segmentation in which the relevant pixels of the object will be grouped and extorted [8]. The extended-maxima transform computes the regional maxima of the H-maxima transform. Here H refers to nonnegative scalar. Regional maxima are connected components of pixels with a constant intensity value, and whose external boundary pixels will have a lower value.

B. Kmeans algorithm

Here K means is used as a two phase iterative algorithm[9] to minimize the sum of point-to centroid distances, summed over all k clusters. The first phase uses each iteration that consists of reassigning points to their nearest cluster centroid. The second phase uses points that are individually reassigned. Arbitrarily choose k data points to act as cluster centers. Until the cluster centers are unchanged following are the steps carried out : Allocate each data -point to cluster whose center is nearest. Replace the cluster centers with the mean of the elements in their clusters end. Clustering in attribute space can lead to unconnected regions in image space (but this may be useful for handling occlusions).The K means image representation groups all feature vectors from all images into K clusters and provides a cluster id for every region of image that represents the salient properties of the region. K means is fast iterative and leads to a local minimum. It looks for unusual reduction in variance. This iterative algorithm has two steps

Assignment step: Assign each observation to the cluster with the closest mean

$$S_i^{(t)} = \{X_j : \left\| X_j - m_i^{(t)} \right\| \leq \left\| x_j - m^{(t)} \right\| \text{ --------- (1)}$$

Update step: Calculate the new means to be centroid of the observations in the cluster

$$m_i^{(t+1)} = \frac{1}{\left| S_i^{(t)} \right|} \sum_{x_j \in S(t)_i} X_j \text{ -}$$

-(2)

C. Threshold intensity distribution algorithm

It can detect object boundaries with low gradient or reduce noise effect in gradient. However, an accurate and stable estimation of intensity distribution is difficult to get from a finite set of 3D image data. To reduce the ”shrink” or ”expand” effect on segmentation results, gradient information is used to calibrate the estimation of intensity distribution in the following. Overlap of image gradient is computed with the boundaries determined by intensity distribution through introducing a probability offset to intensity distribution. The maximum overlap indicates the optimal boundaries of the interested objects. To restate the problem without losing generality, here it use a mixed Gaussian distribution model.

$$P(u) = \sum_{k=1}^{n} \pi_k P(u \left| \lambda_k ; \mu_k, \sigma_k \right.),$$

Where π_k is the prior probability of class λ_k with $\sum_{k=1}^{n} \pi_k = 1$, and μ_k, σ_k are the mean and variance of the Gaussian distribution of the intensity. Intensity distribution inside the region Ω is in Eq below

$$P_{in}(u) = \sum_{k|\lambda_k \in \Omega} \pi_k P\left(u \middle| \lambda_k; \mu_k, \sigma_k\right)$$

And the intensity distribution of the outside region Ω, P_{out} , can be obtained in a similar way. Normally for pixel x on region boundaries with u = I(x),

$$P_{out}\left(I(x)\right) - P_{in}\left(I(x)\right) = 0$$

Threshold Intensity Distribution is a significant aspect in assigning an individual region around an image[4]. Uneven distribution may lead to assignment of two or more regions to an individual pixel. This method reduces processing time by performing gray-level based segmentation that extracts regions of uniform intensity. Subsequently, it is also possible to estimate motion for the regions. It also reduces the computational load, and the region-based estimator gives robustness to noise and changes of illumination. The segmentation of the reference image is designed to group pixels of similar gray-levels. In intensity distribution, B (i,j) is a binary image (pixel are either 0 or 1) created by thresholding F(i,j)

B(i,j) = 1 if F (i,j)<t
B (i,j) =0 if F(I,j)>=t

It is assumed that the 1's are the object pixels and the 0's are the background pixels.

The Histogram (h) - gray level frequency distribution of the gray level image F.

h_F (g) = number of pixels in F whose gray level is g
H_F (g) =number of pixels in F whose gray level is <=g

This method is a probabilistic method that makes parametric assumptions about object and background intensity distributions and then derives "optimal" thresholds.

D. Boundary based algorithm

Here the strategy is to consider each control point in turn and move it to the pixel; in its local neighbourhood which gives us the minimum. For a closed boundary it could make the initial estimate surround the object of interest, and add in another term to the objective function to penalize the total length. A difficulty with this type of strategy is the control points. This method traces the exterior boundaries of objects, as well as boundaries of holes inside these objects, in the binary image. A binary image is considered in which the nonzero pixels belong to an object and 0 pixels constitute the background.

Rules that are worked out to guide the process of true corner point localization.

(i) Select those boundary points which bear significantly large cornerity index by eliminating the boundary points which lie on straight line segments bearing negligibly small cornerity index value.
(ii) Since all points on a smooth curve segment are in general associated with almost same cornerity index, and actual corner points bear cornerity index larger than that of their neighbors, it is suggested to select the set of connected points such that the variations in their cornerity indices are considerably large. This rule helps in selecting only the set of points with in the vicinity of actual corner points by eliminating the points on smooth curve segments.
(iii) Select the points which bear local maximum cornerity index as true corner points. It could be noticed that these rules do not require any priori knowledge in locating true corner points.

Thus, the expected point corresponding to a corner point will have a larger shift when compared to other points on the boundary curve. Therefore, the cornerity index of pi is defined to be the Euclidean distance d between the points pi and its expected point pie and is given by

$$d = \sqrt{\left(x_i - x_{ie}\right)^2 + \left(y_i - y_{ie}\right)^2}$$

The cornerity index indicates the prominence of a corner point. The larger the value of the cornerity index of a boundary point, the stronger is the evidence that the boundary point is a corner.

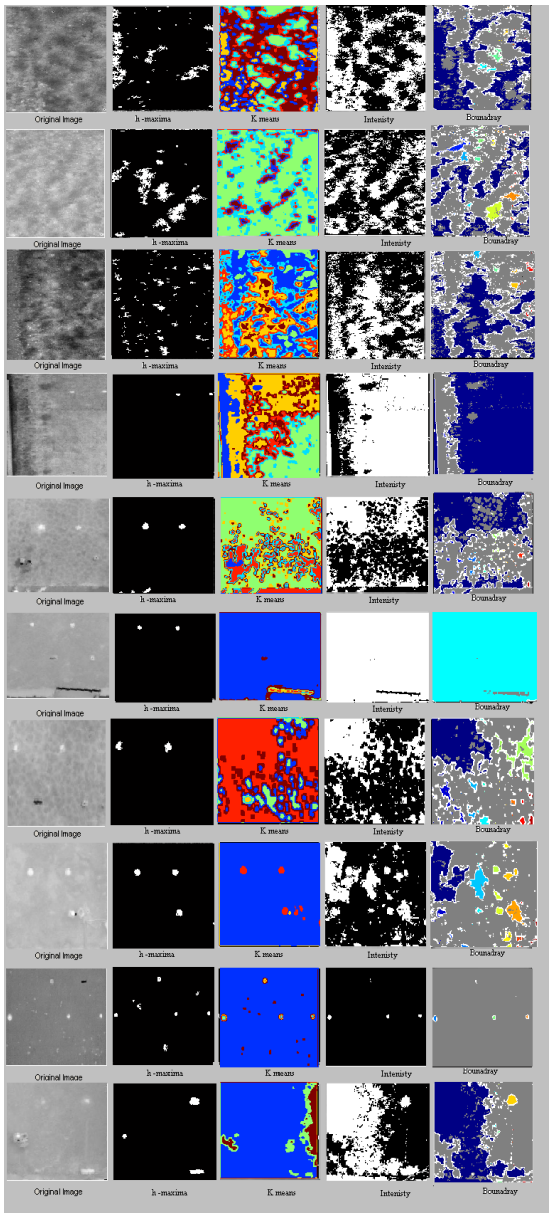


Figure 2: Visual assessment of the segmentation methods

IV. PERFORMANCE EVALUATION

The qualitative and quantitative assessment of segmentation results is carried out here for choosing the appropriate approach for a given segmentation task. Similar to the segmentation theory itself, there is no established standard procedure for the evaluation of its results. For this reason evaluation is done using empirical discrepancy method using the relative ultimate measurement accuracy. Global Consistency Error and Structural Content used as the evaluation parameters to control the segmentation process and dynamically a good number of regions are chosen based on local minima in the segmentation evaluation measure[5] .The uniqueness of each parameter takes the advantage and disadvantages of IR images.

Global Consistency Error (GCE) forces all local refinements to be in the same direction and is defined as:

$$GCE(S,S') = \frac{1}{N}\min\left\{\sum LRE(S,S',x_i), \sum LRE(S',S,x_i)\right\} \quad (3)$$

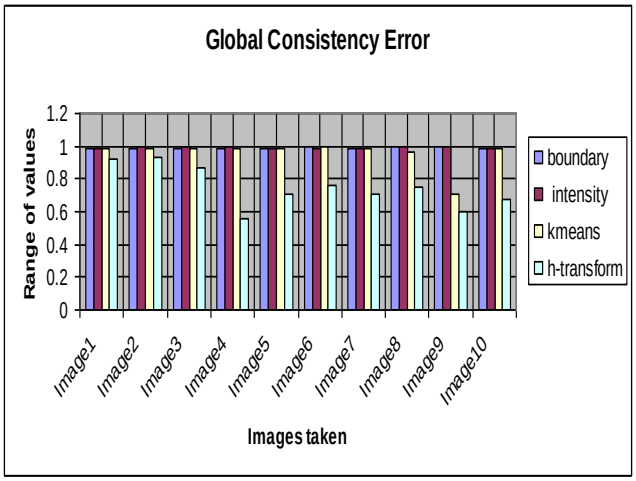


Figure 3: Comparison based on Global Consistency Error

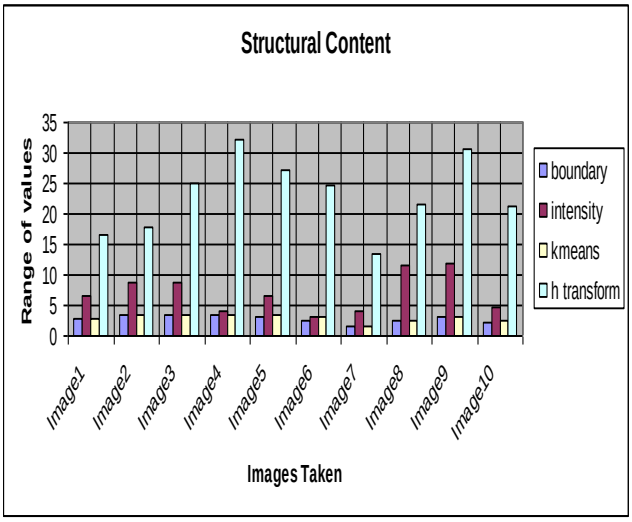


Figure 4: Comparison based on structural content

A structural content is used to find mask structural defects and noise, but are prone to periodic artifacts. This examines to classify image regions hierarchically based on the level of structural content.

$$SC = \sum_{j=1}^{M}\sum_{k=1}^{N} x_{j,k^2}\,2 / \sum_{j=1}^{M}\sum_{k=1}^{N} x'_{j,k^2}\,2 / \ldots\ldots\ldots(4)$$

Structural content of a segmented image is calculated by summation of the original image by the division of the segmented image.

From the above figures 3, 4 the objective evaluation also states that h maxima transform is the desirable method for landmine detection. The Global consistency error rate should always be low and it is the same for the proposed method in comparatively to the conventional methods. The structural content is high for h-maxima which state that the originality of the object and its information remains the same after the segmentation process.

V. CONCLUSION

The method described in this paper provides a relatively simple, extremely fast, and robust method for displaying and performing automatic target identification phase. The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image. Each of the pixels in a region is similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristic(s). From this paper it is concluded that h-maxima is very useful in IR image evaluation but that it should probably restrict studies to similar images and similar processing. Simulations are carried out which demonstrates the proposed method is able to successfully localize landmine objects from different sets of real IR images. Nevertheless, this scheme has some limitations because it is not automatic as different parameters have to be adjusted manually. Future work should incorporate the use of high-level image analysis methods for the identification of the true mine objects among the set of the detected mine cues.

REFERENCES

[1] Frank cremer't, wim de jong and klamer schutte ," processing of polarimetric infrared images for landmine detection",2nd international workshop on advanced gpr. 14-16 may, 2003, the Netherlands

[2] E Ctrmn, W. de long, and K Schutte, "Fusion of polarimetric infrared features and GPR features far landmine detection," in 2nd International Workshop on Advanced Ground Penetrating Radar (IWAGPR), Delft, The Netherlands, May 2003.

[3] Nathir A. Rawashdeh, Shaun T. Love " Hierarchical Image Segmentation by Structural Content" journal of software, vol. 3, no. 2, february 2008

[4] Yongsheng Hu, Qian Chen,"An Infrared Image Preprocessing Method Using Improved Adaptive Neighborhoods ",Proceedings of the 6th World Congress on Intelligent Control and Automation, June 21 - 23, 2006, Dalian, China

[5] Zhou Wang, Alan Conard Bovik, Hamid Rahim Sheik and Erno P Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity", IEEE Trans. Image Processing, and Vol. 13, (2004).

[6] R.M. Haralick, and L.G. Shapiro, "Survey: Image Segmentation Techniques," Computer Vision, Graphics, and Image Processing, Vol. 29, pp. 100-132, 1985.

[7] F. Cremer and W. de Jong and K. Schutte.'Infrared polarization measurements of surface and buried anti-personnel landmines'. In Preprint Proc. SPIE Vol. 4394, Detection and Remediation Techniques for Mines and Minelike Targets VI, Orlando FL, USA, Apr. 2001.

[8] Yan Sun, Chengyi Sun and Wanzhen Wang, "Color Images Segmentation Using New Definition of Connected Components", Proceedings of ICSP2000.

[9] K.S. Ravichandran,B. Ananthi, "Color Skin Segmentation Using K-Means Cluster" International Journal of Computational and Applied Mathematics ISSN 1819-4966 Volume 4 Number 2 (2009), pp. 153–157

[10] Punam K. Saha, "Optimum Image Thresholding via Class Uncertainty and Region Homogeneity",IEEE transactions on pattern analysis and machine intelligence, vol. 23, no. 7, july 2001

[11] Oscar González Merino "Image analysis of Infrared Polarization measurements of Landmines", Academic year: 2000/01

[12] H. Brunzell, "Clutter Reduction and Object Detection in Surface Penetrating Radar," in Proc. Radar.97 Conf., 1997 pp. 688-691.

AUTHORS PROFILE

Dr. Padmavathi Ganapathi is the Professor and Head of Department of Computer Science, Avinashilingam University for Women, Coimbatore. She has 23 years of teaching experience and one year Industrial experience. Her areas of interest include Network security and Cryptography and real time communication. She has more than 108 publications at national and International level. She is a life member of many professional organizations like CSI, ISTE, AACE, WSEAS, ISCA, and UWA.

Dr. Subashini is the Associate professor in Department of Computer Science, Avinashilingam Deemed University for Women, Coimbatore. She has 16 years of teaching experience. Her areas of interest include Object oriented technology, Data mining, Image processing, Pattern recognition. She has 55 publications at national and International level**.**

Ms.M.Krishnaveni has 4 Years of Research Experience Working as Research staff in DRDO and pursuing her Ph.D in Avinashilingam University for Women, Coimbatore. Her research interest are Image Processing, Pattern Recognition and Neural Networks.She has 24 publications at national and international level.

Decision Tree Classification of Remotely Sensed Satellite Data using Spectral Separability Matrix

M. K. Ghose , Ratika Pradhan, Sucheta Sushan Ghose
Department of Computer Science and Engineering
Sikkim Manipal Institute of Technology,
Sikkim, INDIA
mkghose2000@yahoo.com, ratika_pradhan@yahoo.co.in, suchetaghose@gmail.com

***Abstract*— In this paper an attempt has been made to develop a decision tree classification algorithm for remotely sensed satellite data using the separability matrix of the spectral distributions of probable classes in respective bands. The spectral distance between any two classes is calculated from the difference between the minimum spectral value of a class and maximum spectral value of its preceding class for a particular band. The decision tree is then constructed by recursively partitioning the spectral distribution in a Top-Down manner. Using the separability matrix, a threshold and a band will be chosen in order to partition the training set in an optimal manner. The classified image is compared with the image classified by using classical method Maximum Likelihood Classifier (MLC). The overall accuracy was found to be 98% using the Decision Tree method and 95% using the Maximum Likelihood method with kappa values 97% and 94 % respectively.**

Keywords- Decision Tree Classifier (DTC), Separability Matrix, Maximum Likelihood Classifier (MLC), Stopping Criteria.

I. INTRODUCTION

Image classification is one of the primary tasks in geo-computation, that being used to categorize for further analysis such as land management, potential mapping, forecast analysis and soil assessment etc. Image classification is method by which labels or class identifiers are attached to individual pixels on basis of their characteristics. These characteristics are generally measurements of their spectral response in various bands. Traditionally, classification tasks are based on statistical methodologies such as Minimum Distance-to-Mean (MDM), Maximum Likelihood Classification (MLC) and Linear Discrimination Analysis (LDA). These classifiers are generally characterized by having an explicit underlying probability model, which provides a probability of being in each class rather than simply a classification. The performance of this type of classifier depends on how well the data match the pre-defined model. If the data are complex in structure, then to model the data in an appropriate way can become a real problem.

In order to overcome this problem, non-parametric classification techniques such as Artificial Neural Network (ANN) and Rule-based classifiers are increasingly being used. Decision Tree classifiers have, however, not been used widely by the remote sensing community for land use classification despite their non-parametric nature and their attractive properties of simplicity, flexibility, and computational efficiency [1] in handling the non-normal, non-homogeneous and noisy data, as well as non-linear relations between features and classes, missing values, and both numeric and categorical inputs [2].

In this paper, an attempt has been made to develop a decision tree classification algorithm specifically for the classification of remotely sensed satellite data using the separability matrix of spectral distributions of probable classes. The computational efficiency is measured in terms of computational complexity measure. The proposed algorithm is coded in Visual C++ 6.0 language to develop user-friendly software for decision tree classification that requires a bitmap image of the area of interest as the basic input. For the classification of the image, the training sets are chosen for different classes and accordingly spectral separability matrix is obtained. To evaluate the accuracy of proposed method, a confusion matrix analysis was employed and kappa coefficient along with errors of omission and commission were also determined. Lastly, the classified image is compared with the image classified by using classical method MLC.

II. RELATED WORK

The idea of using decision trees to identify and classify objects was first reported by Hunt et al. [3]. Morgan and Sonquist [4] developed the AID (Automatic Interaction Detection) program followed by the THAID developed by Morgan and Messenger [5]. Breiman et al. [6] proposed the CART (Classification and Regression Trees) to solve classification problems. Quinlan [7] developed a decision tree software package called ID3 (Induction of Decision Tree) based on the recursive partitioning greedy algorithm and information theory, followed by the improved version C4.5 addressed in [2]. Buntine [8] developed IND package using the standard algorithms from Brieman's CART and Quinlan's ID3 and C4. It also introduces the use of Bayesian and minimum length encoding methods for growing trees and graphs. Sreerama Murthy [9] reported the decision tree package OC1

(Oblique Classifier 1) which is designed for applications where instances have numeric continuous feature values. Friedl and Brodley [10] have used decision tree classification method to classify land cover using univariate, multivariate and hybrid decision tree as the base classifier and found that hybrid decision tree outperform the other two. Mahesh Pal and Paul M. Mather [11], [12] have suggested boosting techniques for the base classifier to classify remotely sensed data to improve the overall accuracy. Min Xu et al. [13] have suggested decision tree regression approach to determine class proportion within the pixels so as to produce soft classification for remote sensing data. Michael Zambon et al. [14] have used rule based classification using CTA (Classification Tree Analysis) for classifying remotely sensed data and they have found that Gini and class probability splitting rules performs well compared to towing and entropy splitting rules. Mahesh pal [15] have suggested ensemble approaches that includes boosting, bagging, DECORATE and random subspace with univariate decision tree as the base classifier and found that later three approaches works even well compared to boosting. Mingxiang Huang et al. [16] have suggested Genetic Algorithm (GA) based decision tree classifier for remote sensing data SPOT – 5 and found that GA- based decision tree classifier outperform all the classical classification method used in Remote sensing. Xingping Wen et al. [17] have used CART (Classification and Regression Trees) and C5.0 decision tree algorithms for remotely sensed data. Abdelhamid A. Elnaggar and Jay S. Noller [18] have also found that decision tree analysis is a promising approach for mapping soil salinity in more productive and accurate ways compared to classical classification approach.

III. DECISION TREE APPROACH

A decision tree is defined as a connected, acyclic, undirected graph, with a root node, zero or more internal nodes (all nodes except the root and the leaves), and one or more leaf nodes (terminal nodes with no children), which will be termed as an ordered tree if the children of each node are ordered (normally from left to right) (Coreman et al. [19]). A tree is termed as univariate, if it splits the node using a single attribute or a multivariate, if it uses several attributes. A binary tree is an ordered tree such that each child of a node is distinguished either as a left child or a right child and no node has more than one left child or more than one right child. For a binary decision tree, the root node and all internal nodes have two child nodes. All non-terminal nodes contain splits.

A Decision Tree is built from a training set, which consists of objects, each of which is completely described by a set of attributes and a class label. Attributes are a collection of properties containing all the information about one object. Unlike class, each attribute may have either ordered (integer or a real value) or unordered values (Boolean value).

Several methods (Breiman et.al [6], Quinlan [2] and [7]) have been proposed to construct decision trees. These algorithms generally use the recursive-partitioning algorithm, and its input requires a set of training examples, a splitting rule, and a stopping rule. Partitioning of the tree is determined by the splitting rule and the stopping rule determines if the examples in the training set can be split further. If a split is still possible, the examples in the training set are divided into subsets by performing a set of statistical tests defined by the splitting rule. The test that results in the best split is selected and applied to the training set, which divides the training set into subsets. This procedure is recursively repeated for each subset until no more splitting is possible.

Stopping rules vary from application to application but multiple stopping rules can be used across different applications. One stopping rule is to test for the purity of a node. For instance, if all the examples in the training set in a node belong to the same class, the node is considered to be pure (Breinman et.al [6]) and no more splitting is performed. Another stopping rule is by looking at the depth of the node, defined by the length of the path from the root to that node (Aho et.al. [20]). If the splitting of the current node will produce a tree with a depth greater than a pre-defined threshold, no more splitting is allowed. Another common stopping rule is the example size. If the number of examples at a node is below a certain threshold, then splitting is not allowed. Four widely used splitting rules that segregates data includes: gini, twoing, entropy and class probability. The gini index is defined as:

$$gini(t) = \sum p_i (1 - p_i) \qquad (1)$$

where pi is the relative frequency of class i at node t, and node t represents any node (parent or child) at which a given split of the data is performed (Apte and Weiss [21]). The gini splitting rule attempts to find the largest homogeneous category within the dataset and isolate it from the remainder of the data. Subsequent nodes are then segregated in the same manner until further divisions are not possible. An alternative measure of node impurity is the towing index:

$$twoing(t) = \frac{p_L p_R}{4} \left(\sum_i \left| p(i|t_L) - p(i|t_R) \right| \right)^2 \qquad (2)$$

where L and R refer to the left and right sides of a given split respectively, and p(i|t) is the relative frequency of class i at node t (Breiman [22]). Twoing attempts to segregate data more evenly than the gini rule, separating whole groups of data and identifying groups that make up 50 percent of the remaining data at each successive node. Entropy is a measure of homogeneity of a node and is defined as:

$$entropy(t) = -\sum_i p_i \log p_i \qquad (3)$$

where p_i is the relative frequency of class i at node t (Apte and Weiss [21]). The entropy rule attempts to identify splits where as many groups as possible are divided as precisely as possible and forms groups by minimizing the within group diversity (De'ath and Fabricius [23]). Class probability is also based on the gini equation but the results are focused on the probability structure of the tree rather than the classification structure or prediction success. The rule attempts to segregate the data

based on probabilities of response and uses class probability trees to perform class assignment (Venables and Ripley [24]).

From the above discussions, it is evident that a decision tree can be used to classify a pixel by starting at the root of the tree and moving through it until a leaf is encountered. At each non-leaf decision node, the outcome for the test at the node is determined and attention shifts to the root of the sub-tree corresponding to this outcome. This process proceeds until a leaf is encountered. The class that is associated with the leaf is the output of the tree. A class is one of the categories, to which pixels are to be assigned at each leaf node. The number of classes is finite and their values must be established beforehand. The class values must be discrete. A tree misclassifies a pixel if the class label output by the tree does not match the class label. The proportion of pixels correctly classified by the tree is called accuracy and the proportion of pixels incorrectly classified by the tree is called error (Coreman et.al., 1989).

IV. The Proposed Criteria

To construct a classification tree, it is assumed that spectral distributions of each class are available. The decision tree is then constructed by recursively partitioning the spectral distribution into purer, more homogenous subsets on the basis of the tests applied to feature values at each node of the tree, by employing a recursive divide and conquer strategy. This approach to decision tree construction thus corresponds to a top-down greedy algorithm that makes locally optimal decisions at each node. Steps involved in this process can be summarized as below:

- Optimal band selection: By using the separability matrix, a threshold and band is chosen in order to partition the training set in an optimal manner.
- Based on a partitioning strategy, the current training set is divided into two training subsets by taking into account the values of the threshold.
- When the stopping criterion is satisfied, the training subset is declared as a leaf.

The separability matrices are obtained in the respective bands, by calculating the spectral distance between pairs of classes in row and column. The spectral distance between any two classes is calculated from the difference between the minimum spectral value of a class and maximum spectral value of its preceding class for a particular band. More the spectral distance, maximum is the separability. For spectral distance less than zero there will be overlapping between classes.

A. Splitting Criteria

An attempt has been made to define splitting rules by calculating the separability matrix to split the given set of classes into two subsets where the separability is maximum or the amount of spectral overlapping is minimum between two subsets. That is, the split group together a number of classes that are similar in some characteristic near the top of the tree and isolate single class in the bottom of the tree. First the spectral classes are arranged in ascending order based on the value of their midpoints. Lower the value of the midpoint lowers the class orders. Here the midpoint of a class is calculated by (Min+ Max)/2. This will result in reduction of the matrix computation/memory allocation, wherein only the upper diagonal elements are to be considered for finding the threshold. The threshold value is calculated, as the midpoint between the spectral distributions of the classes in the band where, the separability is maximum or overlapping is minimum. Once the split is found, this is used to find the subsets at each node.

B. Determining the Terminal Node or Stopping Criteria

When a subset of classes becomes pure, create a node and label it by the class number of the pure subset having only one class. If a subset having more than one class, apply the splitting criteria till it becomes purer.

V. The Proposed Algorithm

Notations and Assumptions:

S – *the set of spectral distributions of all classes with class label.*

n – *the number of classes.*

K – *the number of bands.*

$M_k = \left(m_{ij}^k\right), (1 \le i \le n, 1 \le j \le n)$ – *the separability matrix in band k.*

C_{kl}^1 – *the range* minimum *of the spectral distribution in band k for the class l.*

C_{kl}^2 – *the range* maximum *of the spectral distribution in band k for the class l.*

Initialize the root by all the classes.

$\mathrm{DTree}\,(N, S, n)$

Step 1: For each band $1 \le k \le K$, sort the spectral distributions with respect to the midpoint.

If $n \le 2$ and overlapping, then construct the tree as in special case.

else

goto step2.

Step 2:

$$For\, 1 \le k \le K\ ,\ 1 \le i \le n\ ,\ 1 \le j \le n$$

$$M_k = \left(m_{ij}^k\right) = \begin{cases} 0 & ,\ if\ \ i = j \\ C_{kj}^1 - C_{ki}^2 & ,\ otherwise \end{cases}$$

Step 3: Find the threshold value, which will divide the set of classes S into two subsets of classes, say S_L and S_R, such that the separability between a pair of classes is maximum. The procedure is as follows.

Case 1: In all the K matrices, consider only the rows having all the elements to the right of the diagonal element are positive. If no such row exists in all the matrices go to case 2.

Find the minimum element from each such row. Find a maximum from all such minimum elements. Let this element be m_{rc}^{b} which is at the r^{th} row and c^{th} column of the matrix M_b. That is,

$$If\ \exists\, b,r,c\ such\ that\ m_{rc}^{b} = \max_{1\le k\le K}\left\{\max_{1\le i<n}\left\{\min_{i+1\le j\le n}\left\{m_{ij}^{k} \;/\; m_{ij}^{k}\ge 0\ \ for\ all\ j\right\}\right\}\right\}$$

then the split is in between the classes represented by the row(r) and column(c) in band b.

$$\text{Threshold(T)} = \frac{1}{2}\left[\begin{Bmatrix}\text{range maximum of the } class\\ distribution\ represented\\ by\ the\ row\ in\ band\ b\end{Bmatrix} + \begin{Bmatrix}\text{range minimum of the } class\\ distribution\ represented\\ by\ the\ column\ in\ band\ b\end{Bmatrix}\right]$$

$$= \frac{1}{2}\left\{C_{br}^{2} + C_{bc}^{1}\right\}$$

BAND = b, go to step 4.

Case 2: In all K matrices, consider only the rows having at least one positive element. Find the minimum element from each such row. Find a maximum from all such minimum elements. Let this element be m_{rc}^{b} which is at the r^{th} row and c^{th} column of the matrix M_b. That is,

$$If\ \exists\, b,r,c\ such\ that\ m_{rc}^{b} = \max_{1\le k\le K}\left\{\max_{1\le i<n}\left\{\min_{i+1\le j\le n}\left\{m_{ij}^{k} \;/\; m_{ij}^{k}\ge 0\right\}\right\}\right\}$$

then the split is in between the classes represented by the row(r) and column(c) in band b. Find the threshold (T = T2) as in case 1. Let BAND = BAND2 = b.

Check whether the threshold T lies in any of the spectral range in the band b, except for the class distributions represented by r and c.

If yes, compute $\min\left\{\left|\,T - C_{bi}^{1}\,\right|, \left|\,T - C_{bi}^{2}\,\right|\right\}$ for each class i, $i \ne r$ and $i \ne c$, such that $T \in (C_{bi}^{1}, C_{bi}^{2})$. Find a maximum of all such minima. Call it as EF2 and then go to case 3.

i) If no, go to step 4.

Case 3: From all the upper triangular elements of all K matrices, find the minimum negative element; say m_{rc}^{b} which is at the r^{th} row and c^{th} column of the matrix M_b.

That is, find $m_{rc}^{b} = \max_{1\le k\le K}\left\{\max_{i<j}\left\{m_{ij}^{k}\right\}\right\}$

Calculate the threshold (T = T3) as in above cases. Let BAND = BAND3 = b.

Compute $\min\left\{\left|\,T - C_{bi}^{1}\,\right|, \left|\,T - C_{bi}^{2}\,\right|\right\}$ for each class i, $i \ne r$ and $i \ne c$, such that $T \in (C_{bi}^{1}, C_{bi}^{2})$. Find the maximum of all such minima. Call it as EF3.

If EF2 ≤ EF3, let $T = T_2$ and BAND = BAND2, go to step 4.

Step 4: Assign the threshold (T) & BAND to the node N.

Step 5: Find the subsets of classes; say left subset (S_L) and the right subset (S_R) as follows.

S_L = Set of classes having distributions with range maximum or mid-point ≤ T in band BAND and let n_L be the number of classes in S_L.

S_R = Set of classes having distributions with range minimum or mid-point > T in band BAND and let n_R be the number of classes in S_R.

Step 6: Initialize the left node (N_L) of N by all the classes in S_L and the right node (N_R) of N by all the classes in S_R.

Step 7: If $n_L \le 1$, terminate splitting and return a node with the corresponding class Label. Else

$$\text{DTree}(N_L, S_L, n_L).$$

If $n_R \le 1$, terminate splitting and return a node with the corresponding class Label. Else

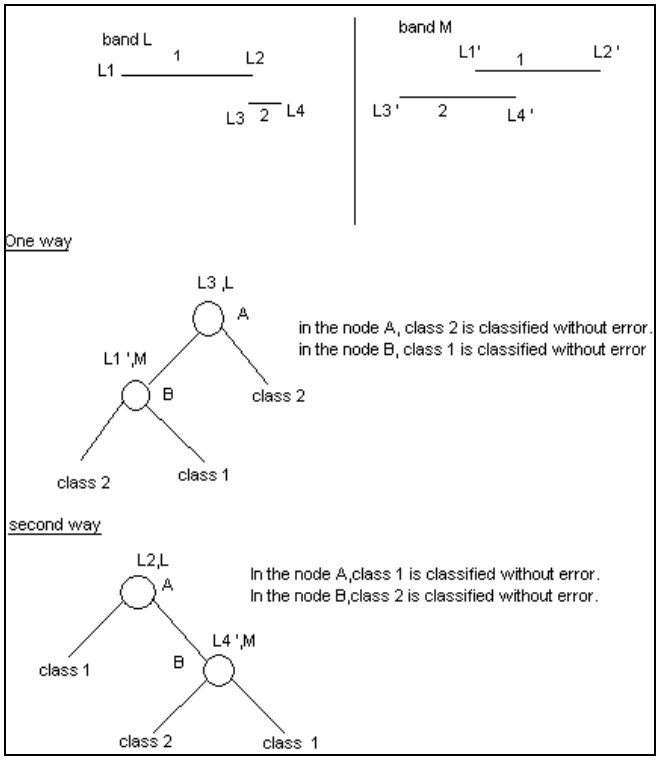
$$\text{DTree}(N_R, S_R, n_R)$$

A. Special Case

When n=2, and these two classes are overlapping in all the bands, we can construct a decision tree in the following ways:

Consider the following distributions:

We choose way 1, when the node A is the left child of its parent and way 2, when it is the right child of its parent. Because, when the node A is on the left branch and the threshold in node A and the threshold in its parent are chosen from the same band, then it will reduce the error in classification and the same reason applies for the other one.



VI. Application and Case Study

To validate the applicability of the proposed decision tree algorithm, a case study is presented in this section, which is carried out on IRS-1C/LISS III sample image with 23m resolution. The FCC (False Color Composite) of the input image (Figure 1) belongs to the Mayurakshi reservoir (Latitude: 240 09' 47.27'' N and Longitude: 87017'49.41" E) of Jharkhand state (INDIA) and band used to prepare FCC includes – Red (R), Green (G), Near Infrared (NIR). The IRS-IC/LISS III data of Mayurakshi reservoir was acquired on October, 2004. The input image is first converted into a bitmap, and then used as input for the DTCUSM software. As per the software requirement the training set along with test set were selected. For the present study, eleven classes viz., Turbid water, Clear water, Forest, Dense Forest, Upland Fallow, Lateritic Cappings, River Sand, Sand, Drainage, Fallow, and Wetland are considered. The training set for these eleven classes is chosen considering the prior knowledge of the hue, tone and texture of these classes, in addition, physical verification on ground were also done for each of these class. The spectral class distributions for the training set taken from the input image are shown in Table I.

Once the training for all the classes is set, the proposed decision tree algorithm is applied to classify the image. The decision tree construction steps for the spectral class distributions given in Table 1 are shown in Figure 2. The output image after applying the proposed Decision Tree Classification method is shown in Figure 3.

To assess accuracy of the proposed technique, the confusion matrix along with the errors of omission and commission and overall accuracy, kappa coefficient (Jensen, 1996) are obtained and shown in Table II. For the comparison purpose, the same image is classified again by using Maximum Likelihood Classifier (Jensen, 1996) with same training sets as used in Decision tree classification. The classified image by Maximum Likelihood method is shown in Figure 4. The percentage of pixels in each class is given for both the cases in Figure 5.

TABLE I. Spectral class distributions for the training set taken from the image in Figure 1.

	Turbid Water	Clear Water	Forest	Dense Forest	Upland Fallow	Lateritic Cappings	River Sand	Sand	Drainage	Fallow	Wetland
Band 1	**0-11**	**0-31**	**140-226**	**83-136**	**48-110**	**49-81**	**223-255**	**158-238**	**131-208**	**113-195**	**56-164**
Band 2	**17-64**	**2-57**	**6-103**	**8-52**	**82-131**	**32-76**	**230-255**	**220-255**	**105-177**	**149-230**	**82-222**
Band 3	**67-162**	**0-70**	**0-104**	**0-42**	**44-88**	**4-40**	**236-255**	**223-255**	**127-195**	**123-197**	**117-255**

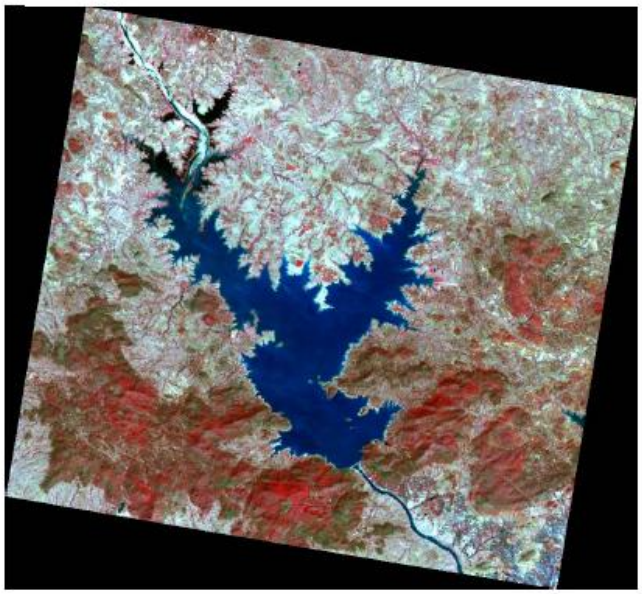


Figure 1. FCC (IRS LISS III Image – OCT 2004) of study area – Mayurakshi Reservoir, Jharkhand, INDIA, Band used – Red (R), Green (G), Near Infrared (NIR).

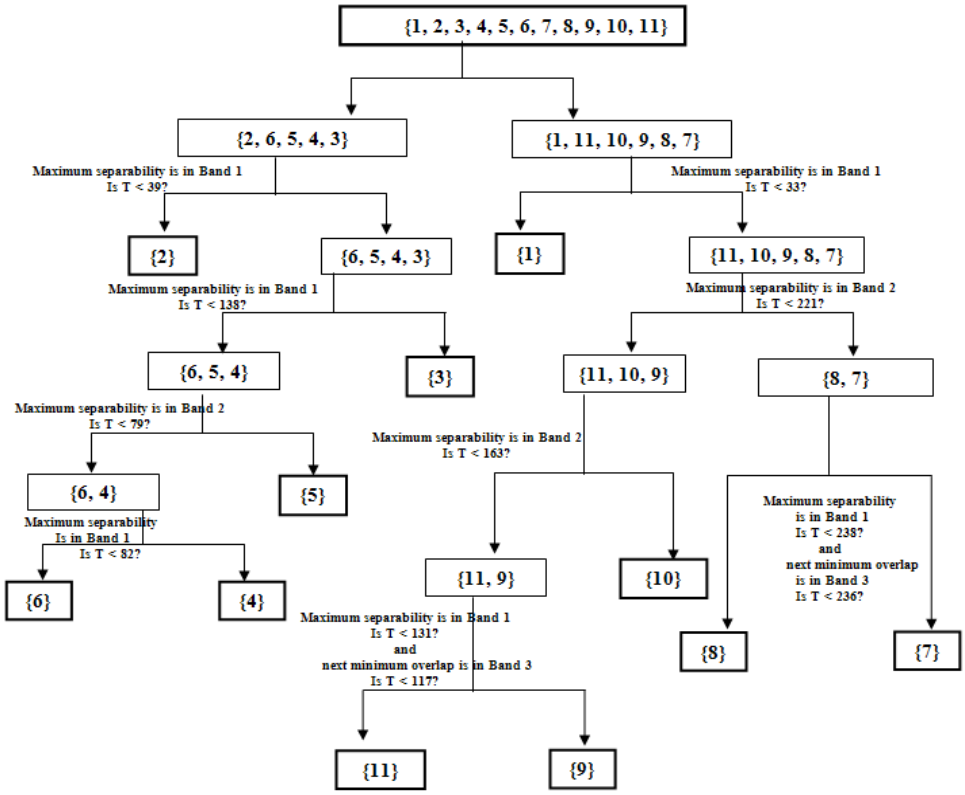


Figure 2. Decision tree construction steps for distribution given in table 1.

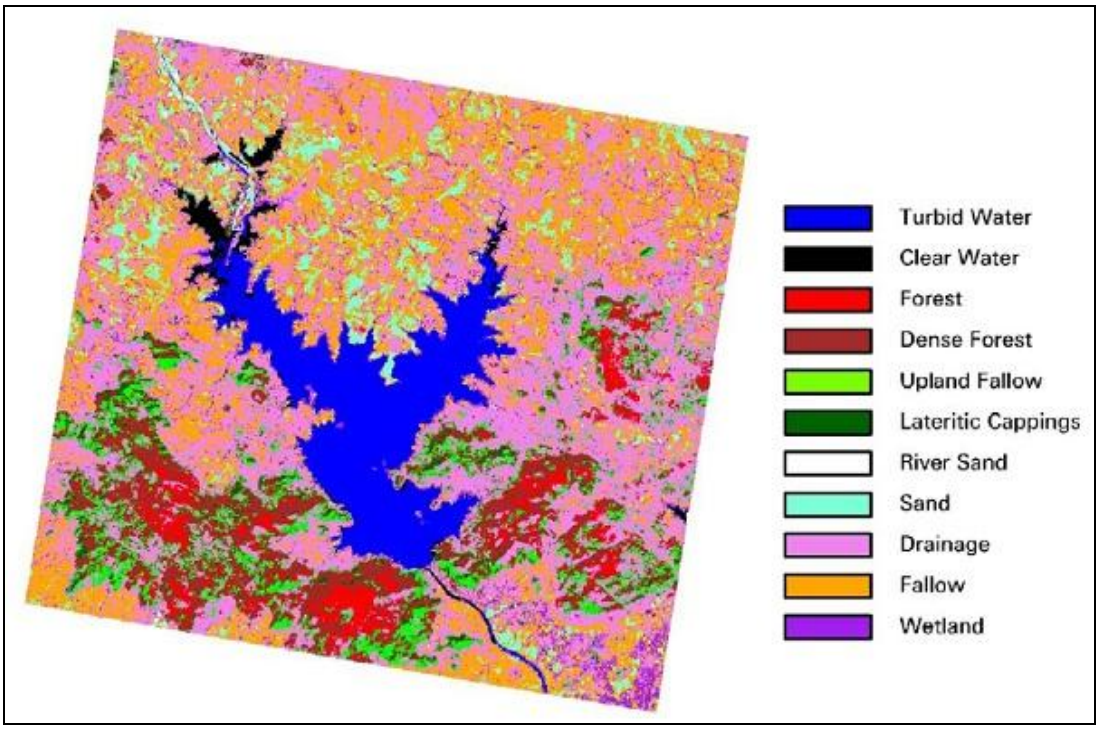


Figure 3. Classified image using the proposed Decision Tree classifier.

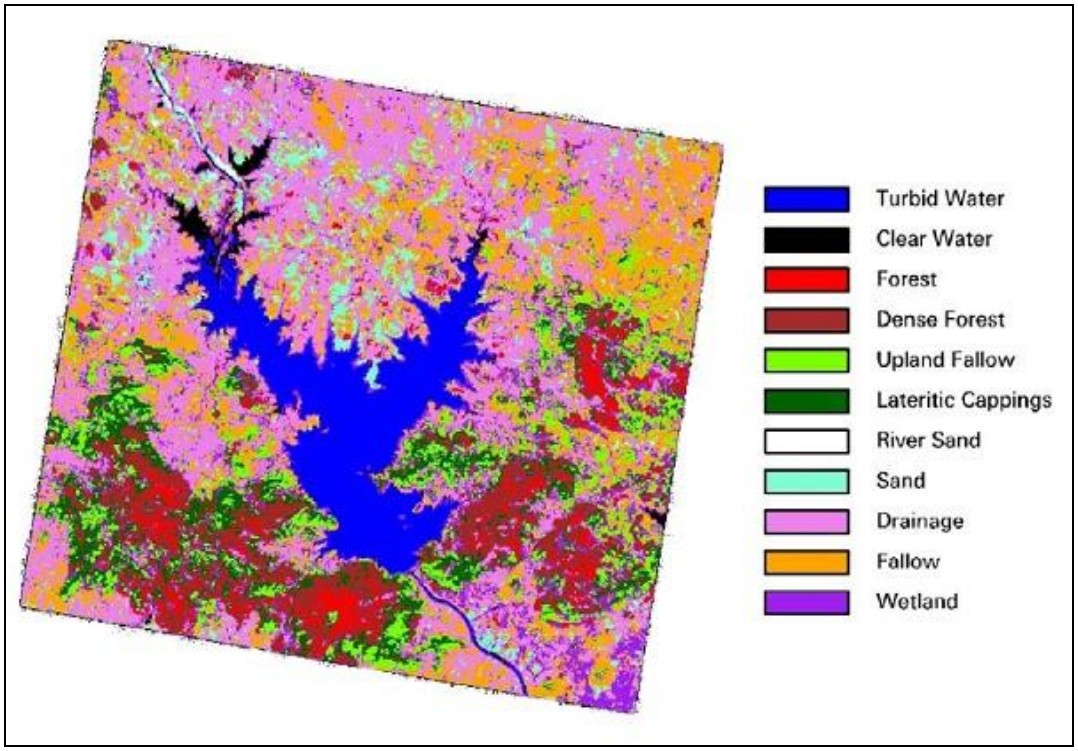


Figure 4. Classified image using standard Maximum Likelihood classifier.

TABLE II. CONFUSION MATRIX AND ACCURACY CALCULATIONS FOR DECISION TREE METHOD

	Turbid Water	Clear Water	Forest	Dense Forest	Upland Fallow	Lateritic Cappings	River Sand	Sand	Drainage	Fallow	Wetland	Error of omission (in %)
Turbid Water	323	0	0	0	0	0	0	0	0	0	0	**0.00**
Clear Water	1	34	0	0	0	0	0	0	0	0	0	**2.86**
Forest	0	0	42	0	0	0	0	0	0	0	0	**0.00**
Dense Forest	0	0	0	35	0	0	0	0	0	0	0	**0.00**
Upland Fallow	0	0	0	0	15	0	0	0	1	0	0	**6.25**
Lateritic Cappings	0	0	0	0	0	9	0	0	0	0	0	**0.00**
River Sand	0	0	0	0	0	0	12	3	0	0	0	**20.00**
Sand	0	0	0	0	0	0	0	16	0	0	0	**0.00**
Drainage	0	0	0	0	0	0	0	0	15	0	0	**0.00**
Fallow	0	0	0	0	0	0	0	0	0	12	0	**0.00**
Wetland	0	0	0	0	0	0	0	0	0	2	4	**33.33**
Error of commission (in %)	**0.31**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**15.79**	**6.25**	**14.29**	**0.00**	

Overall Accuracy = 98.66% Kappa = 97.77%

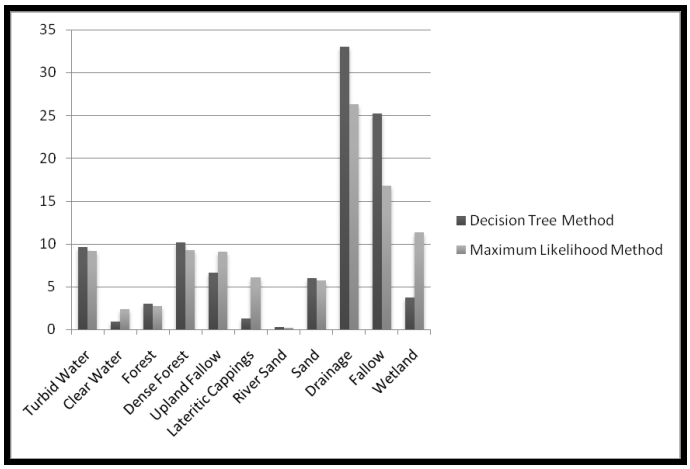


Figure 5. Class wise percentage of pixels for Decision Tree and Maximum likelihood methods

VII. SUMMARY AND CONCLUSION

In this paper, a decision tree classification algorithm for remotely sensed satellite data using separability matrix of spectral distributions of probable classes has been developed. To test and validate the proposed Decision Tree algorithm, the sample image taken into consideration is multi-spectral IRS-1C/LISS III of Mayurakshi reservoir of Jharkhand state. The proposed Decision Tree classifier can also be used for hyper-spectral remote sensing data considering the best bands as input for preparing spectral class distribution. The sample image is classified by both Decision Tree method and Maximum Likelihood method and then the overall accuracy, kappa coefficients were calculated. The overall accuracy for the sample test image was found to be 98% using the Decision Tree method and 95% using the Maximum Likelihood method with kappa values 97% and 94 % respectively. The reason for high accuracy may be to some extent attributed for the reason

that the part of the training set is being considered as ground truths instead of actual data. Since the accuracy of the results depends only upon the test set chosen, the efficiency of any algorithm shall not be considered on the accuracy measure alone. Out of eleven classes considered for the sample image, many classes were found to be closely matching in both the methods. However, differences are observed in certain classes in both the methods. The classified images shall also be compared with the input image (FCC) and collecting ground truth information physically. From the comparison, it is found that both the methods are equally efficient, but the decision tree algorithm will have an edge over its statistical counterpart because of its simplicity, flexibility and computational efficiency.

REFERENCES

[1] Friedman, J. H., 1977. A Recursive Partitioning Decision Rule for Nonparametric Classification, IEEE Trans. Computers, Vol C-26, pp 404-408.

[2] Quinlan, J.R., 1993. C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, California, 302 p.

[3] Hunt, E.B, Marin, J., and Stone, P. J., 1966. Experiments in Induction, Academic Press.

[4] Morgan, J. N., and Sonquist, J. A, 1963. Problems in the Analysis of Survey Data and A Proposal, Journal of American Statistics Association, Vol. 58, pp 415-434.

[5] Morgan, J. N., and Messenger, R. C., 1973. THAID- A Sequential Search Program for the Analysis of Nominal Scale Dependent Variables, Institute for Social Research, University of Michigan, Ann Arbor.

[6] Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J., 1984. Classification and Regression Trees, Wadsworth International Group, Belmont, California, 358 p.

[7] Quinlan, J. R., 1975. Machine Learning, Vol. 1, No. 1, University of Sydney.

[8] Buntine, W., 1992. Tree Classification Software, Third National Technology Transfer Conference and Exposition, Baltimore, MD.

[9] Murthy, S.K., Kasif, S., and Salburg, S. 1994. A System for Induction of Oblique Decision Tree, Vol.2, pp. 1-32.

[10] Friedl, M.A., and Brodley, C.E., 1997. Decision tree classification of land cover from remotely sensed data, Remote Sensing of Environment, Vol. 61, pp 399–409.

[11] Mahesh Pal, and Paul M. Mathers, 2001. Decision Tree Based Classification of Remotely Sensed Data, 22nd Asian Conference on Remote Sensing.

[12] Mahesh Pal, and Paul M. Mathers, 2003. An assessment of the effectiveness of decision tree methods for land cover classification, Remote Sensing of Environment, Vol. 86, pp 554–565.

[13] Min Xu, Pakorn Watanachaturaporn, Pramod K. Varshney, and Manoj K. Arora, 2005. Decision tree regression for soft classification of remote sensing data, Remote Sensing of Environment, Vol. 97, No. 3, pp 322-336.

[14] Michael Zamban, Rick Lawrence, Andrew Bunn, and Scott Powell, 2006. Effect of Alternative Splitting Rules on Image Processing Using Classification Tree Analysis, Photogrammetric Engineering and Remote Sensing, Vol. 72, No. 1, January 2006, pp 25-30.

[15] Mahesh Pal, 2007. Ensemble Learning with Decision Tree for Remote Sensing Classification, World Academy of Science, Engineering and Technology, pp 258-260.

[16] Mingxiang Huang, Jianghua Gong, Zhou Shi, Chunbo Liu, and Lihui Zhang, 2007. Genetic algorithm-based decision tree classifier for remote sensing mapping with SPOT-5 data in the HongShiMao watershed of the loess plateau, China, Neural Computing and Applications archive, Vol. 16, No. 6 pp 513 – 517.

[17] Xingping Wen, Guangdao Hu and Xiaofeng Yang, 2008. CBERS-02 remote sensing data mining using decision tree algorithm, International Conference on Forensic Applications and Techniques in Telecommunications, Information, and Multimedia and Workshop, Article No. 58.

[18] Abdelhamid A. Elnaggar and Jay S. Noller, 2010. Application of Remote-sensing Data and Decision-Tree Analysis to Mapping Salt-Affected Soils over Large Areas, Remote Sensing, Vol. 2, pp 151-165.

[19] Cormen, T. H., Leiserson, C. E. and Rivest, R. L. 1989. Introduction to Algorithms, MIT Press, Cambridge, MA.

[20] Aho, A. V., Hopcroft, J. E., and Ullman, J. D., 1974. The Design and Analysis of Computer Algorithms, Addison-Wesley Publishing Company, Reading, MA.

[21] Apte, C., and Weiss, S., 1997. Data mining with decision trees and decision rules, Future Generation Computer Systems, Vol. 13, pp 197–210.

[22] Breiman, L., 1978. Description of Chlorine Tree Development and Use, Technical Report, Technology Service Corporation, Santa Monica, CA.

[23] De'ath, G., and Fabricius, K., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis, Ecology, Vol. 81-11, pp 3178–3192.

[24] Venables, W.N., and B.D. Ripley, 1999. Modern Applied Statistics with S-Plus, Third Edition, Springer-Verlag, New York, 501 p.

AUTHORS PROFILE

Dr. M. K. Ghose is the Professor and Head of the Department of Computer Science & Engineering at Sikkim Manipal Institute of Technology, Mazitar, Sikkim, India since June, 2006. Prior to this, Dr. Ghose worked in the internationally reputed R & D organization ISRO – during 1981 to 1994 at Vikram Sarabhai Space Centre, ISRO, Trivandrum in the areas of Mission simulation and Quality & Reliability Analysis of ISRO Launch vehicles and Satellite systems and during 1995 to 2006 at Regional Remote Sensing Service Centre, ISRO, IIT Campus, Kharagpur in the areas of RS & GIS techniques for the natural resources management. His area of research interest are Data Mining, Simulation & Modeling, Network & Information Security, Optimization & Genetic Algorithm, Digital Image processing, Remote Sensing & GIS and Software Engineering. Dr. Ghose has conducted quite a number of Seminars, Workshop and Training programmes in the above areas and published 100 technical papers in various national and international journals in addition to presentation/ publication in several international/ national conferences. Presently 9 scholars are pursuing Ph.D work under his guidance and he is having 8 sponsored projects.

Mrs. Ratika Pradhan is the Associate Professor of the Department of Computer Science & Engineering at Sikkim Manipal Institute of Technology, Mazitar, Sikkim, India. She completed her Bachelor of Engineering in Computer Science and Engineering from Adhiparasakthi Engineering College, Madras University, Tamilnadu in the year 1998. She completed her Master of Engineering in Computer Science and Engineering from Jadavpur University in the year 2004. Her research interest is mainly in Digital Image Processing and Remote Sensing & GIS.

Ms. Sucheta Sushan Ghose is the Assitant Professor of the Department of Computer Science & Engineering at Sikkim Manipal Institute of Technology, Mazitar, Sikkim, India. She completed her Bachelor of Technology in Information Technology, Sikkim Manipal University, Sikkim in the year 2009.

Texture Based Segmentation using Statistical Properties for Mammographic Images

H.B.Kekre
Department of Computer Engineering
Senior Professor, MPSTME, SVKM's NMIMS University
Mumbai, India
hbkekre@yahoo.com

Saylee Gharge
Department of Electronics and telecom. Engineering
Ph.D. Scholar, MPSTME, SVKM's NMIMS University
Mumbai, India
Sayleegharge73@yahoo.co.in

***Abstract*— Segmentation is very basic and important step in computer vision and image processing. For medical images specifically accuracy is much more important than the computational complexity and thus time required by process. But as volume of data of patients goes on increasing then it becomes necessary to think about the processing time along with accuracy. Here in this paper, new algorithm is proposed for texture based segmentation using statistical properties. For that probability of each intensity value of image is calculated directly and image is formed by replacing intensity by its probability . Variance is calculated in three different ways to extract the texture features of the mammographic images. These results of proposed algorithm are compared with well known GLCM and Watershed algorithm.**

Keywords- Segmentation, Variance, Probability, Texure

I. INTRODUCTION

Segmenting mammographic images into homogeneous texture regions representing disparate tissue types is often a useful preprocessing step in the computer-assisted detection of breast cancer. With the increasing size and number of medical images, the use of computers in facilitating their processing and analysis has become necessary. Estimation of the volume of the whole organ, parts of the organ and/or objects within an organ i.e. tumors is clinically important in the analysis of medical image. The relative change in size, shape and the spatial relationships between anatomical structures obtained from intensity distributions provide important information in clinical diagnosis for monitoring disease progression. Therefore, radiologists are particularly interested to observe the size, shape and texture of the organs and/or parts of the organ. For this, organ and tissue morphometry performed in every radiological imaging centre. Texture based image segmentation is area of intense research activity in the past few years and many algorithms were published in consequence of all this effort, starting from simple thresholding method up to the most sophisticated random field type method. The repeating occurrence of homogeneous regions of images is texture. Texture image segmentation identifies image regions that have homogeneous with respect to a selected texture measure. Recent approaches to texture based segmentation are based on linear transforms and multiresolution feature extraction [1], Markov random filed models [2,3], Wavelets [4 – 6] and fractal dimension [7]. Although unsupervised texture-based image segmentation is not a novel approach, these have limited adoption due to their high computational complexity.

Segmentation methods are based on some pixel or region similarity measure in relation to their local neighborhood. These similarity measures in texture segmentation methods use some textural spatial-spectral-temporal features such as Markov random field statistics (MRF) [8-10], co-occurrence matrix based features [11], Gabor features [12], local binary pattern (LBP) [13], autocorrelation features and many others. A number of image processing methods have been proposed to perform this task. S**.** M. Lai et al. [14] and W. Qian et al. [15] have proposed using modified and weighted median filtering, respectively, to enhance the digitized image prior to object identification. D. Brzakovic et al. [16] used thresholding and fuzzy pyramid linking for mass localization and classification. Other investigators have proposed using the asymmetry between the right and left breast images to determine possible mass locations. Yin ***et al.*** [17] uses both linear and nonlinear bilateral subtractions while the method by Lau *et al.* [18]. relies on "structural asymmetry" between the two breast images. Recently Kegelmeyer has reported promising results for detecting spiculated lesions based on local edge characteristics and Laws texture features [19-21].The above methods produced a true positive detection rate of approximately 90%. Various segmentation techniques have been proposed based on statistically measurable features in the image [22-27] Clustering algorithms, such as k-means and ISODATA, operate in an unsupervised mode and have been applied to a wide range of classification problems.

For this paper gray level co-occurrence matrix based features and watershed algorithm are considered for comparison with proposed algorithm which is based on statistical properties for segmentation of mammographic images. In section II different algorithms for texture based segmentation are explained in detail. Section III shows results for those methods and section IV concludes the work.

II. ALGORITHMS FOR TEXTURE BASED SEGMENTATION

Texture is one of the most important defining characteristics of an image. It is characterized by the spatial distribution of gray levels in a neighborhood. In order to capture the spatial dependence of gray-level values which contribute to the perception of texture, a two dimensional dependence texture analysis matrix are discussed for texture

consideration. Since texture shows its characteristics by both each pixel and pixel values. There are many approaches using for texture classification.

A. Gray Level Co-occurrence Matrix(GLCM)

The gray-level co-occurrence matrix seems to be a well-know statistical technique for feature extraction. However, there is a different statistical technique using the absolute differences between pairs of gray levels in an image segment that is the classification measures from the Fourier spectrum of image segments. Haralick suggested the use of gray level co-occurrence matrices (GLCM) for definition of textural features. The values of the co-occurrence matrix elements present relative frequencies with which two neighboring pixels separated by distance *d* appear on the image, where one of them has gray level *i* and other *j*. Such matrix is symmetric and also a function of the angular relationship between two neighboring pixels. The co-occurrences matrix can be calculated on the whole image, but by calculating it in a small window which scanning the image, the co-occurrence matrix can be associated with each pixel.

For a 256 gray levels image one should compute 256x256 co-occurrence matrices at all positions of the image. It is obvious that such matrices are too large and their computation becomes memory intensive. Therefore, it is justified to use a less number of gray levels, typically 64 or 32. There is no unique way to choose the values of distance, angle and window, because they are in relationship with a size of pattern.

Using co-occurrence matrix textural features are defined as:

Maximum Probability: $\max(P_{ij})$ (1.1)

$$\text{Variance} = (\sum (i-\mu_i)^2 \sum P_{ij})(\sum (j-\mu_j)^2 \sum P_{ij}) \quad (1.2)$$

$$\text{Correlation} = \sum_i \sum_j (i-\mu_X)(j-\mu_Y) P_{ij} / \sigma_X \sigma_Y \quad (1.3)$$

where μ_x and μ_y are means and σ_x, σ_y are standard deviation

$$\text{Entropy} = \sum_i \sum_j P_{ij} \log(P_{ij}) \quad (1.4)$$

$$\text{Energy} = \sum_i \sum_j p_{i,j}^2 \quad (1.5)$$

Amongst all these features variance, probability has given the best results. Hence results for these extracted features using gray level co-occurrence matrix are displayed in section III.

B. Watershed Algorithm

The watershed transformation is a powerful tool for image transformation .Beucher and Lantuejoul were the first to apply the concept of watershed and divide lines to segmentation problems [28].They used it to segment images of bubbles and metallographic pictures. The watershed transformation considers the gradient magnitude of an image as a topographic surface. Watershed segmentation [29] classifies pixels into regions using gradient descent on image features and analysis of weak points along region boundaries. The image feature space is treated, using a suitable mapping, as a topological surface where higher values indicate the presence of boundaries in the original image data. It uses analogy with water gradually filling low lying landscape basins. The size of the basins grow with increasing amounts of water until they spill into one another. Small basins (regions) gradually merge together into larger basins. Regions are formed by using local geometric structure to associate the image domain features with local extremes measurement.

Watershed techniques produce a hierarchy of segmentations, thus the resulting segmentation has to be selected using either some prior knowledge or manually with trial and error. Hence by using this method the image segmentation can not be performed accurately and adequately, if we do not construct the objects we want to detect. These methods are well suited for different measurements fusion and they are less sensitive to user defined thresholds. In this approach, the picture segmentation is not the primary step of image understanding. On the contrary, a fair segmentation can be obtained only if we know exactly what we are looking for in the image. For this paper, watershed algorithm for mammographic images is implemented as mentioned in [30] and displayed as a result in Figure1(b) and 2(b) .

C. Proposed Algorithm

From the previous section it can be inferred that even though variance using GLCM gives proper tumor demarcation for mammographic images it require huge computation time to calculated statistical properties for the image. Watershed algorithm is comparatively less complex hence less computation time is required but this method gives over segmentation. Hence to achieve proper segmentation with less complexity, new algorithm has been proposed. In this proposed algorithm statistical properties such as variance, probability for grouping pixels into regions and then images are formed for each statistical property.

1) Probability

Images are modeled as a random variable. A full understanding of the properties of images and of the conclusions has drawn from them thus demand accurate statistical models of images. In this paper, probability of image is considered for extraction of the features of an image.

For complete image, probability of particular i^{th} gray level is calculated which is given by:

$$\text{Probability P (i)} = \frac{X_i}{MXN} \quad (1.6)$$

Where Xi is number of pixels for i th gray levels, M and N are no. of rows and columns of the image.

After calculating this the image is formed which contains probability values for that particular gray level instead of gray level in the image .Since the values of probabilities are too

small they are invisible. For perceptibility of this image histogram equalization is preferred and displayed as equalized probability image as shown in Figure 1(e) and 2(e) for mammographic images.

2) Variance

Variance is a measure of the dispersion of a set of data points around their mean value. It is a mathematical expectation of the average squared deviations from the mean. The variance of a real-valued random variable is its second central moment, and it also happens to be its second cumulant. The variance of random variable is the square of its standard deviation.

$$\text{Variance}(X) = E[(X - \mu)^2] \quad (1.7)$$

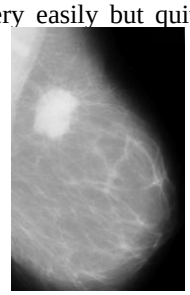
if $\mu = E(X)$, where $E(X)$ is the expectation (mean) of the random variable X. That is, it is the expectation of the square of the deviation of X from its own mean. It can be expressed as "The average of the square of the distance of each data point from the mean", thus it is the mean squared deviation. This same definition is followed here for images. Initially probability of image is calculated. Since the probability values are very small, equalized probability image is applied as an input image to find variance of probability image by using 3x3 window size as given by Equation 1.7. Results are shown in the section III.

By using this approach any abnormality in the image can be observed very easily but quite often the radiologist need other details. In this case original image is used instead of using equalized probability image as an input image. Thus variance of original image is calculated using same Equation 1.7 for window size 3x3 and results are shown as direct variance image in the section III. In the third approach , variance is calculated using probability of the image as given by equation 1.8.Results are shown as variance using probability Figure1(h) and 2(h)

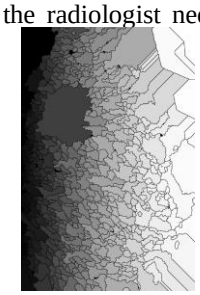
$$\text{Variance using probability}(X) = E[(X - \mu)^2 \times P(X)] \quad (1.8)$$

III. RESULTS

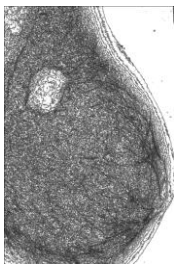
Mammography images from mini-mias database were used in this paper for implementation of GLCM, Watershed and proposed algorithm for tumor demarcation. Fig.2 (a) shows original image with tumor. It has fatty tissues as background. Class of abnormality present is CIRC which means well-defined/ circumscribed masses. Image 1 and Image 2 (mdb184 and mdb028 from database) have malignant abnormalities.Figure 1(a) and 2 (a) show original mammographic images. Figure 1(b) and 2 (b) indicates segmentation using watershed algorithm .Figure 1 and 2 (c)-(d) show results for probability and variance using GLCM. Figure 1 and 2 (e)-(h) show equalized probability, variance of probability ,direct variance and variance using probability image for image 1 and image 2.



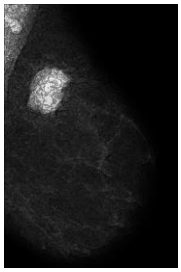
(a)Original image1



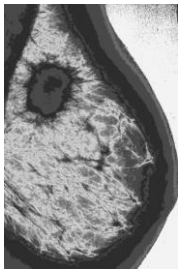
(b) Using Watershed



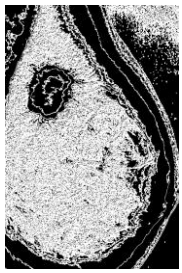
(c) Probability (GLCM)



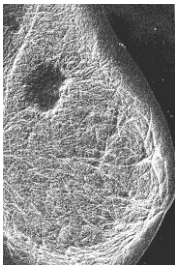
(d) Variance (GLCM)



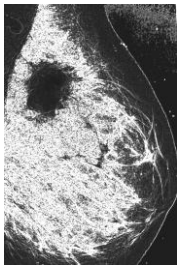
(e) Eq. Probability Image



(f) Variance of probability

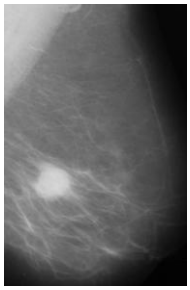


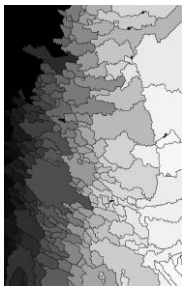
(g) Direct Variance

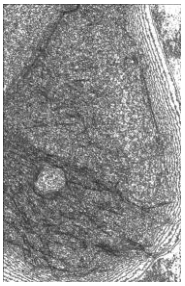


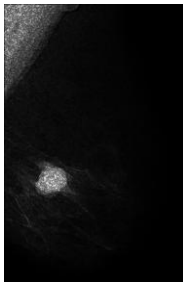
(h) Variance using probability

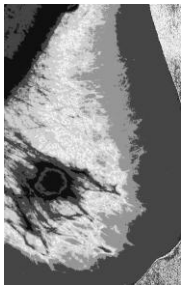
Figure 1 Results of Watershed, GLCM and proposed algorithm for image 1

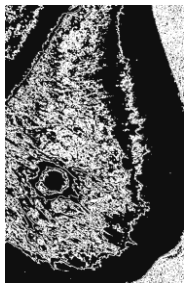

(a) Original image2

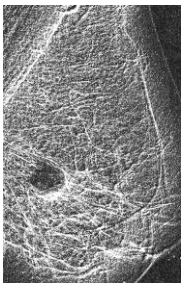

(b) Using Watershed


(c) Probability (GLCM)


(d) Variance (GLCM)


(e)Eq. Probability Image


(f) Variance of probability


(g) Direct Variance

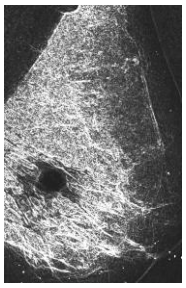

(h) Variance using probability

Figure 2.Results of Watershed, GLCM and proposed algorithm for image 2

Table 1: Performance comparison of GLCM, Watershed and Proposed Algorithm

Method	Tumor demarcation	Tumor details	Other details	Remarks
(1)GLCM				
Variance	Clear	Clear	Obscure	Acceptable
Probability	Clear	Clear	Obscure	Acceptable
(2)Watershed Algorithm	Clear	Not clear	Obscure	Over segmentation
(3)Proposed Algorithm				
Variance	Clear	Clear	Clear	**Recommended**
Eq. Probability	Clear	Clear	Obscure	Acceptable

IV. CONCLUSION

From Table 1 it can be inferred that GLCM method results are not very good but acceptable but have high computational complexity. As far as watershed algorithm is concerned the results are not acceptable because of over segmentation. The results of proposed methods using statistical parameters such as variance, probability are all

acceptable, amongst which direct variance method gives the best results for mammographic images which are verified by radiologist.

REFERENCES

[1] M. Unser, "Texture classification and segmentation using wavelet frames," *IEEE transaction on Image Processing*, vol. 4, no. 11, pp. 1549–1560, 1995.

[2] B.Wang and L. Zhang, "Supervised texture segmentation using wavelet transform," in *Proceedings of the 2003 International Conference on Neural Networks and Signal Processing*, 2003, vol. 2, pp. 1078–1082.

[3] Chang-Tsun Li and Roland Wilson, "Unsupervised texture segmentation using multiresolution hybrid genetic algorithm," in *Proc. IEEE International Conference on Image Processing ICIP03*, 2003, pp. 1033–1036.

[4] T.R. Reed and J. M. H. Du Buf, "A review of recent texture segmentation, feature extraction techniques," in *CVGIP Image Understanding*, 1993, pp. 359–372.

[5] A.K. Jain and K. Karu, "Learning texture discrimination masks," *IEEE transactions of Pattern Analysis and Machine Intelligence*, vol. 18, no. 2, pp. 195–205, 1996.

[6] Eldman de Oliveira Nunes and Aura Conci, "Texture segmentation considering multi band, multi resolution and affine invariant roughness.," in *SIBGRAPI*, 2003, pp. 254–261.

[7] L.F. Eiterer, J. Facon, and D. Menoti, "Postal envelope address block location by fractal-based approach," in *17th Brazilian Symposium on Computer Graphics and Image Processing*, D. Coppersmith, Ed., 2004, pp. 90–97.

[8] Haindl, M. & Mikeš, S. (2005). Colour texture segmentation using modelling approach. *Lecture Notes in Computer Science*, (3687), 484–491.

[9] Haindl, M. & Mikeš, S. (2006a). Unsupervised texture segmentation using multispectral modelling approach. In Y. Tang, S.Wang, D. Yeung, H. Yan, & G. Lorette (Eds.), *Proceedings of the 18th International Conference on Pattern Recognition, ICPR 2006*, volume II pp. 203–206. Los Alamitos: IEEE Computer Society.

[10] Haindl, M. & Mikeš, S. (2007). Unsupervised texture segmentation using multiple segmenters strategy. In M. Haindl, J. Kittler, & F. Roli (Eds.), MCS 2007, volume 4472 of Lecture Notes in Computer Science pp. 210–219.: Springer.

[11] Robert M. Haralick, Statistical and Structural Approaches to Texture, IEEE Proceedings Of vol. 67, no. 5, May 1979.

[12] Thomas P., William E., "Efficient Gabor Filter Design for Texture segmentation," Pattern Recognition 1996,Pattern Rec.Soc.

[13] Ojala, T. & Pietikainen, M. (1999). Unsupervised texture segmentation using feature distributions. *Pattern Recognition*, 32(477-486).

[14] **S.** M. Lai, X. Li, and W. F. Bischof, " On techniques for detecting circumscribed masses in mammograms," *IEEE Trans. Med. Zmag.*, vol. 8, no. 4, pp. 377-386, Dec. 1989.

[15] W. Qian, L. P. Clarke, M. Kallergi, and R. A. Clark, " Tree-structured nonlinear filters in digital mammography, *IEEE Trans. Med. Imag.*, vol.13, no. 1, pp. 25-36, Mar. 1994.

[16] D. Brzakovic, **X.** M. Luo, and P. lBzrakovic, "An approach to automated detection of tumors in mammography," *IEEE Trans. Med. Imag.*, vol. 9, no. 3, pp. 233-241, Sept. 1990.

[17] F. F. Yin, M. L. Giger, K. Dol, C. E. Metz, R. A. Vyborny, and C. J. Schmidt, "Computerized detection of masses in digital mammograms: Analysis of bilateral subtraction images,'' *Med. Phys.*, vol. 18, no. 5, pp. 955-963, Sept. 1991.

[18] T. K. Lau and W. F. Bischof, "Automated detection of breast tumors using the asymmetry approach,' *Comput. Biomed. Res.*, vol. 24, pp.273-295, 1991.

[19] W. P. Kegelmeyer Jr., **J.** M. Pruneda, P. D. Bourland, A. Hillis, M. W. Riggs, and M. L. Nipper, "Computer-aided mammographic screening for spiculated lesions," *Radiol.*, vol. 191, no. 2, pp. 331-337, May 1994.

[20] D. Marr and E. Hildreth, "Theory of edge detection," in Proceeding Royal Society, London., vol. 207, pp. 187-217 , 1980.

[21] **J.** Lunscher and M. P. Beddoes, "Optimal edge detector design: Parameter selection and noise effects," IEEE Trans. Pattem Anal. Machine Intell., vol. 8, no. 2, pp. 154-176, Mar. 1986.

[22] H. B. Kekre , Saylee Gharge, "Segmentation of MRI Images using Probability and Entropy as Statistical parameters for Texture analysis," Advances in Computational sciences and Technology(ACST),Volume 2,No.2,pp:219-230,2009, http://www.ripublication.com/acst.htm

[23] H. B. Kekre , Saylee Gharge , "Selection of Window Size for Image Segmentation using Texture Features," International Conference on Advanced Computing &Communication Technologies(ICACCT-2008) Asia Pacific Institute of Information Technology SD India, Panipat ,08-09 November,2008.

[24] H. B. Kekre , Saylee Gharge , "Image Segmentation of MRI using Texture Features," International Conference on Managing Next Generation Software Applications ,School of Science and Humanities, Karunya University, Coimbatore, Tamilnadu ,05-06 December,2008.

[25] H. B. Kekre , Saylee Gharge , "Statistical Parameters like Probability and Entropy applied to SAR image segmentation," International Journal of Engineering Research & Industry Applications (IJERIA), Vol.2,No.IV, pp.341-353.

[26] H. B. Kekre , Saylee Gharge , "SAR Image Segmentation using co-occurrence matrix and slope magnitude," ACM International Conference on Advances in Computing, Communication & Control (ICAC3-2009), pp.: 357-362, 23-24 Jan 2009, Fr. Conceicao Rodrigous College of Engg. Available on ACM portal.

[27] H. B. Kekre, Tanuja K. Sarode ,Saylee Gharge, "Detection and Demarcation of Tumor using Vector Quantization in MRI Images" ,International Journal of Engineering Science and Technology(IJEST),Volume 2,No.2,pp:59-66,2009.

[28] Serge Beucher and Christian Lantuejoul, "Use of watersheds in contour detection". In *International workshop on image processing, real-time edge and motion detection* (1979).

[29] Leila Shafarenko and Maria Petrou, " Automatic Watershed Segmentation of Randomly Textured Color Images", *IEEE Transactions on Image Processing*, Vol.6, No.11, pp.1530-1544, 1997.

[30] Basim Alhadidi, Mohammad H. et al, " Mammogram Breast Cancer Edge Detection Using Image Processing Function" Information Technology Journal 6(2):217-221,2007,ISSN-1812-5638.

AUTHORS PROFILE

Dr. H. B. Kekre has received B.E. (Hons.) in Telecomm. Engg. from Jabalpur University in 1958, M.Tech (Industrial Electronics) from IIT Bombay in 1960, M.S.Engg. (Electrical Engg.) from University of Ottawa in 1965 and Ph.D. (System Identification) from IIT Bombay in 1970. He has worked Over 35 years as Faculty of Electrical Engineering and then HOD Computer Science and Engg. at IIT Bombay. For last 13 years worked as a Professor in Department of Computer Engg. at Thadomal Shahani Engineering College, He has guided several B.E/B.Tech projects, more than 100 M.

E./M. Tech projects and 17 Ph. D. projects. Mumbai. He is currently Senior Professor working with Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS University, Vile Parle(w), Mumbai, INDIA. His areas of interest are Digital Signal processing and Image Processing. He has more than 350 papers in National / International Conferences / Journals to his credit. Recently ten students working under his guidance have received best paper awards. Two research scholars under his guidance have been awarded Ph. D. by NMIMS. Currently he is guiding ten Ph.D. students.

Ms. Saylee M. Gharge has received M.E.(Electronics and telecomm.) degree from Mumbai University in 2007, currently Pursuing Ph.D. from Mukesh Patel School of Technology, Management and Engineering, SVKM'S NMIMS University, Vile-Parle (W), Mumbai. She has more than 9 years of experience in teaching. Currently working as a lecturer in Department of Electronics and Telecommunication in Vivekanand Institute of Technology, Mumbai. Her areas of interest are Image Processing, Signal Processing. She has 32 papers in National /International Conferences/journal to her credit.

Enhancement of Passive MAC Spoofing Detection Techniques

Aiman Abu Samra and Ramzi Abed
Islamic University of Gaza
Gaza, Palestine
{aasamra, rabed}@iugaza.edu.ps

***Abstract*- Failure of addressing all IEEE 802.11i Robust Security Networks (RSNs) vulnerabilities enforces many researchers to revise robust and reliable Wireless Intrusion Detection Techniques (WIDTs).**

In this paper we propose an algorithm to enhance the performance of the correlation of two WIDTs in detecting MAC spoofing Denial of Service (DoS) attacks. The two techniques are the Received Signal Strength Detection Technique (RSSDT) and Round Trip Time Detection Technique (RTTDT). Two sets of experiments were done to evaluate the proposed algorithm. Absence of any false negatives and low number of false positives in all experiments demonstrated the effectiveness of these techniques.

Keywords: Intrusion Detection, RSS, RTT, Denial of Service.

I. INTRODUCTION

Due to the vast interest in WLAN technologies, these wireless networks have matured a lot since ratification of the first 802.11 standard in 1997 [2, 4]. Since then, several amendments have been made to the base standard[1], out of which most have been to the physical (PHY) layer to increase the operating speeds and throughput of WLANs [12]. However, one amendment -IEEE 802.11i was ratified in 2004 to address the threats of confidentiality, integrity and access control in WLANs [16].

As a result of the failure of the WLAN standards to address the lack of authentication of 802.11 Management frames and network card addresses, it is possible for adversaries to spoof the identity of legitimate WLAN nodes and take over their associations. Such attacks, where the attacker assumes the identity of another WLAN node, are referred to as MAC spoofing or simply spoofing based attacks. Such attacks are of grave concern as they can lead to unauthorized access and leakage of sensitive information. MAC spoofing is the root of almost all RSN attacks. Without the ability to inject forged frames using a spoofed MAC address, none of the RSN attacks can be launched. WIDS should be passive, accurate and sensitive [19]. Unfortunately, few intrusion detection techniques are available for reliably and accurately detecting MAC spoofing. The few that exist are not very robust and reliable. Given the enormous impact MAC spoofing has on WLAN security, wireless intrusion detection techniques are required to reliably and accurately detect MAC spoofing activity in WLANs.

II. WIRELESS INTRUSION DETECTION TECHNIQUES FOR MAC SPOOFING

The intrusion detection systems can be divided into two main categories depending on how their events of interest [7, 20, 21]:

Misuse-Based IDSs: Require that patterns representing security events be explicitly defined. This pattern is usually referred to as a signature. The IDS monitors computer systems and networks looking for these signatures and raises an alert when it finds a match.

Anomaly-Based IDSs: Anomaly-based IDSs on the other hand, do not require explicit signatures of security events. They use expected or non-malicious behavior and raise any deviations from this behavior as security events.

RSNs suffer from a number of security vulnerabilities; out of which the ability to spoof a WLAN node's MAC address is the most serious one. MAC spoofing allows an adversary to assume the MAC address of another WLAN node and launch attacks on the WLAN using the identity of the legitimate node. Without this vulnerability, an adversary will not be able to inject forged frames (Management, Control, EAP) into the WLAN and all attacks based on injection of such frames would be impossible [1]. Some of these attacks are Man-in-the-Middle, Session Hijacking, Rogue AP, Security Level Rollback, RSN IE Poisoning, EAP based DoS attacks, Management and Control frame based DoS attacks. Even exploiting the unprotected MAC frame Duration field to cause a DoS (virtual jamming) is also only possible in combination with MAC Spoofing. Software Implementation Based Attacks are also launched when an adversary injects forged frames containing exploit code into the WLAN using the MAC address of another WLAN node. The 4-Way Handshake Blocking and Michael Countermeasures DoS attacks are also launched using forged frames with spoofed MAC addresses.

The use of CCMP for confidentiality and integrity protection in RSNs has removed the threat of eavesdropping based passive attacks such as brute force and other key discovery attacks on the captured WLAN traffic. Hence, most attacks in RSNs are performed using active injection of forged frames into the WLAN using spoofed identity (MAC address)

[1] http://standards.ieee.org/getieee802/802.11.html

of other WLAN nodes. Even attacks that do not use MAC Spoofing directly exploit it in post attack activity. For instance, after an adversary has successfully discovered key material using the Dictionary Attack, it would use MAC Spoofing to authenticate to the WLAN using the key material and the MAC address of the victim node.

Hence, MAC Spoofing is responsible for majority of the attacks on RSNs. Spoofing based attacks in WLANS are possible as the existing WLAN standards fail to address the lack of authentication of unprotected WLAN frames and network card addresses. To further exacerbate the problem, almost all WLAN hardware provides a mechanism to change its MAC address; hence trivializing changing identities.

MAC Spoofing is the root cause of all injection based attacks on RSNs. A number of different techniques have been suggested to detect MAC spoofing activity in a WLAN. These are discussed below:

A. Sequence Number Monitoring:

This approach was first suggested by Wright [102] and was later used by Godber and Dasgupta [10] for detecting rogue APs. Kasarekar and Ramamurthy [13] have suggested using a combination of sequence number checks along with ICMP augmentation for detecting MAC spoofing. The idea is that an adversary spoofing the MAC address of a legitimate node will be assigned the same IP address as the legitimate node by the DHCP server of the WLAN. Hence, an ICMP ping to that IP address will return two replies; clearly identifying existence of MAC spoofing.

Guo and Chiueh [11] extended sequence number based MAC spoofing detection by monitoring patterns of sequence number changes. Rather than raising an alarm if a single sequence number gap is detected for a MAC address, the MAC address is transitioned to a verification mode and the subsequent sequence numbers of that MAC address are monitored for any anomalous gaps. In this manner, false positives raised due to lost and out of order frames are avoided. Their system also caches the last few frames for each MAC address to verify retransmissions and out of order frames.

Their solution also uses regular ARP requests to all STAs to synchronize with their sequence numbers based on ARP responses. This is done to defeat an adversary successfully injecting frames with correct sequence numbers somehow and detect the spoofing even if the legitimate node is no longer transmitting.

Madory [15] suggests a technique called Sequence Number Rate Analysis (SNRA) to detect MAC spoofing using sequence numbers. This technique calculates a transmission rate for a MAC address. If the calculated transmission rate is greater than the theoretical transmission limit for PHY of the WLAN it is considered to be an indication of a MAC spoof.

B. Fingerprinting

Fingerprinting MAC addresses based on their unique characteristics. The combination of device driver, radio chipset and firmware provides each WLAN node a unique fingerprint of its 802.11 implementation. Ellch [5] suggests using CTS frame responses and 802.11 Authentication and Association frames to fingerprint 802.11 implementations of WLAN nodes. He also suggests using the Duration field values in 802.11 frames to fingerprint WLAN nodes in a particular WLAN. Such fingerprints can be used to detect MAC spoofing activity as the fingerprint for the adversary would be different from the legitimate node. Franklin et al. [6] also suggest similar fingerprinting of 802.11 device drivers. Their technique exploits the fact that most 802.11 drivers implement the active scanning algorithm differently. They suggest that each MAC address could be mapped to a single device driver fingerprint and hence could be used for detecting MAC spoofing.

C. Location Determination

Location of the WLAN nodes can also be used to detect MAC spoofing. Location of a particular node is usually determined using its signal strength values as a location dependent metric.

Once the location of a MAC address is known, any changes in its location can be used as an indication of MAC spoofing activity. Bahl and Padmanabhan [4] record the received signal strength (RSS) values of each node on each AP and then compare these against a pre-calculated database that maps these RSS values to physical locations. Smailagic and Kogan [17] improve on this system and use a combination of triangulating WLAN nodes' RSS values from multiple APs and lookups in a database that maps RSS values to physical locations. Many other systems have also been proposed that establish location of a WLAN node using its RSS values and hence can be used for detecting MAC spoofing in a WLAN [4, 6,].

D. Signal Strength Fourier Analysis

Madory [15] also suggests a statistical technique called the Signal Strength Fourier Analysis (SSFA) to detect MAC spoofing using received signal strength (RSS) values of a WLAN node. It performs Discrete Fourier Transform on a sliding window of RSSs and uses the statistical variance of the high-frequencies which result from the interference between the attacker and the victim to detect MAC spoofing.

Some of the techniques for detecting spoofing based attacks have been implemented in some open source WIDSs such as Snort-Wireless [7]. Snort-Wireless claims to be capable of detecting MAC spoofing by monitoring for inconsistencies in MAC frame sequence numbers.

III. RELATED WORK

R. Gill et al. [8] address this issue by proposing two wireless intrusion detection techniques (WIDTs): Received Signal Strength Based Intrusion Detection Technique (RSSDT), and Round Trip Time Based Intrusion Detection Technique (RTTDT). These WIDTs are capable of detecting the spoofing based attacks reliably, and meet many of the desirable characteristics as they: are based on unspoofable characteristics of the PHY and MAC layers of the IEEE 802.11 standard; are passive and do not require modifications to the standard, wireless card drivers, operating system or

client software; are computationally inexpensive; and do not interfere with live traffic or network performance.

In [8] RSSDT and RTTDT work effectively against session hijacking attacks where the attacker and the legitimate STA are geographically separated and the differences in observed RSS and RTT between the attacker and the STA are significant. For more reliable intrusion detection, the RSSDT and RTTDT were not used in isolation from each other. Rather results from both the techniques were correlated to provide more confidence in the generated alarms [7].

IV. EXPERIMENTS

This section demonstrates the accuracy and utility of RSSDT and RTTDT through empirical data and use of correlation techniques. The RSSDT and the RTTDT, both use threshold values, namely the RSSdiff threshold and the RTTdiff threshold respectively. The RTTdiff and RSSdiff values greater than these thresholds are considered anomalous. In these experiments, RSSdiff threshold and the RTTdiff threshold were set to the value of 5.

To assist empirical analysis, eight experiment scenarios per attack were designed to study the effectiveness of the RSSDT and the RTTDT in the presence of an attacker, who launches three different new attacks against a legitimate station (STA); TKIP Cryptographic DoS attack [9, 18], Channel Switch DoS attack [14], and Quite DoS attack [14].

A. Equipment and Preparation

The experiments were carried out in a lab environment. The same networking hardware/software was used in all experiment scenarios. The following four parties took part in the scenarios: a legitimate client (STA), an access point (AP), a passive Intrusion Detection System (IDS) sensor, and an attacker.

B. Correlation Engine

Results from both the RTTDT and the RSSDT were correlated to provide more confidence in the generated alarms. The correlation engine used was event based i.e. if one of the detection techniques detected an anomaly (an alert), the correlation engine activated and waited until it obtained the detection results from the other technique, before making a decision on whether or not to raise an alarm. If both the detection techniques detected the anomaly, then an alarm will be raised.

C. Experimentation -Set1

In *Set1* of the experiments, robustness and reliability of the RTTDT and the RSSDT were tested when none of the participants were in motion. The AP, the IDS sensor, the STA and the attacker were all stationary in this set. In all scenarios, the AP was placed in close proximity to the IDS sensor. In Fig. 1, points A, B and C represent location of the STA, the AP and the IDS sensor respectively. Experimentation set1 has 4 scenarios. Points X,Y and Z represent location of the attacker in Scenarios Two, Three and Four respectively. While scenario four has no attaker.

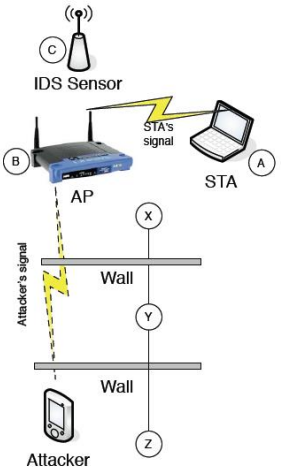


Figure 1. Experimentation Set-1

1) Scenario One

In *Scenario One*, there was no attacker present and the AP and the STA were placed in close proximity to each other at points B and A respectively (see Fig. 1). Network traffic was generated from the STA to the AP. The IDS sniffer was used to capture this WLAN traffic between the STA and the AP. After examination of 1000 captured frames, the *correlation engine* did not raise any alarms. As there was no attacker present, both the detection techniques and the correlation engine correctly did not generate any false positives.

2) Scenario Two

In *Scenario Two*, the AP and the STA were placed in close proximity to each other at points B and A respectively. The attacker was placed in line of sight of the STA at point X (Fig. 1). Then network traffic was generated between the STA and the AP. The attacker then launched the attack on the STA.

In this scenario three different experiments were carried out; in the first one, the attacker launched a TKIP DoS Attack [9, 18], in the second experiment he launched a Channel Switch DoS attack [14], while in the third experiment the attacker launched a Quite DoS attack [14]. For each experiment, traffic was captured using *Wireshark*[2], after that the captured traffic was examined using the IDS Sensor which is based on the *correlation engine*, which resulted in two alarms.

3) Scenarios Three and Four

Similar to scenario two except the location of the attaker it is in point Y for scenario three where three alarms were generated when running. And in point Z for scenario four where tow alarms were generated.

D. Experimentation -Set2

In experiments *Set2*, the robustness and reliability of the RTTDT and the RSSDT were tested with the attacker stationary and the STA in motion between a point closer

[2] http://www.wireshark.org

to the AP and another point far away from it. The AP, the IDS sensor, and the attacker were all stationary at locations B, C and D (or G) in Fig. 2. In all scenarios, the IDS sensor was placed in close proximity to the AP.

1) Scenario Five

In *Scenario Five*, the AP and the attacker were stationary and were placed in close proximity to each other (in line of sight at points B and D in Fig. 2). Network traffic was then generated from the STA to the AP. The STA then started traveling (at walking pace) from a point close to the AP to a point far away from it (i.e. from point E to F in Fig. 2). Towards the end of the STA's journey, the attacker then launched three different attacks on the STA as described in Scenario Tow. After capturing the traffic; executing IDS Sensor over the captured traffic resulted in two alarms.

2) Scenario Six

Similar to scenario Five except the STA waking from point F to E and one alarm was raised.

3) Scenario Seven

In *Scenario Seven*, the AP and the attacker were stationary and were placed far away from each other (not in line of sight, at points B and G in Fig. 2). Then similar to scenario Five with one alarm.

4) Scenario Eight

Similar to scenario Seven except the STA waking from point F to E. Also one alarm was raised.

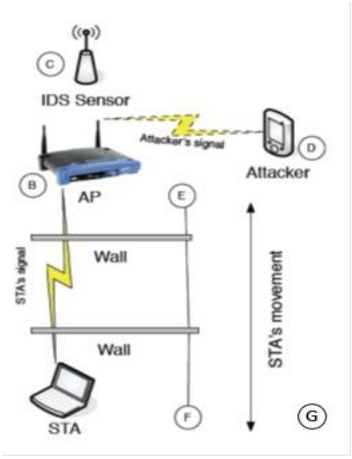


Figure 2. Experimentation Set-2

V. ANALYSIS

True Positives and False Positives

In our experiments, no false negatives were registered. However, some false positives were raised by the correlation engine.

Tables 1, 2, and 3 summarize the true positives raised by the correlation engine when applying TKIP DoS attack, Channel Switch DoS attack, and Quite DoS attack respectively in all eight scenarios. For instance, the entry for Scenario Two in Table 1 shows that when applying the first attack, the true alarm was raised by the correlation engine at frame 499. This alarm was caused by a RSS fluctuation (RSSdiff =19) at frame 499 and a RTT spike at frame 510 (RTTdiff =29.651) for the STA. Frame 510 was the very next RTS-CTS handshake event for the STA after frame 499. Hence, both the RSSDT and the RTTDT sensors reported the anomaly and the TKIP DoS attack was identified correctly and accurately. Also, the entry for Scenario Two in Table 2 shows that when applying the second attack in Scenario Two, the true alarm was raised by the correlation engine at frame 520. This alarm was caused by a RSS fluctuation (RSSdiff =16) at frame 520 and a RTT spike at frame 543 (RTTdiff =25.556) for the STA. Frame 543 was the very next RTS-CTS handshake event for the STA after frame 520. Hence, both the RSSDT and the RTTDT sensors reported the anomaly and the Channel Switch DoS attack was identified correctly and accurately. Moreover, the entry for Scenario Two in Table 3 shows that when applying the third attack, the true alarm was raised by the correlation engine at frame 602. This alarm was caused by a RSS fluctuation (RSSdiff =16) at frame 602 and a RTT spike at frame 620 (RTTdiff =26.447) for the STA. Frame 620 was the very next RTS-CTS handshake event for the STA after frame 602. Hence, both the RSSDT and the RTTDT sensors reported the anomaly and the Quite DoS attack was identified correctly and accurately.

In *Scenario One*, *Scenario Two*, *Scenario Three* and *Scenario Four*, as expected, the *RSSdiff* and *RTTdiff* values increased as the attacker was placed further away from the STA. In *Scenario Five* and *Scenario Six*, the AP, the IDS sensor and the attacker were located in close proximity of each other and as expected, the *RSSdiff* and *RTTdiff* values increased as the STA moved away from them and decreased as the STA moved closer. In *Scenario Seven* and *Scenario Eight*, the attacker was located further away from the IDS sensor and the AP. The observed *RTTdiff* and *RSSdiff* values increased as the STA moved away from the attacker, and decreased as it moved closer to the attacker (see Tables 1, 2, and 3).

An interesting observation was made that in *Scenarios Two*, *Scenario Three* and *Scenario Four*, where all parties were stationary; all the false positives were detected in frames generated after the attack had commenced (Tables 1 - 6). This means that all the false positives were caused by abnormal fluctuations in observed RSS and RTT values for the attacker. This observation was most likely the result of increasing distance between the attacker and the passive IDS monitor from *Scenario Two* to *Scenario Four*. Lack of line of sight connectivity and presence of various obstacles (walls, doors etc.) most likely acted as contributing factors to random fluctuations in observed RSS and RTT values for the attacker. Being positioned in close proximity of the sensor in all these scenarios, the STA did not suffer such random fluctuations and hence did not generate any false positives before the attack was launched.

However, in *Scenario five* to *Scenario Eight*, just the opposite was observed. The false positives were detected in frames generated before the attack had commenced, which meant that the source of these abnormalities was the STA and not the attacker. In these scenarios, the attacker was always stationary and the STA was in motion. These false positives

can be attributed to the fluctuations in observed RSS and RTT values for the STA as a result of it being in motion.

The correlation technique successfully managed to keep the number of these false positives fairly low. The RSSDT and the RTTDT both successfully detected the performed attacks.

TABLE 1. True Positives for TKIP DoS Attack experiments

Scenario	RSS diff	Frame number	RTT diff	Frame number
One	NA	NA	NA	NA
Two	19	499	29.651	510
Three	34	550	39.204	585
Four	47	606	51.014	657
Five	32	554	43.751	590
Six	23	622	31.098	634
Seven	27	530	38.773	549
Eight	30	583	40.007	603

TABLE 2. True Positives for Channel Switch DoS Attack experiments

Scenario	RSS diff	Frame number	RTT diff	Frame number
One	NA	NA	NA	NA
Two	16	520	25.556	543
Three	37	603	40.002	630
Four	42	583	49.099	599
Five	27	532	40.071	545
Six	26	601	37.655	628
Seven	27	530	38.773	549
Eight	30	583	40.007	603

TABLE 3. True Positives for Quite DoS Attack experiments

Scenario	RSS diff	Frame number	RTT diff	Frame number
One	NA	NA	NA	NA
Two	16	602	26.447	620
Three	34	603	39.077	626
Four	38	593	47.014	639
Five	30	495	38.977	512
Six	28	598	36.007	620
Seven	22	617	31.055	633
Eight	26	589	32.411	603

Fig.s 3, 4, and 5 show the distribution of true positives and false positives registered by the RSSDT and the RTTDT when running the TKIP DoS Attack, the Channel Switch Attack, and the Quite Attack experiments respectively. It is clear from these Fig.s that the used techniques are effective in detecting the applied attacks because of the absence of false negatives.

Single Anomalies

A single anomaly would occur if a RSS alert was registered by the RSSDT, while the RTTDT did not register an anomaly in the next RTS-CTS event for that MAC address. Another example would be if a RTT alert was raised by the RTTDT but the next RSS reading for that MAC address was below the threshold. The RSSDT and the RTTDT only raise an alert if the difference between the last observed and current characteristic is above a threshold. Single anomalies were ignored by the correlation engine and an alarm was only raised if both the detection techniques register an alert.

Fig. 3, 4, and 5 show the distribution of single anomalies registered by the RSSDT and the RTTDT when running the TKIP DoS Attack, the Channel Switch Attack, and the Quite Attack experiments respectively, where the correlation engine did not raise any alarm.

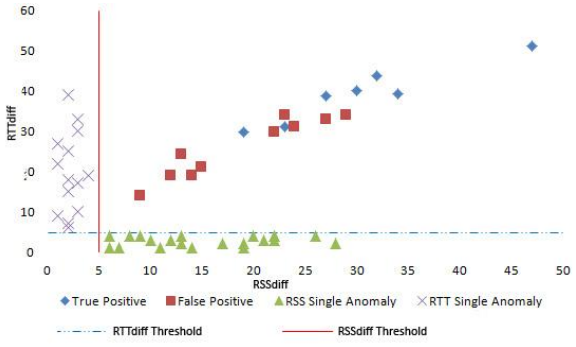


Figure 3. Alarms and Single Anomalies for TKIP DoS Attack Experiment

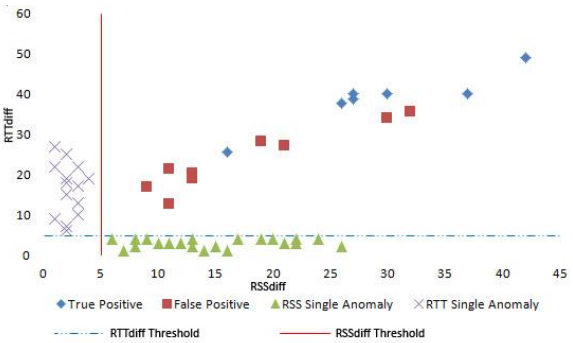


Figure 4. Alarms and Single Anomalies for Channel Switch DoS Attack Experiment

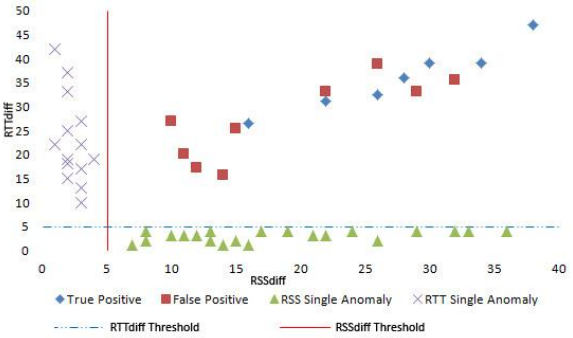


Figure 5. Alarms and Single Anomalies for Quite DoS Attack Experiment

TABLE 4. False Positives for TKIP DoS Attack experiments

Scenario	RSS diff	Frame number	RTT diff	Frame number
One	NA	NA	NA	NA
Two	14	730	19.011	744
Three	24 27	698 811	31.112 33.009	718 840
Four	23	800	34.093	832
Five	13	389	24.225	411
Six	12	370	18.999	393
Seven	15	278	21.112	307
	9	396	14.001	420
Eight	29 22	290 340	34.112 29.901	312 366

TABLE 5. False Positives for Channel Switch DoS Attack experiments

Scenario	RSS diff	Frame number	RTT diff	Frame number
One	NA	NA	NA	NA
Two	13	698	20.332	709
Three	30 32	721 826	33.878 35.552	754 865
Four	19	748	28.221	773
Five	11	335	21.222	378
Six	9	293	16.988	316
Seven	13 11	271 377	18.909 12.727	300 399
Eight	21	389	27.101	409

TABLE 6. False Positives for Quite DoS Attack experiments

Scenario	RSS diff	Frame number	RTT diff	Frame number
One	NA	NA	NA	NA
Two	11	765	19.991	789
Three	29 32	777 870	33.117 35.550	808 897
Four	26	781	38.887	800
Five	10	278	27.002	301

Six	15	390	25.333	409
Seven	22	410	33.011	437
Eight	12 14	300 378	17.112 15.553	329 404

VI. THRESHOLD OPTIMIZATION

In experiments discussed above the RSSdiff threshold and the RTTdiff threshold were set to an initial constant value of 5. This means a RSS anomaly was only registered if the RSSdiff was greater than 5 and a RTT anomaly was only acknowledged if the RTTdiff value was greater than 5. An alarm was raised by the correlation engine only when both the RSSdiff threshold and the RTTdiff threshold were exceeded. The threshold value 5 was thought to be just low enough to avoid a high number of false negatives and just high enough to avoid a large volume of false positives. Since ideally both the techniques should exhibit the same level of accuracy, the same threshold value was used for both. In these experiments (in sections 4.3 and 4.4), both the thresholds were set to the same value, in fact there is no need for the RTTdiff threshold and the RSSdiff threshold to be with the same value.

In reality, we need to optimize these threshold values to ensure the lowest possible number of false positives and false negatives. Choosing the best threshold value for each detection technique can be performed using the algorithm presented in Fig. 6.

Figure 6. RSSdiff threshold and RTTdiff threshold optimization algorithm

Table 7 represents the optimized thresholds after applying the algorithm in Fig. 6 on the results the experiments section. Referring to Table 7 we found that the optimized thresholds are very close for the three attacks experiments. In our opinion, from a general intrusion detection perspective, it is far more critical for an IDS to minimize the false negative rate than to maintain a low false positive rate. The cost of missing an attack is much higher than the cost of raising a false alarm. Therefore, we choose the minimum threshold to avoid the false negatives i.e. 15 for RSSdiff threshold and 25 for RTTdiff threshold. Fig.s 3, 4, and 5 represent the true positives, false positives and single anomalies registered by the IDS when using the initial threshold settings. In Fig.s 3, 4, and 5; RSSdiff Threshold and RTTdiff Threshold refer to initial values used for thresholds (i.e. 5 for all scenarios).

Fig. 7, 8, and 9 represent the true positives, false positives and single anomalies registered by the IDS when using the optimum threshold settings. In Fig. 7, 8, and 9, RSSdiff Threshold and RTTdiff Threshold refer to the optimized values of RSSdiff and RTTdiff thresholds respectively. These values were generated by applying the Algorithm in Fig. 6 to minimize the number of false positives and false negatives.

Table 7 demonstrates that RSSdiff threshold of 15 and RTTdiff threshold of 25 which are the optimum choice for the thresholds of the detection techniques. Fig.s 7, 8, and 9 show how the single anomalies, false positives and true positives are affected by the new optimum threshold values. As a result of the new thresholds, some false positives became RSS single anomalies or RTT single anomalies. The new thresholds did not introduce any false negatives since there are no true positives became false negatives. Moreover, no single anomaly was converted into a false positive as a result of the new threshold. In fact, some of the single anomalies (both RSS and RTT single anomalies) became normal events. Hence with 100% true positive detection, RSSdiff Optimized Threshold of 5 and RTTdiff Optimized Threshold of 25 is proved to be the optimum threshold values for the resented test scenarios.

1. *Read the captured packets from the dump file*
2. *Filter the points according to their class (true positive, false positive…)*
3. *Let* $F_{min}=(RSSdiff_{min}, RTTdiff_{min})$ *be the frame of class "true positive" such that it has the least RSSdiff and RTTdiff values*
4. *If* $RSSdiff_{min} \epsilon \mathbb{N}$

   Set the optimized RSSdiff threshold = $RSSdiff_{min} - 1$

   Else If $RSSdiff_{min} \epsilon \mathbb{R}$

   Set the optimized RSSdiff threshold = $\lfloor RSSdiff_{min} \rfloor$
5. *If* $RTTdiff_{min} \epsilon \mathbb{N}$

   Set the optimized RTTdiff threshold = $RTTdiff_{min} - 1$

   Else If $RSSdiff_{min} \epsilon \mathbb{R}$

   Set the optimized RTTdiff threshold = $\lfloor RTTdiff_{min} \rfloor$

Where $\mathbb{N}$ is the set of all natural numbers, and $\mathbb{R}$ is the set of all real numbers

TABLE 7. Optimized RSSdiff and RTTdiff thresholds

Experiment	**New *RSSdiff* threshold**	**New *RTTdiff* threshold**
TKIP DoS Attack	18	29
Channel Switch DoS Attack	15	**25**

Quite DoS Attack	**15**	26

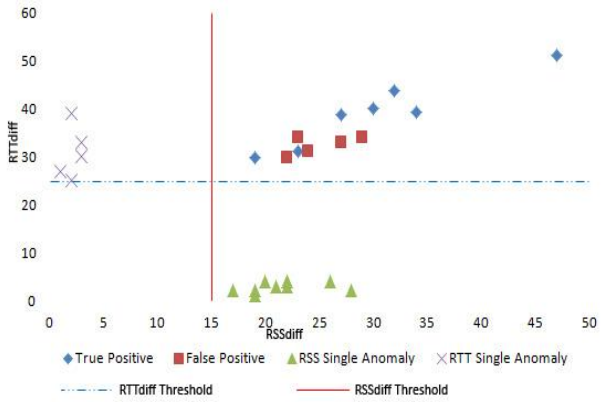


Figure 2. Alarms and Single Anomalies for TKIP DoS Experiment when applying the optimized threshold

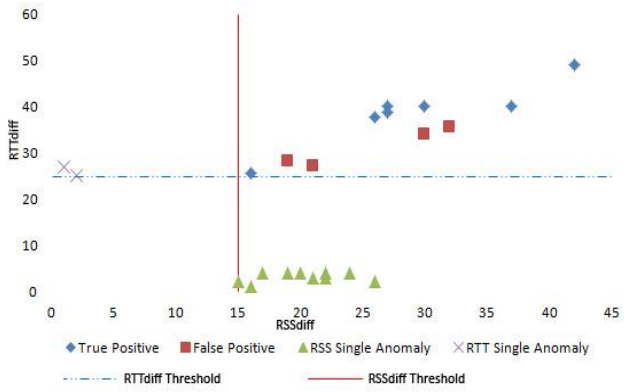


Figure 3. Alarms and Single Anomalies for Channel Switch DoS Experiment when applying the optimized threshold

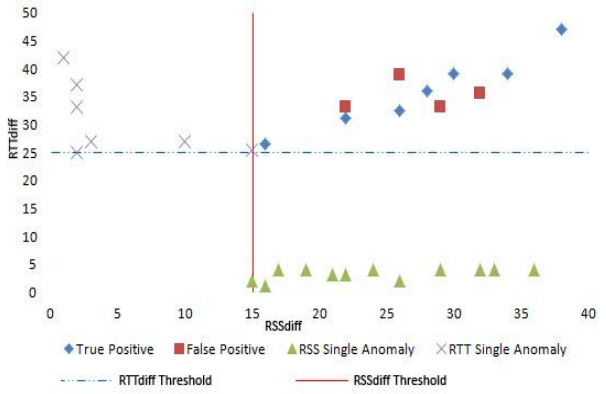


Figure 4. Alarms and Single Anomalies for Quite DoS Experiment when applying the optimized threshold

Accuracy and efficiency of the RSSDT and the RTTDT depends on the choice of suitable threshold values and hence places a large expectation on these threshold values to be optimally calculated. This increases the importance of our developed algorithm (Fig. 6).

Thresholds are unique to each WLAN environment and can also change frequently. Hence, the thresholds should be regularly calculated to optimum values. Using a distributed approach and deploying multiple distributed co-operating IDS sensors can decrease this expectation on the accuracy of the threshold values. Rather than relying on the alarms generated by a single IDS sensor, the intrusion detection process can be enhanced by correlating detection results across multiple sensors.

This also makes it a much harder job for the attacker to launch a successful spoofing attack as they will have to guess and spoof the RSS and the RTT values for the legitimate nodes, as observed by each IDS sensor. This will require the attacker to be at multiple locations at the same time, hence making it very hard for the attacker to launch an undetected attack.

VII. Conclusion

Despite using a number of preventative security measures, IEEE 802.11i RSNs still suffer from multiple vulnerabilities that can be exploited by an adversary to launch attacks. This underlines the need for using a monitoring framework as a second layer of defense for WLANs. Such monitoring capability can be implemented using a wireless intrusion detection system.

This paper verifies the effectiveness of RSSDT and RTTDT wireless intrusion detection techniques that address majority of RSN attacks. The paper proposed an algorithm to enhance the performance of the correlation of the Received Signal Strength Detection Technique (RSSDT) and Round Trip Time Detection Technique (RTTDT) in detecting MAC spoofing Denial of Service (DoS) attacks. The proposed algorithm is enhanced the performance by optimizing the value of the detection threshold.This paper also demonstrates that the detection results can be correlated across the WIDS sensors and also the detection techniques themselves to provide greater assurance in the reliability of the alarms and enable automatic attack scenario recognition.

The experiments presented in Section 2 demonstrate the feasibility of using RSS and RTT monitoring as wireless intrusion detection techniques since they did not produce any false negatives, while the correlation between the RSSDT and RTTDT and the self-adaptation for both RSSDT and RTTDT thresholds results was feasible in lowering the number of false positives.

References

[1] B. Aboba, L. Blunk, J. Vollbrecht, J. Carlson, and H. Levkowetz. Extensible Authentication Protocol (EAP). The Internet Engineering Task Force-Request for Comments, RFC 3748, 2004.

[2] R. Ahlawat and K. Dulaney. Magic Quadrant for Wireless LAN Infrastructure, 2006. Gartner Research, 2006.

[3] J. Anderson. Computer Security Threat Monitoring and Surveillance. Report 79F296400, James P.Anderson Co., 1980.

[4] P. Bahl and V. Padmanabhan. RADAR: an in-building RF-based user location and tracking system. INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, 2, 2000.

[5] J. Ellch. Fingerprinting 802.11 Devices. PhD thesis, Naval Postgraduate School; Available from National Technical Information Service, 2006.

[6] J. Franklin, D. McCoy, P. Tabriz, V. Neagoe, J. Van Randwyk, and D. Sicker. Passive Data Link Layer 802.11 Wireless Device Driver Fingerprinting. Proceedings of the 15th Usenix Security Symposium, 2006.

[7] R. Gill, J. Smith, and A. Clark. Experiences in Passively Detecting Session Hijacking Attacks in IEEE 802.11 Networks. In R. Safavi-Naini, C. Steketee, and W. Susilo, editors, Fourth Australasian Information Security Workshop (Network Security) (AISW 2006), volume 54 of CRPIT, pages 221–230, Hobart, Australia, 2006. ACS.

[8] R. Gill, J. Smith, M. Looi, and A. Clark. Passive Techniques for Detecting Session Hijacking Attacks in IEEE 802.11 Wireless Networks. In Proceedings of AusCERT Asia Pacific Information Technology Security Conference (AusCert 2005): Refereed R&D Stream, pages 26 – 38, 2005.

[9] S. Glass and V. Muthukkumarasamy. A Study of the TKIP Cryptographic DoS Attack. 15th IEEE International Conference on Networks, 2007.

[10] A. Godber and P. Dasgupta. Countering rogues in wireless networks. International Conference on Parallel Processing Workshops, pages 425–431, 2003.

[11] F. Guo and T. Chiueh. Sequence Number-Based MAC Address Spoof Detection. In 8th

[12] IEEE. IEEE Standard for Information technology-Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 6: Medium Access Control (MAC) Security Enhancements, 2004. Institute of Electrical and Electronics Engineers.

[13] V. Kasarekar and B. Ramamurthy. Distributed hybrid agent based intrusion detection and real time response system. First International Conference on Broadband Networks (BroadNets 2004), pages 739–741, 2004.

[14] B. Konings, F. Schaub, F. Kargl, and S. Dietzel. Channel switch and quiet attack: New DoS attacks exploiting the 802.11 standard. IEEE 4th Conference on Local Computer Networks, 2009.

[15] D. Madory. New Methods of Spoof Detection in 802.11 b Wireless Networking. PhD thesis, Dartmouth College, 2006.

[16] NIST. Establishing Wireless Robust Security Networks: A Guide to IEEE 802.11i, Feb 2007. National Institute of Standards and Technology. Special Publication 800-97. [Online] Available: http://csrc.nist.gov/publications/ nistpubs/800-97/SP800-97.pdf

[17] A. Smailagic and D. Kogan. Location sensing and privacy in a context-aware computing environment. Wireless Communications, IEEE Personal Communications, 9(5):10–17, 2002.

[18] Tkiptun-ng, 2010, [Online] Available: http://www.aircrack-ng.org

[19] NSA. Guidelines for the Development and Evaluation of IEEE 802.11 Intrusion Detection Systems (IDS), Nov 2005. National Security Agency. Network Hardware Analysis and Evaluation Division.[Online] Available:http://www.nsa.gov/notices/notic00004.cfm?Address=/snac/wireless/I332-005R-2005.pdf

[20] R. Bace and P. Mell. NIST Special Publication on Intrusion Detection Systems. National Institute of Standards and Technology, draft document, February, 2001.

[21] H. Debar and J. Viinikka. Intrusion detection: Introduction to intrusion detection and security information management. FOSAD, 2005:2005, 2004.

A framework for Marketing Libraries in the Post-Liberalized Information and Communications Technology Era

Author: Garimella Bhaskar Narasimha Rao
Associate Professor, MBA Department
DIET, Anakapalle, Visakhapatnam, India

***Abstract*- The role of library is professional and is fast customizing to altering technological platforms. Our subscriber's perceptions of the nature of our libraries are also altering like never before – particularly in Universities. In this dynamic environment the efficient presentation of the library is an essential survival tool. This holds true whether a library is contributing effectively to an institution's overall management effort or, sensitively and subtly, promoting rational attitudes in the minds of its local constituency. Citations, references and annotations referred here are intended to provide background, ideas, techniques and inspiration for the novice as well as the experienced personnel. We hope you will find this paper as interesting and useful as we trust it would be, and that libraries may gain in useful collaborations and reputation from the application of information provided by resources identified here. Result oriented marketing is not on its own - some kind of substitute for a well-run service provider or service the demands of its user population, but in the 21st century even the best-run centers of learning or information service will only prosper if effort and talent are devoted to growth orientation. Almost all of us can render to this effort, and this paper can help coax the roots of marketing talent into a complete, harvestable and fragrant flower.**

Keywords-Digital Libraries, University Learining, Adaptability

I. INTRODUCTION

This paper is intended to serve as a source of inspiration, innovation and resources for librarians who are setting out to administer and roll-out projects in a fast-moving, ever-changing and potentially dynamic arena. A new perspective, addressing the marketing of library information, products and services has been moving up in the chronological list of challenges to be addressed on a priority. The whole area of marketing and "selling the library" is a concept which many information professionals still feel nests uneasily with our core professional values of service and empathy. Yet, the library and information science (LIS) professionals can no longer look at reputed publications which emphasize this introduction and think it has no bearing on our own.

II. BACKGROUND & RELATED WORK

Ghosh Maitrayee (March 2009). This research presented the developments in ETD repositories and performed a preliminary study to explore the possibilities for creating a national repository for the submission, identification, use and long-term care of research theses in an open access environment. The authors look at the current state of deployment of ETD repositories in the academia and discuss the subject coverage, number of items, access policy, search/browse option, and value added services. This study posed questions about policies and strategies that, research funding, national higher education and policy-making bodies, as well as individual institutional communities within the higher education might want to consider. This research paper was titled "A case study of nine ETD digital libraries and formulation of policies for a national service and published in the International Information & Library Review, Volume 41, Issue 1.

Bárbara L. Moreira, Gonçalves A. Marcos, Laender Alberto H.F, Fox Edward A. (April 2009).

This research work which was published under the title "Automatic evaluation of digital libraries with 5SQual and published in the Journal of Informetrics, Volume 3, Issue 2 was targeted to alter the situation with the creation of 5SQual, a tool which provides ways to perform systemized and configurable evaluations of some of the most important Digital Library components, encompassing metadata, digital objects and services. In total, the main contributions of this work include: (i) the applicability of the application in several usage scenarios; (ii) the implementation and design of 5SQual; and (iii) the evaluation (including specialists usability) of its graphical interface specially designed to guide the configuration of 5SQual evaluations. The results present the analysis of interviews conducted with administrators of real Digital Libraries regarding their opinions and expectations about 5SQual.

Frias-Martinez Enrique, Chen Sherry Y., Liu Xiaohui (February 2009).

This research asserts that, personalization can be accounted for by adaptability and adaptivity, which have different pros and cons. This study investigated how digital library (DL) users react to these two techniques. The research was titled "Evaluation of a personalized digital library based on cognitive styles": Adaptivity vs. Adaptability was published in

the International Journal of Information Management, Volume 29, Issue 1 and the authors developed a personalized DL to address the needs of different cognitive styles based on the findings of their previous work [Frias-Martinez, E., Chen, S. Y., & Liu, X. (2008) Investigation of behavior and perception of digital library users: A cognitive style perspective: International Journal of Information Management. The personalized DL included two versions: adaptable version and adaptive version. Results reflected that users not only performed better in the adaptive version, but also perceived more positively to the adaptive version. The primary assertion and conclusion was that cognitive styles have great effects on users' responses to adaptivity and adaptability. Results provided were aimed to render guidance for designers to select suitable techniques to structure personalized DLs.

Porcel C., Moreno J.M., Herrera-Viedma E. (December 2009)

The main theme was that the internet is one of the most important information media and it is influencing in the development of other media, as for example, newspapers, journals, books, and libraries. In this paper, The authors analyzed the logical extensions of traditional libraries in the Information Society. and present a model of a fuzzy linguistic recommender system to help the University Digital Libraries users to access for their research resources. This system recommended researchers specialized in complementary resources to discover collaboration possibilities to form multi-disciplinar groups. The research was titled "a multi-disciplinar recommender system to advice research resources in University Digital Libraries" and published in the Journal of Expert Systems with Applications, Volume 36, Issue 10.

III. Domains That Require Analysis – A Perspective Of Challenges

The Digital Library of a University requires a unified portal that would enable comprehensive searching and access rights to all library electronic resources from a single entry point. Disparate purchasing, management, and work related policies have previously led to a need to lower user confusion. The library with a sophisticated system to manage the different licensing agreements and access rights matrices of various units.

The Library of a University primarily serves the purpose of providing access to its complete collection. The implemented technology platform assists and enables the search and the mentioned access to electronic and digital publications. Based on centuries of educational tradition (like the University of Taxila, India), a large University student body imbibes a resident city with a youthful and lively character. Numerous graduate, postgraduate students and teaching staff make use of a University's huge library information system--based on the network of various branches and allied libraries for the automation and access to print collections. Once the basic project of collating and establishing the printed books collection gets over, a librarian needs to turn his focus towards the management of the libraries electronic resources.

When electronic resources first became available, many librarians had to address the challenges of uniform purchasing, management, and the creation of usage policies for academic libraries that were served. As a result, each university faculty was required to provide a list of publications and e-resources relevant to its area of study that needed purchasing. These materials were often restricted to the users associated with that specific faculty, and were only made available to them via a network or any specific other search systems developed by that department. A few of these systems assisted virtual private network access to the library resources while others required users to access the resources from within the library.

Researchers wishing to browse through any library resources could be faced with different entry points and interfaces rather than a consolidated listing of all library holdings. This left users frustrated and led to under-use of many valuable high-cost resources. This quickly established that a unified library portal, via which users could perform a federated searching across all of the resources, was deemed essential. Drawing up a list of requirements for choosing a library solution that would help us to meet the goals of the digital library project, would be; on one hand, libraries need systems sophisticated enough to accommodate the different copyright agreements and access rights entitlements of various converging units. On the other hand, the system has to provide a single point of entry through which all users, regardless of access authority, who access the library's complete collection displayed via a list and a catalog of resources.

The University library exists to benefit a diverse set of people. The associates mentioned here are the societies, undergrads to post grads, teaching and non-teaching fraternity. These groups may be further be segmented by roles, characteristics, or interests, members for whom, English is a second language for the purposes of specific marketing activities. As an institution which is established from student fees and non-governmental grants and as a repository for government information, librarians also have an added responsibility to serve the local society. Again, within the local society, there are multiple subsets that could serve our local society. Again within the local society there are multiple subsets that could be of particular interest to librarians as they market services and resources. It is here that, technology makes an impact and assists selective promotion on a global platform. Research Scholars interested in a library's unique collections may also be audience for a library's long-term marketing plan. In this context it is apt to quote that, the university libraries have been presented both in the library profession and in the higher education segment as a showcase for its architecture and information technologies.

IV. Libraries In The Post Liberalized Era – A Few Thought Starters

Libraries are deemed to posses reliable and convenient information selected to meet a users research needs and learning styles. Also libraries have unique and valuable regional materials on issues of national interest such as rural development, and crops management in a dry environment, mobile computing, nuclear power generation architecture and health development to name a few.

Libraries assist and cater to the development of literacy skills crucial for extended learning and professional success, namely the ability to ascertain and access, evaluate and effectively utilize information in all of its various formats. Helpful library staff are to be made available and accessible within the facility and remotely, to collaborate with patrons in a one on one set-up or in group settings, to provide subject specific or general assistance. Association with the library in the form, of an ID (physical or electronic or biometric) can help in authenticating users. Libraries cater to their users with sound-proof venues, inviting with comfortable spaces to read, research, study, relax, explore, learn, assimilate with convenient access to print (Paid and free) and electronic information with the latest hardware and software that today's users need and expect.

For librarians engaged in strategic planning encompassing information and communications technology (ICT), guidance from confirmed sources can result valuably. There can unfortunately be a great deal of challenges associated with much written components regarding the strategic use to which modern ICT can be utilized in organizations throughout the Globe these days. Re-sounding yet vague terms such as 'Information Technology Enabled Services' and 'Business Process Outsourcing' frequent the popular media channels on the subject. The Journal of Strategic Information Systems takes a more critical outlook. In publishing papers authored from scholarly study and utilizing research from all nooks and corners of the globe, modern libraries provide a truly impressive array of information on how centers of learning can utilize ICT strategically to avoid associated pitfalls.

To identify the importance of the marketing plan for the overall success of a library and emphasize the need to include a marketing plan within a library's strategic plan. Further we need to identify components to include in a marketing plan and hence provide a procedure book to generate such a plan. Quoting many eminent librarians, components essential to a marketing plan encompass determining what to promote, defining a target audience, choosing types of outreach, and evaluation.

Publicity Guidelines: Many companies use public relations to assist other promotional efforts. It is a very potent tool that could always be part of a library's marketing mix. The specified 5-step guideline can assist a librarian's public relations efforts and can closely coordinate with promotional campaigns so as to achieve a greater impact.

1. Deliberate on and choose the right media.
2. Do have a good grounding on the media type.
3. Timing & Packaging your news is pivotal
4. Ensure you follow-up in person.
5. Establish a strong relationship and cement the same.

For a library to market itself in the Post Liberalized world, it needs to have a focus to achieve some or all of the goals identified as below.

Goals of a Library - Not limited to

1. Sponsorship of events, aimed primarily at students and teachers to promote the library's resources and services and to address library anxiety at the start of each session as a follow-up to participation in orientation activities.

a. To solicit for volunteers twice a year to form event planning teams lead by a member from the corporate communications committee.

b. Choose a key theme for each such event.

c. Establish monetary budgets for the event and associate promotional materials.

2. Collate a rule book in conjunction with allied departmental libraries, and an informational guide for students and users as an annual publication for donors and community leaders and associates of the library highlighting the accomplishments, thanking donors and gift contributors.

a. Seek general suggestions / comments from

i. Within departments of the library.

ii. Specific groups (if any) in the libraries.

iii. External participants to the Library.

b. Define and establish editorial responsibilities.

c. Identify potential contributors within the Institution

3. To collaborate on strategic alliances with on and off campus groups to promote the library via pre-existing vehicles and to provide more responsive services to these campus groups and their constituents.

a. Identify key opportunities for expansion

b. Seek for input and help from the libraries' staff

c. Seek for opportunities to publish in the newsletter of student services offices and other academic departments

4. To advertise campus-wide and in the community. Efficient advertising will be essential to the library's success in attracting and retaining users. Potential users need to be made aware of the high quality services and products available to them-mostly without a monetary charge.

a. Utilize relevant payments of library campaigning for campus and community advertizing.

b. Link into promotions and events sponsored by Library associations when applicable.

c. Utilize media: Issue press releases and public announcements for promotional events and to add to the depiction of libraries and allied resources.

5. Allocate resources to the annual reception for faculty authors

a. Define authorship to include the creation of books and performing accomplishments.

b. In cooperation with the Rector's office and library administration, establish date and place and budget of reception.

c. Seek for participant lists and other campus publications noting faculty work.

V. Future Trends

Library services are bound to increase in the ICT segment viz., electronic databases and electronic journals. The World Wide Web will continue to act as a change agent in the usage of library print and subscription to its database resources. The trend towards electronic development and maintenance of library materials would make library resources accessible readily from home or other off campus systems. Promotional material and activities need to create a desire or need to visit the library in person for instruction and educational support services. One of the primary challenges is to market a library's resource to online users through the central command office of the host university. Another notable challenge is to market library services to patrons on and off campus.

VI. Conclusion

To be effective in achieving the library's aims, any university needs to understand the overall environment in which it functions. For an established University, this means knowing its customers and competitors; for modern libraries, it means mainly (but not limited) understanding its users and their potential demands. A thorough environmental scan will hand-hold a library in developing and marketing its services as deemed necessary in a measurable environment, and can be crucial to its health and efficiency. The references quoted in this paper give a bit of theory and practical advice regarding how to discover the demographics of a library's universe; the process of ascertaining the needs and preferences of its actual and potential users; and the methodology of analyzing all available / possible factors, external and internal, to cater to the library's strategic planning process. All types of institutions (academic and non-academic) providing information need to pay attention to the process of serving customers and manage with media and public relations. In order to provide good customer service, a library must primarily make sure that its customers know of the available services and rendering schedule (time-frames). Whether an information organization is an academic or professional library, any or all publishing houses or a patent office, needs to get word out about services available and take care of customers and prioritize their primary needs. By seeking to understand how diverse information is stored in today's ICT world, a library needs to structure its rendered services and manage public relations, sharing of best practices may be outlined, identified and advantages gained for a better tomorrow.

Interactivity is a mandated aspect of student-centered course design framework, especially in the internet era. Students more often than not, require learning institutions to generate graduates who are problem solvers and creative thinkers; personnel who are literate enough to function well in a knowledge-centric economy. To achieve this educational goal, ICT needs to reform traditional methods of instruction, shifting from a more passive method of teaching to more interactive methodologies.

However, the concern of how to evaluate the coveted goal of interactivity in a class is not often fully addressed, as asserted by this study. Based on four years of ICT enabled design and delivery, this paper has proposed multiple factors and criteria to that can be considered in determining how interactivity can be improved and evaluated. Fundamentally, teachers ought to consider their teaching environment as one of a conversation between an instructor and learner in what this research terms "Conversational Learning Society". Imperative components in this environment include Learners, instructor(s), course materials, and links to remote experts and resources enabled by ICT. All these components are fastened intact by instructional interactivity. Three types of instructional interactivity ought to be recognized. These are learner-learner, instructor-learner and learner-resource interactivity.

To evaluate interactivity one could mull-over both quantitative and qualitative criteria. On the qualitative aspects a teacher needs to pay attention to issues such as initiative, critical thinking and academic rigor coupled with the factor "how far the students seem to be showing signs of these". On the quantitative aspects it may be helpful to consider various log-on statistics on the course homepage and other related statistics about student appraisal of the coursework.

The criteria of how to successfully evaluate the magnitude to which a course design facilitates interaction is bound to remain for a long time - a contentious one. Regardless of the case with regard to the issue of evaluation, this study asserts and concludes that interactivity on the web enhances even tutorial sessions and traditional classroom. ICT enabled Interactive web-based design can allow teachers to achieve a better functional management of their courses, leading to a more effectual scholarship of learning and teaching.

REFERENCES

[1] Altman, M., Andreev, L., Diggory, M., Krot, M., King, G., Kiskis, D., Kolster, E., Sone, A.,&Verba, S. (2001). An introduction to theVirtual Data Center project and software. In, Proceedings of the first ACM+IEEE Joint Conference on Digital Libraries. New York: ACM.

[2] Andreasen, A. R., & Kotler, P. (2002). Strategic marketing for nonprofit organizations (6th ed.) Upper Saddle River, NJ: Prentice Hall Press.

[3] Barnard, R. (1995) Interactive Learning: A key to successful distance delivery, The American Journal of Multimedia, 12, 45 - 47.

[4] Berry, L. L. (1995). On great service: A framework for action. New York: Free Press.

[5] Blurton, C. (1999) New directions of ICT-use in education, UNESCO's World Communication and Information Report 1999. URL: http://www.unesco.org/education/educprog/lwf/d1/edict.pdf

[6] Boaden, S. (2005). Building public library community connections through cultural planning. Australasian Public Libraries and Information Services, 18 (1), 29-36.

[7] Bodomo, A. (2000) Interactivity in Web-based (Distance Education) Courses, CD-ROM, Proceedings of the Ghana Computer Literacy & Distance Education (GhaCLAD) Conference, 26th – 29th July, 2000, Accra, Ghana.

[8] Bodomo, A. (2001) Interactivity in Web-based courses, Paper for the WebCT Asia Pacific Conference, 9th-11th April 2001, Adelaide, Australia. Brogan, Pat. (1999) 'Using the web for interactive teaching and learning: An imperative for the new millennium', A white paper for the Macromedia's Interactive Learning Division

[9] Bruner, J. (1966) Toward a Theory of Instruction, Cambridge, MA: Harvard University Press.

[10] Bruner, J. (1983) Child's Talk: Learning to Use Language, New York: Norton.

[11] Bruner, J. (1986) Actual Minds, Possible Worlds, Cambridge, Harvard University Press.

[12] Bruner, J. (1990) Acts of Meaning, Cambridge, MA: Harvard University Press.

[13] Ceci, S., & Walker, E. (1983). Private archives and public needs. American Psychologist, 38, 414-423.

[14] Cowan, R. S. (1997). A social history of American technology. New York: Oxford University Press.

[15] Davis-Kahl, S. (2004). Creating a marketing plan for your academic and research library. Retrieved from Illinois Wesleyan University website

[16] Daniel, J. and C. Marquis. (1983) Independence and interaction: Getting the mix right, Teaching at a Distance, 15, 445-460.

[17] Daniel, R. (1997). A trivial convention for using HTTP in URN resolution [Online] (Internet Engineering Task Force, No. RFC 2169). Retrieved from the World Wide Web: http://www.ietf.org/.

[18] Davis, J. R.,&Lagoze, C. (2000). NCSTRL: Design and deployment of a globally distributed digital library. JASIS, 51, 273-280.

[19] Feigenbaum, S., & Levy, D. M. (1993). The market for (ir)reproducible econometrics. Social Epistemology, 7, 215-292.

[20] Freire, Paulo. (1970) The Adult Literacy Process as Cultural Action for Freedom, Harvard Educational Review, 40, 205-212

[21] Griffin, S. M. (1998). NSF/DARPA/NASA digital libraries initiative: A program manager's perspective. D-Lib Magazine [Online serial], 4(7). Retrieved May 1, 2001, from the World Wide Web: http://www.dlib.org/dlib/july98/07griffin.html

[22] Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics, 5, 299-314.

[23] Jackson, Robert H. (2000) Web-based Learning Resources Library, Available HTTP:http://www.outreach.utk.edu/weblearning /(email:rhjackson@utk.edu)

[24] Kearsley, Greg. (1994-2000) Explorations in Learning & Instruction: Theory in Practice Database, Available HTTP:http://www.gwu.edu/~tip/ (May 21, 2000).(email:gkearsley@sprynet.com) http://home.sprynet.com/~gkearsley

[25] Kotler, P., Bloom, P. N., & Hayes, T. (2002). Marketing professional services: Forward-thinking

[26] strategies for boosting your business, your image, and your profits (2nd ed.). Paramus, NJ: Prentice Hall Press.

[27] Laurie, B., Laurie, P., & Denn, R. (Ed.). (1998). Apache: The definitive guide. Sebastapol, CA: O'Reilly and Associates. Altman et al. / VIRTUAL DATA CENTER 469

[28] Lee, D. (2005). Can you hear me now? Using focus groups to enhance marketing research. Library Administration & Management, 19 (2), 100-101.

[29] Lessig, L. (1999). Code, and other laws of cyberspace. New York: Basic Books.

[30] Markwood, R. and S. Johnstone. (1994) New Pathways to a degree: Technology opens the college, Western Cooperative for Educational Telecommunications, Western Interstate Commission for Higher Education, Boulder, CO.

[31] McCullough, B. D.,&Vinod, H. D. (1999). The numerical reliability of econometric software. Journal of Economic Literature, 37, 633-665.

[32] McKenna, R. (2002). Total access: Giving customers what they want in an anytime, anywhere world. Boston: Harvard Business School Press.

[33] Moore, M. (1991) Editorial: distance education theory, The American Journal of Distance Education, 5(3), 1- 6.

[34] Moore, M. (1992) Three types of interaction, The American Journal of Distance Ed, 3 (2), 1-6.

[35] Moore, M. (1993) Theory of transactional distance, in Desmond Keegan (Eds.) Theoretical Principles of Distance Education, London & New York: Routledge.

[36] Parker, Angie. (1999) Interaction in Distance Education: The critical conversation, Education Technology Review, 13.

[37] Pask, G. (1975) Conversation, Cognition, and Learning, New York: Elsevier.

[38] Piaget, J. (1973) To Understand is to Invent, New York: Grossman.

[39] Plosker, G. (2005). The information strategist: Revisiting library funding: What really works? Online, 29 (2), 48-53.

[40] Strauss, M. J. (1994) A constructivist dialogue, Journal of Humanistic Education and Development, 32 (4), 183 – 187.

[41] Schmitt, B. H. (2003). Customer experience management: A revolutionary approach to connecting with your customers. Hoboken, NJ: John Wiley & Sons.

[42] Thenell, J. (2004). The library's crisis communications planner: A PR guide for handling every emergency. Chicago: American Library Association.

[43] Vygotsky, L. S. (1962) Thought and Language, Cambridge, MA: MIT Press.

[44] Vygotsky, L. S. (1978) Mind in Society, Cambridge, MA: Harvard University Press.

[45] Wagner, J. (1994) Learning from a distance, The International Journal of Multimedia, 19 (2): 12 – 20.

[46] Walter, S. (in press). Moving beyond collections: Academic library outreach to multicultural student centers. Reference Services, Review, 33 (4).

[47] Weiner, S. A. (in press). Library quality and impact: Is there a relationship between new measures and traditional measures? The Journal of Academic Librarianship.DOI: 10.1016/j.acalib.2005.05.004

[48] Weldon, S. (2005). Collaboration and marketing ensure public and medical library viability. Library Trends, 53 (3), 411-421.

[49] Williams, R., Bunn, J., Moore, R.,&Pool, J.C.T. (1998).Workshop on interfaces to scientific data archives [Online] Pasadena, CA: California Institute of Technology, Center for Advanced Computing Research. Retrieved May 10, 1999, from the World Wide Web: http://www.cacr.caltech.edu/isda.

AUTHORS PROFILE

The author is 34 years old and has 6 years of MBA level teaching experience and has presented 5 Research Papers in International Conferences, and have 5 accepted papers for publication in International Journals. The author also has six years of working experience in Banking and Finance, predominantly in Multi-Site, transnational operations encompassing Operations Management and Investment banking. The author is a keen planner, strategist & implementer with skills in handling complete responsibility for management of departmental administration, organization of study tours, Sports, NSS activities, organizing Medical Camps, training and spearheading campus connect meetings.

OFDM System Analysis for reduction of Inter symbol Interference Using the AWGN Channel Platform

D.M. Bappy, Ajoy Kumar Dey, Susmita Saha
Department of Signal Processing
Blekinge Tekniska Hogskola (BTH)
Karlskrona, Sweden
E-mail: dey.ajoykumar@yahoo.com

Avijit Saha, Shibani Ghosh
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
E-mail: avijit003@gmail.com

***Abstract*— Orthogonal Frequency Division Multiplexing (OFDM) transmissions are emerging as important modulation technique because of its capacity of ensuring high level of robustness against any interferences. This project is mainly concerned with how well OFDM system performs in reducing the Inter Symbol Interference (ISI) when the transmission is made over an Additive White Gaussian Noise (AWGN) channel. When OFDM is considered as a low symbol rate and a long symbol duration modulation scheme, it is sensible to insert a guard interval between the OFDM symbols for the purpose of eliminating the effect of ISI with the increase of Signal to Noise Ratio (SNR).**

Keywords- OFDM; Inter symbol Interference; AWGN; matlab; algorithm.

I. INTRODUCTION *(HEADING 1)*

This paper is intended for those engineers whose research interests are related to digital communication, particularly the terrestrial digital video broadcasting (DVB-T) and the digital audio broadcasting (DAB) as Orthogonal Frequency Division Multiplexing(OFDM) system has been effectively used in both digital television and radio broadcasting system. The OFDM usually operates with high data rate and hence it is also widely used in asymmetric digital subscriber lines (ADSL) in order to achieve a high speed data connection. Considering the cost as a factor, this paper excludes the need of an equalizer and shows that it is still possible to reduce the ISI in an OFDM system without the equalizer[6].

This project investigates the OFDM system performance in eliminating the Inter symbol Interference (ISI) effect which, in our experiment, is caused by the noise generated from the AWGN channel. In the case of using the single carrier modulation, frequency selective fading and inter symbol interference occur, which leads to high probability of errors, consequently, affecting on the system performance. So, it is crucially important to study the elimination process of ISI effect. Additionally, in our project, we have considered OFDM system because OFDM is a form of digital multi-carrier modulation method that is widely used in wireless communication in order to solve the problem. In order to reducing the ISI effects, a 16-QAM modulation technique and an AWGN channel are used in our experiment.

II. LITERATURE REVIEW

Assuming the ideal channel or noiseless environment, we do so in order to focus attention on the effects of imperfections in the frequency response of the channel, on data transmission through the channel[4].

The receiving filter output, we may be written as

$$y(t) = \mu \sum_{k=-x}^{x} A_k \, p(t - kT_b) \qquad (1)$$

where μ, is a scaling factor. The pulse p(t) has a shape different from that of g(t), but it is normalized such that,

$$p(0) = 1.$$

The pulse μA_k p(t) is the response of the cascade connection of the transmitting filter, the channel, and receiving filter, which is produced by the pulse $A_k g(t)$ applied to the input of this cascade connection. Therefore, we may relate p(t) to g(t) in the frequency domain as follows (after cancelling the common factor A_k)

$$\mu P(f) = G(f)H_T(f)H_C(f)H_R(f) \qquad (2)$$

where P(f) and G(f) are the Fourier transform of p(t) and g(t), respectively. The receiving filter output y(t) is sampled at time $t_i = iT_b$ (with *i* taking on integer values), yielding

$$y(t_i) = \mu \sum_{k=-\infty}^{\infty} A_k \, p[(i-k)T_b]$$

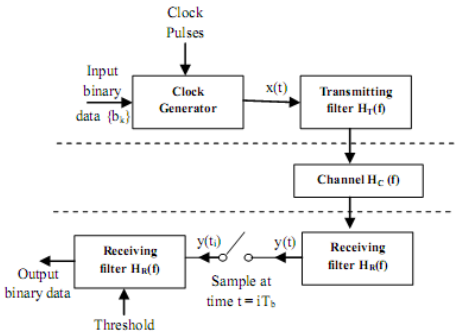
$$\mu A_i + \mu \sum_{\substack{k=-\infty \\ k \neq i}}^{\infty} A_k p[(i-k)T_b] \quad i = 0, \pm 1, \pm 2 \qquad (3)$$

=

In Equation 3, the first term μA_i represents the contribution of the *i*th transmitted bit. The second term represents the residual effect of all other transmitted bits on the decoding of the *i*th received bit; this residual effect is called inter symbol

interference (ISI). In the absence of ISI, we observe from the equation 3 that

Figure 1. Baseband Transmission System



$$y(t_i) = \mu A_i \qquad (4)$$

Which shows that, under these conditions, the *i*th transmitted bit can be decoded correctly. The unavoidable presence of ISI in the system, however, introduces errors in the decision device at the receiver output. Therefore, in the design of the transmitting and receiving filters, the objective is to minimize the effects of ISI, and thereby deliver the digital data to its destinations with the smallest error rate.

Two previous researches demonstrate that ISI can be handled effectively in OFDM system by using the 4-QAM technique [1] [2]. ISI effect can be reduced by inserting a guard interval at low symbol rate instead of using pulse shaping filter at high symbol rate and this idea is drawn from [3].

III. Research Rationale

In an OFDM system with a higher symbol rate, reflected signals can interfere with consequent symbols. It creates complexity as the received signal can be distorted by this interference. So our research question arises, ‘‘how can we reduce ISI in an OFDM system?’’

In the OFDM system we can mitigate this problem by modulating multiple sub-carriers per channel at lower symbol rate (long symbol period). By using a smaller symbol rate, the signal reflections are found occupying only a small portion of the total symbol period. Thus, it is possible to just insert a guard interval between consecutive OFDM symbols in order to remove interference from reflections. And then it becomes feasible for ISI to be removed from signals by increasing the value of SNR [5].

The main contributions of this project are:

1. Designing a new method that reduces the ISI in OFDM system very effectively. The method uses a guard interval which is equal to one-fourth of the symbol period, and also which includes a 16-QAM slicer, the slicer which is mainly used to compare the output constellation.

2. Implementing the whole model on MATLAB Workspace. In order to validate our experiment, the result of the simulation is shown in a plot of the SNR versus SER (Symbol Error Rate).

IV. Analysis of Findings

A. OFDM signal generation and reception over AWGN channel

The system model of the project is shown in Fig 1 where the upper part is the transmitter block and the lower part is the receiver block.

After working with the transmitter block, we have the transmitted signal *s(t)* with the carrier frequency f_c,

$$s(t) = u_o ft_I(t)\cos(2\pi f_c t) + u_o ft_Q(t)\sin(2\pi f_c t) \qquad (5)$$

From (1), the quadrature signal accords with the imaginary part of the complex modulation symbols and in our project, these are the 16-QAM symbols. The signal is then transmitted through an AWGN channel and the purpose of using this channel is to add a noise *n(t)* to the transmitted signal *s(t)*. In the receiver part, the corrupted OFDM signal is first filtered by using a low pass filter in order to get the original baseband signal and then sampled. Finally, the output from FFT modulation gives the received constellation. The received constellation passes through a 16-QAM slicer and inserts the received symbols into the sixteen possible constellation points.

B. ISI analysis and Validation

In our experiment, we have considered only one path (AWGN channel) between OFDM transmitter and receiver which corrupts the transmitted signal. After getting the output from FFT block, it is possible to make a comparison between the original and received 16-QAM constellation. By observing the received constellation, a large amount of symbol errors arises which leads to high noise power, consequently, generates the ISI in the system. By increasing the value of SNR up to a certain level, we can reduce the Symbol Error Rate (SER) as well as the ISI.

Table 1 presents the symbol errors for different values of SNR.

Fig 3 shows the ideal 16QAM constellation that we get in the transmitter part. Fig 4 shows that the noise power and the ISI are found higher when the SNR value is fixed at 0 dB. With the increase of SNR, ISI is found to be reduced effectively and this is shown in Fig 5. Finally, Fig 6 illustrates the system performance as good as we expected.

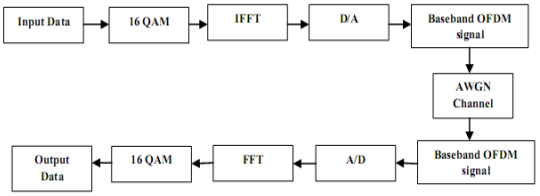


Figure 2. OFDM system model for 16-QAM

TABLE 1. NO. OF SYMBOL ERRORS FOR DIFFERENT SNR

SNR[dB]	Symbol Error	SNR[dB]	Symbol Error
0	612	12	44
4	362	14	12
8	203	16	5

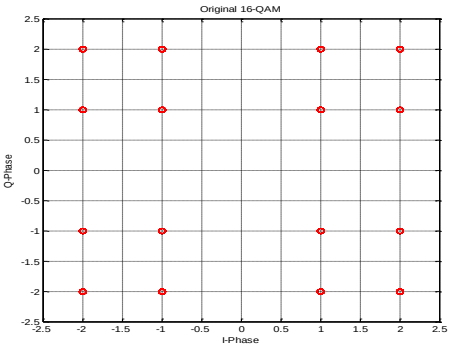


Figure 3. Original 16-QAM constellation

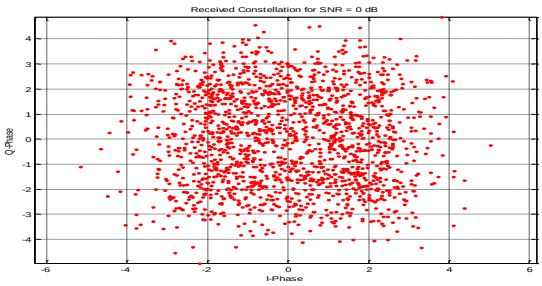


Figure 4. Received 16-QAM constellation with SNR =0 dB

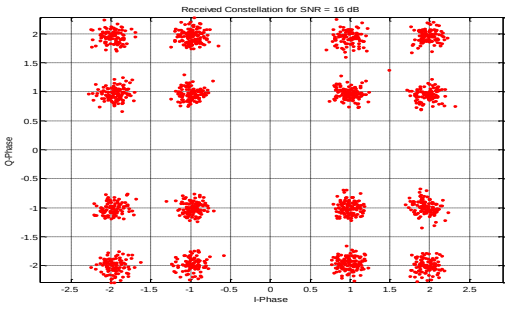


Figure 5. Received 16-QAM constellation with SNR =16 dB

V. CONCLUSION

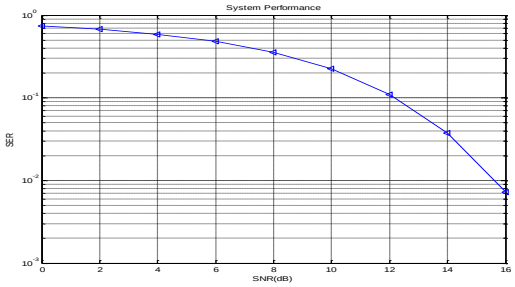


Figure6. Simulated SER Performance For 16-QAM

In this paper, we have demonstrated the application of a 16-QAM modulation technique in the OFDM system with a view to reducing the ISI. The experiment has been successful through adding the guard intervals between the consecutive symbols. Although we could not bring the SER reduced to zero our designed system still works properly. The system performance curve demonstrates that it can still be used for DVB-T.

For future work, improvement can be made by coding the original information; over and above, the ISI effect could be further removed by using an equalizer at the receiver end.

REFERENCES

[1] Mary Ann Ingram and Guillermo Acosta, *Smart Antenna Research Laboratory using Matlab*, 2000. [E-book] Available: toodoc e-book [Accessed Sep. 23, 2009].

[2] Long Bora and Heau-Jo Kang, "A Study on the Performance of OFDM Transmission Scheme for QAM Data Symbol," *2008 Second International Conference on Future Generation Communication and Networking (FGCN)*, vol.2, pp.267 – 270, Dec. 2008.

[3] National Instruments, "OFDM and Multi-Channel Communication systems,"2007.[Online]Available:http://zone.ni.com/devzone/cda/tut/p/id/37 40 [Accessed Oct. 5, 2009]

[4] Simon Haykin "An introduction to Analog & Digital communications", 1994, pp. 228-231

[5] K. Takashahi, T. Saba, "A novel symbol synchronization algorithm with reduced influence of ISI for OFDM systems",2001 Global Telecommunication Conference, Globcom"01, IEEE, vol.1, pp.524-528, Nov.2001 USA.

[6] Dukhyun Kim, Stuber, G.L., "Residual ISI cancellation for OFDM with applications to HDTV broadcasting", IEEE Journals on Communication, vol.16, pp.1590-1599, Oct. 1998