IJACSA

WHERE WISDOM SHARES

ISSN : 2156-5570

Volume 1 Issue 3 2010

INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS

IJACSA Publications 2010
New York,
United States of America

http://ijacsa.thesai.org

Thank You For Sharing Wisdom!

IJACSA Editorial

From the Desk of Managing Editor…

With this issue we aim to reach out to a larger number with an expectation that more and more authors join in our mission to provide an outlet for quality research to be publicised and published to a global audience.

IJACSA wishes to become a sole conduit for sharing the research work done by the authors in the field of Advanced Computer Science and Applications.

We strive to produce the journal of the highest quality, and at the same time, to keep our journal accessible to researchers and students at institutions of all types and sizes, everywhere in the world, through on-going efforts to reduce production costs and through policies that bring the wisdom world closer and integrated.

As the editor my responsibilities will continue towards the advancement of the Journal and its promotion and distribution in a manner that it becomes an integrated component of Scientific Community

IJACSA expresses its appreciation to all Reviewers, whose efforts in peer review not only keep the standards of the journal high, but also help authors to improve the readability of their articles.
The Associate Editors deserve special regards for being exceptionally helpful in assessing manuscripts for publication in the IJACSA.

Since the integration of IJACSA into The Science and Information (SAI) Organization we made significant strides in our behind-the-scenes editorial work as well, decreasing the IJACSA Acceptance rate to almost 30 per cent.
I know that amongst the readers there is a good deal of intellectual thinking in the area that our journal focuses on, I, therefore respectfully ask that you consider preparing a submission for our Journal.

The Authors are to be congratulated and thanked for their outstanding contribution to the scientific community. We hope that the relationships we have cultivated will continue and expand.

Thank You for Sharing Wisdom!

IJACSA Managing Editor
September 2010
editorijacsa@thesai.org

Editorial Board

IJACSA Reviewer Board

CONTENTS

A Comparative Study of Gaussian Mixture Model and Radial Basis Function for Voice Recognition

Fatai Adesina Anifowose
Center for Petroleum and Minerals, The Research Institute
King Fahd University of Petroleum and Minerals
Dhahran 31261, Saudi Arabia
anifowo@kfupm.edu.sa

***Abstract*—A comparative study of the application of Gaussian Mixture Model (GMM) and Radial Basis Function (RBF) in biometric recognition of voice has been carried out and presented. The application of machine learning techniques to biometric authentication and recognition problems has gained a widespread acceptance. In this research, a GMM model was trained, using Expectation Maximization (EM) algorithm, on a dataset containing 10 classes of vowels and the model was used to predict the appropriate classes using a validation dataset. For experimental validity, the model was compared to the performance of two different versions of RBF model using the same learning and validation datasets. The results showed very close recognition accuracy between the GMM and the standard RBF model, but with GMM performing better than the standard RBF by less than 1% and the two models outperformed similar models reported in literature. The DTREG version of RBF outperformed the other two models by producing 94.8% recognition accuracy. In terms of recognition time, the standard RBF was found to be the fastest among the three models.**

Keywords- Gaussian Mixture Model, Radial Basis Function, Artificial Intelligence, Computational Intelligence, Biometrics, Optimal Parameters, Voice Pattern Recognition, DTREG

I. INTRODUCTION

Biometrics is a measurable, physical characteristic or personal behavioral trait used to recognize the identity, or verify the claimed identity, of a candidate. Biometric recognition is a personal recognition system based on “who you are or what you do” as opposed to “what you know” (password) or “what you have” (ID card) [17]. The goal of voice recognition in biometrics is to verify an individual's identity based on his or her voice. Because voice is one of the most natural forms of communication, identifying people by voice has drawn the attention of lawyers, judges, investigators, law enforcement agencies and other practitioners of forensics.

Computer forensics is the application of science and engineering to the legal problem of digital evidence. It is a synthesis of science and law [8]. A high level of accuracy is required in critical systems such as online financial transactions, critical medical records, preventing benefit fraud, resetting passwords, and voice indexing.

In view of the importance of accurate classification of vowels in a voice recognition system, the need for a well-trained computational intelligence model with an acceptable percentage of classification accuracy (hence a low percentage of misclassification error) is highly desired. Gaussian Mixture Models (GMMs) and Radial Basis Function (RBF) networks have been identified in both practice and literature as two of the promising neural models for pattern classification.

The rest of this paper is organized as follows. Section II reviews the literature on voice recognition; overview and application of GMM and RBF in biometric voice recognition; and an overview of the RBF component of DTREG software. A description of the data and tools used in the design and implementation of this work are discussed in Section III. Section IV describes the experimental approach followed in this work and the criteria for quality measurement used to evaluate its validity. The results of the experiment are discussed in section V while conclusions are drawn in section VI.

II. LITERATURE SURVEY

A. Voice Recognition

A good deal of effort has been made in the recent past by researchers in their attempt to come up with computational intelligence models with an acceptable level of classification accuracy.

A novel suspect-adaptive technique for robust forensic speaker recognition using Maximum A-Posteori (MAP) estimation was presented by [1]. The technique addressed Likelihood Ratio computation in limited suspect speech data conditions obtaining good calibration performance and robustness by allowing the system to weigh the relevance of the suspect specificities depending on the amount of suspect data available via MAP estimation. The results showed that the proposed technique outperformed other previously proposed non-adaptive approaches.

[2] presented three mainstream approaches including Parallel Phone Recognition Language Modeling (PPRLM), Support Vector Machine (SVM) and the general Gaussian Mixture Models (GMMs). The experimental results showed that the SVM framework achieved an equal error rate (EER) of

4.0%, outperforming the state-of-art systems by more than 30% relative error reduction. Also, the performances of their proposed PPRLM and GMMs algorithms achieved an EER of 5.1% and 5.0% respectively.

Support Vector Machines (SVMs) were presented by [3] by introducing a sequence kernel used in language identification. Then a Gaussian Mixture Model was developed to do the sequence mapping task of a variable length sequence of vectors to a fixed dimensional space. Their results demonstrated that the new system yielded a performance superior to those of a GMM classifier and a Generalized Linear Discriminant Sequence (GLDS) Kernel.

Using a vowel detection algorithm, [4] segmented rhythmic units related to syllables by extracting parameters such as consonantal and vowel duration, and cluster complexity and modeled with a Gaussian Mixture. Results reached up to 86 ± 6% of correct discrimination between stress-timed, mora-timed and syllable-timed classes of languages. These were then compared with that of a standard acoustic Gaussian mixture modeling approach that yielded 88 ± 5% of correct identification.

[9] presented an additive and cumulative improvements over several innovative techniques that can be applied in a Parallel Phone Recognition followed by Language Modeling (PPRLM) system for language identification (LID), obtaining a 61.8% relative error reduction from the base system. They started from the application of a variable threshold in score computation with a 35% error reduction, then a random selection of sentences for the different sets and the use of silence models, then, compared the bias removal technique with up to 19% error reduction and a Gaussian classifier of up to 37% error reduction, then, included the acoustic score in the Gaussian classifier with 2% error reduction, increased the number of Gaussians to have a multiple-Gaussian classifier with 14% error reduction and finally, included additional acoustic HMMs of the same language with success gaining 18% relative improvement.

B. *Gaussian Mixture Model (GMM)*

From a clustering perspective, most biometric data cannot be adequately modeled by a single-cluster Gaussian model. However, they can often be accurately modeled via a Gaussian Mixture Model (GMM) i.e., data distribution can be expressed as a mixture of multiple normal distributions [7].

Basically, the Gaussian Mixture Model with k components is written as:

$$p(y|\mu_1,\ldots,\mu_k,s_1,\ldots,s_k,\pi_1,\ldots,\pi_k)=\sum_{j=1}^{k}\pi_j\mathcal{N}\left(\mu_j,s_j^{-1}\right) \quad (1)$$

where μ_j are the means, s_j the precisions (inverse variances), π_j the mixing proportions (which must be positive and sum to one) and N is a (normalized) Gaussian with specified mean and variance. More details on the component parameters and their mathematical derivations can be found in [10-13, 25, 26].

[5] presented a generalized technique by using GMM and obtained an error of 17%. In another related work, [10] described two GMM-based approaches to language identification that use Shifted Delta Costar (SDC) feature vectors to achieve LID performance comparable to that of the best phone-based systems. The approaches included both acoustic scoring and a GMM tokenization system that is based on a variation of phonetic recognition and language modeling. The results showed significant improvement over the previously reported results.

A description of the major elements of MIT Lincoln Laboratory's Gaussian Mixture Model (GMM)-based speaker verification system built around the likelihood ratio test for verification, using simple but effective GMMs for likelihood functions, a Universal Background Model (UBM) for alternative speaker representation, and a form of Bayesian adaptation to derive speaker models from the UBM were presented by [6]. The results showed that the GMM-UBM system has proven to be very effective for speaker recognition tasks.

[12] evaluated the related problem of dialect identification using the GMMs with SDC features. Results showed that the use of the GMM techniques yields an average of 30% equal error rate for the dialects in one language used and about 13% equal error rate for the other one.

Other related works on GMM include [11, 13].

C. *Radial Basis Function (RBF)*

A RBF Network, which is multilayer and feedforward, is often used for strict interpolation in multi-dimensional space. The term 'feedforward' means that the neurons are organized in the form of layers in a layered neural network. The basic architecture of a three-layered neural network is shown in Fig. 1.

A RBFN has three layers including input layer, hidden layer and output layer. The input layer is composed of input data. The hidden layer transforms the data from the input space to the hidden space using a non-linear function. The output layer, which is linear, yields the response of the network.

The argument of the activation function of each hidden unit in an RBFN computes the Euclidean distance between the input vector and the center of that unit. In the structure of RBFN, the input data X is an I-dimensional vector, which is transmitted to each hidden unit. The activation function of hidden units is symmetric in the input space, and the output of each hidden unit depends only on the radial distance between the input vector X and the center for the hidden unit. The output of each hidden unit, h_j, j = 1, 2, . . ., k is given by:

$$h_j(x)=\phi(\|x-c_j\|) \quad (2)$$

Where || || is the Euclidean Norm, c_j is the center of the neuron in the hidden layer and Φ() is the activation function.

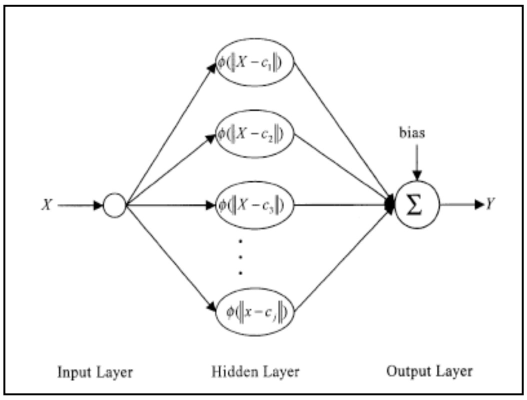


Figure 1. The architecture of an RBF network.

The activation function is a non-linear function and is of many types such Gaussian, multi-quadratic, thinspline and exponential functions. If the form of the basis function is selected in advance, then the trained RBFN will be closely related to the clustering quality of the training data towards the centers.

The Gaussian activation function can be written as:

$$\phi_j(x) = \exp\left[-\frac{\|x - c_j\|^2}{2\rho^2} \right] \quad (3)$$

where x is the training data and ρ is the width of the Gaussian function. A center and a width are associated with each hidden unit in the network. The weights connecting the hidden and output units are estimated using least mean square method. Finally, the response of each hidden unit is scaled by its connecting weights to the output units and then summed to produce the overall network output. Therefore, the k^{th} output of the network $\hat{y}_k$ is:

$$\hat{y}_k = w_0 + \sum_{j=1}^{M} w_{jk}\phi_j(x) \quad (4)$$

where $\Phi_j(x)$ is the response of the j^{th} hidden unit, w_{jk} is the connecting weight between the j^{th} hidden unit and the k^{th} output unit, and w_0 is the bias term [18].

RBF model, with its mathematical properties of interpolation and design matrices, is one of the promising neural models for pattern classification and has also gained popularity in voice recognition [14].

[15] presented a comparative study of the application of a minimal RBF Neural Network, the normal RBF and an elliptical RBF for speaker verification. The experimental results showed that the Minimal RBF outperforms the other techniques. Another work for explicitly modeling voice quality variance in the acoustic models using RBF and Hidden Markov Models, in order to improve word recognition accuracy, was demonstrated by [16]. They also presented SVM and concluded that voice quality can be classified using input features in speech recognition.

Other related works have been found in the fields of medicine [14], hydrology [18], computer security [19], petroleum engineering [20] and computer networking [21].

D. DTREG Radial Basis Function (DTREG-RBF)

DTREG software builds classification and regression Decision Trees, Neural and Radial Basis Function Networks, Support Vector Machine, Gene Expression programs, Discriminant Analysis and Logistic Regression models that describe data relationships and can be used to predict values for future observations. It also has full support for time series analysis. It analyzes data and generates a model showing how best to predict the values of the target variable based on values of the predictor variables. DTREG can create classical, single-tree models and also TreeBoost and Decision Tree Forest models consisting of ensembles of many trees. It includes a full Data Transformation Language (DTL) for transforming variables, creating new variables and selecting which rows to analyze [27].

One of the classification/regression tools available in DTREG is Radial Basis Function Networks. Like the standard RBF, DTREG-RBF has an input layer, a hidden layer and an output layer. The neurons in the hidden layer contain Gaussian transfer functions whose outputs are inversely proportional to the distance from the center of the neuron. Although the implementation is very different, RBF neural networks are conceptually similar to K-Nearest Neighbor (K-NN) models. The basic idea is that a predicted target value of an item is likely to be about the same as other items that have close values of the predictor variables.

DTREG uses a training algorithm developed by [28]. This algorithm uses an evolutionary approach to determine the optimal center points and spreads for each neuron. It also determines when to stop adding neurons to the network by monitoring the estimated leave-one-out (LOO) error and terminating when the LOO error begins to increase due to overfitting. The computation of the optimal weights between the neurons in the hidden layer and the summation layer is done using ridge regression. An iterative procedure developed by an author in 1966 is used to compute the optimal regularization lambda parameter that minimizes the Generalized Cross-Validation (GCV) error. A more detailed description of the DTREG can be found in [27].

III. DATA AND TOOLS

The training and testing data were obtained from an experimental 2-dimensional dataset available in [22]. The training data consists of 338 observations while the testing data consists of 333 observations. There are 2 input variables and each observation belongs to one of 10 classes of vowels to be classified using the trained models.

The GMM and RBF classifiers were implemented in MATLAB with the support of NETLAB toolbox obtained as freeware from [23] while the DTREG-RBF was implemented using the DTREG software version 8.2. The descriptive statistics of the training and test data are shown in table I and II while the scatter plots of the training and test data are shown in Fig. 2 respectively.

IV. Experimental Approach and Criteria for Performance Evaluation

The methodology in this work is based on the standard Pattern Recognition approach to classification problem using GMM and RBF. For training the models, Expectation Maximization (EM) algorithm was used for efficient optimization of the GMM parameters. The RBF used forward and backward propagation to optimize the parameters of the neurons using the popular Gaussian function as the transform function in the hidden layer as is common in literature. The parameters of the models were also tuned and varied and those with maximum classification accuracy were selected. The DTREG-RBF was run on the same dataset with the default parameter settings.

For the GMM, several runs were carried out using the "diag" and "full" covariance types and with number of centers ranging from 1 and 10 while for the RBF, several runs were carried out with different numbers of hidden neurons ranging from 1 and 36.

TABLE I. Descriptive Statistics of Training data

	X1	X2
Average	567.82	1533.18
Mode	344.00	2684.00
Median	549.00	1319.50
Std Dev	209.83	673.94
Max	1138.00	3597.00
Min	210.00	557.00

TABLE II. Descriptive Statistics of Testning data

	X1	X2
Average	565.47	1540.38
Mode	542.00	2274.00
Median	542.00	1334.00
Std Dev	216.40	679.79
Max	1300.00	3369.00
Min	198.00	550.00

The DTREG-RBF is not flexible; only one variable can be set as the target at a time. It is most ideal for one-target classification problems. For this work, 10 different models were trained with each output column as the target. This was very cumbersome.

The most commonly used accuracy measure in classification tasks is the classification/recognition rate. This is calculated by:

$$\frac{p}{q} x\, 100$$

where *p* is the number of correctly classified points and *q* is the total number of data points.

For the purpose of evaluation in terms of speed of execution, Execution Time for training and testing was also used in this study.

V. Discussion of Results

For the GMM, generally, it was observed that the execution time increased as the number of centers was increased from 2, but with a little dip at 1. Similarly, the training and testing recognition rates increased as the number of centers was increased from 1 to 2 but decreased progressively when it was increased from 3. Fig. 3 and 4 show the plots of the different runs of the "diag" and "full" covariance types and how execution time and recognition rates vary with the number of centers. The class boundaries generated by the GMM Model for training and testing are shown in Fig. 5.

The results for GMM above showed that the average optimal performance was obtained with the combination of "full" covariance type and number of centers chosen to be 2.

For the RBF, generally, the training time increased as the number of hidden neurons increased while the testing time remained relatively constant except for little fluctuations. Also, the training and testing times increased gradually as the number of hidden neurons increased until up to 15 when they began to fall gradually at some points and remained relatively constant except for little fluctuations at some other points. Fig. 6 shows the decision boundaries of the RBF-based classifier using the same training and testing data applied on the GMMs while Fig. 9 shows the contour plot of the RBF model with the training data and the 15 centers.

The results for RBF above showed that the average optimal performance was obtained when the number of hidden neurons is set to 15.

As mentioned earlier in section IV, one disadvantage of the DTREG-RBF is that it accepts only one variable as the target. This constitutes a major restriction and poses a lot of difficulties. For each of the 10 vowel classes, one model was built by training it with the same dataset but with its respective class for classification. There is no automated way of doing this. For the purpose of effective comparison, the average of the number of neurons, training times and training and testing recognition rates were taken. Fig. 7 and 8 show the relationship between the number of hidden neurons and the execution time and classification accuracy respectively. They both indicate that the optimal performance in terms of execution time and classification accuracy is obtained approximately at the point where the number of hidden neurons is set to 15.

Comparatively, in terms of execution time, RBF clearly outperforms GMM and DTREG-RBF, but in terms of recognition rate, it was not clearly visible to see which is better between GMM and RBF since GMM (79.6%) is better in training than RBF (78.1%) while RBF (80.8%) is better in recognition than GMM (79.9%). To ensure fair judgment, the average of the training and testing recognition rates of the two models shows that GMM (79.7%) performs better than RBF (79.4%) by a margin of 0.3%. It is very clear that in terms of recognition accuracies, the DTREG-RBF model performed best with an average recognition rate of 94.79%. This is clearly shown in Fig. 10.

VI. CONCLUSION

A comparative study of the application of Gaussian Mixture Model (GMM) and Radial Basis Function (RBF) Neural Networks with parameters optimized with EM algorithm and forward and backward propagation for biometric recognition of vowels have been implemented. At the end of the study, the two models produced 80% and 81% maximum recognition rates respectively. This is better than the 80% recognition rate of the GMM proposed by Jean-Luc et al. in [4] and very close to their acoustic GMM version with 83% recognition rate as well as the GMM proposed by [5]. The DTREG version of RBF produced a landmark 94.8% recognition rate outperforming the other two techniques and similar techniques earlier reported in literature.

This study has been carried out using a vowel dataset. The DTREG-RBF models were built with the default parameter settings left unchanged. This was done in order to establish a premise for valid comparison with other studies using the same tool. However, as at the time of this study, the author is not aware of any similar study implemented with the DTREG software, hence there is no ground for comparison with previous studies.

Further experimental studies to evaluate the classification and regression capability of DTREG will be carried out to use each of its component tools such as Support Vector Machines, Probabilistic and General Regression Neural Networks, Cascaded Correlation, Multilayer Perceptron, Decision Tree Forest, and Logistic Regression for various classification and prediction problems in comparison with their standard (usually MATLAB-implemented) versions.

Furthermore, in order to increase the confidence in this work and establish a better premise for valid comparison and generalization, a larger and more diverse dataset will be used. In order to overcome the limitation of the dataset used where a fixed data was preset for training and testing, we plan for a future study where stratified sampling approach will be used to divide the datasets into training and testing sets as this will give each row in the dataset an equal chance of being chosen for either training or testing each time the implementation is executed.

With our previous work on the hybridization of machine learning techniques [29], a study has commenced for the combination of GMM and RBF as a single hybrid model to achieve better learning and recognition rates. It has been reported [30-33] and confirmed [29] that hybrid techniques perform better than their individual components used separately.

ACKNOWLEDGMENT

The author is grateful to the Department of Information and Computer Science and the College of Computer Sciences & Engineering of King Fahd University of Petroleum and Minerals for providing the computing environment and the licensed DTREG software for the purpose of this research. The supervision of Dr. Lahouari Ghouti and the technical evaluation Dr. Kanaan Faisal are also appreciated.

REFERENCES

[1] D. Ramos-Castro, J. Gonzalez-Rodriguez, A. Montero-Asenjo, and J. Ortega-Garcia, "*Suspect-adapted map estimation of within-source distributions in generative likelihood ratio estimation*", Speaker and Language Recognition Workshop, 2006. IEEE Odyssey 2006: The , vol., no., pp.1-5, June 2006.

[2] H. Suo, M. Li, P. Lu, and Y. Yan, "*Automatic language identification with discriminative language characterization based on svm*", IEICE-Transactions on Info and Systems, Volume E91-D, Number 3 , Pp. 567-575, 2008.

[3] T. Peng, W., and B. Li, "*SVM-UBM based automatic language identification using a vowel-guided segmentation*", Third International Conference on Natural Computation (ICNC 2007), ICNC, pp. 310-314, 2007.

[4] J. Rouas, J. Farinas, F. Pellegrino, and R. Andre-Obrecht, "*Rhythmic unit extraction and modeling for automatic language identification",* Speech Communication, Volume 47, Issue 4, December 2005, Pages 436-456.

[5] P.A. Torres-Carrasquillo, D.A. Reynolds, and J.R. Deller, "*Language identification using Gaussian mixture model tokenization*", IEEE International Conference on Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP '02), vol.1, no., pp. I-757-I-760 vol.1, 2002.

[6] A.D. Reynolds, T.F. Quatieri, and R.B. Dunn, "*Speaker verification using adapted gaussian mixture models*", Digital Signal Processing, Vol. 10, 19–41 (2000).

[7] S.Y. Kung, M.W. Mak, and S.H. Lin, "*Biometric authentication: a machine learning approach*", Prentice Hall, September 14, 2004, Pp. 496.

[8] T. Sammes and B. Jenkinson, "*Forensic computing: a practitioner's guide*", Second Edition, Springer-Verlag, 2007, Pp. 10.

[9] R. Córdoba, L.F. D'Haro, R. San-Segundo, J. Macías-Guarasa, F. Fernández, and J.C. Plaza, "*A multiple-Gaussian classifier for language identification using acoustic information and PPRLM scores*", IV Jornadas en Tecnologia del Habla, 2006, Pp. 45-48.

[10] P.A. Torres-Carrasquillo, E. Singer, M.A. Kohler, R.J. Greene, D.A. Reynolds, and J.R. Deller, "*Approaches to language identification using gaussian mixture models and shifted delta cepstral features*", Proceedings of International Conference on Spoken Language Processing, 2002.

[11] T. Chen, C. Huang, E. Chang, and J. Wang, "*Automatic accent identification using Gaussian mixture models*", IEEE Workshop on Automatic Speech Recognition and Understanding, 2001 (ASRU '01), Pp. 343-346, 9-13 Dec. 2001.

[12] P.A. Torres-Carrasquillo, T.P. Gleason, and D.A. Reynolds, "*Dialect identification using gaussian mixture models*", In Proc. Odyssey: The Speaker and Language Recognition Workshop in Toledo, Spain, ISCA, pp. 297-300, 31 May - 3 June 2004.

[13] T. Wuei-He and C. Wen-Whei, "*Discriminative training of gaussian mixture bigram models with application to chinese dialect identification*", Speech Communication, Volume 36, Issue 3, March 2002, Pp. 317 – 326.

[14] S. Miyoung and P. Cheehang, "*A radial basis function approach to pattern recognition and its applications*", ETRI Journal, Volume 22, Number 2, June 2000.

[15] L. Guojie, "*Radial basis function neural network for speaker verification*", A Master of Engineering thesis submitted to the Nanyang Technological University, 2004.

[16] T. Yoon, X. Zhuang, J. Cole, and M. Hasegawa-Johnson, "*Voice quality dependent speech recognition*", In Tseng, S. (Ed.), Linguistic Patterns of Spontaneous Speech, Special Issue of Language and Linguistics, Academica Sinica, 2007.

[17] A.K. Jain, "*Multimodal user interfaces: who's the user?*", International Conference on Multimodal Interfaces, Documents in Computing and Information Science, 2003.

[18] L. Gwo-Fong, and C. Lu-Hsien, "A non-linear rainfall-runoff model using radial basis function network", Journal of Hydrology 289, 2004.

[19] B. Azzedine, "*Behavior-based intrusion detection in mobile phone systems*", Journal of Parallel and Distributed Computing 62, 1476–1490, 2002.

[20] A.I. Fischetti and A. Andrade, "*Porosity images from well logs*", Journal of Petroleum Science and Engineering 36, 2002, 149–158.

[21] D. Gavrilis, and E. Dermatas, "Real-time detection of distributed denial-of-service attacks using RBF networks and statistical features", Computer Networks 48, 2005, 235–245.

[22] http://www.eie.polyu.edu.hk/~mwmak/Book

[23] Neural Computing Research Group, Information Engineering, Aston University, Birmingham B4 7ET, United Kingdom, http://www.ncrg.aston.ac.uk/netlab

[24] J. Han, and M. Kamber, "*Data mining concepts and techniques*", Second Edition, Morgan Kaufmann, 2006, Pp. 361.

[25] C.E. Rasmussen, "T*he infinite gaussian mixture model*", in Advances in Neural Information Processing Systems, Volume 12, Pp. 554–560, MIT Press, 2000.

[26] X. Yang, F. Kong, W Xu, and B. Liu, "*gaussian mixture density modeling and decomposition with weighted likelihood*", Proceedings of the 5th World Congress on Intelligent Control and Automation, June 15-19, 2004.

[27] P.H. Sherrod, "*DTREG predictive modeling software*", Users' Guide, 2003-2008, www.dtreg.com.

[28] S. Chen, X. Hong, and C.J. Harris, "Orthogonal forward selection for constructing the radial basis function network with tunable nodes", ICIC 2005, Part I, LNCS 3644, pp. 777–786, @ Springer-Verlag, Berlin, Heidelberg 2005.

[29] F. Anifowose, "Hybrid ai models for the characterization of oil and gas reservoirs: concept, design and implementation", VDM Verlag, Pp. 4 - 17, 2009.

[30] C. Salim, "A fuzzy ART versus hybrid NN-HMM methods for lithology identification in the Triasic province", IEEE Transactions, 0-7803-9521-2/06, 2006.

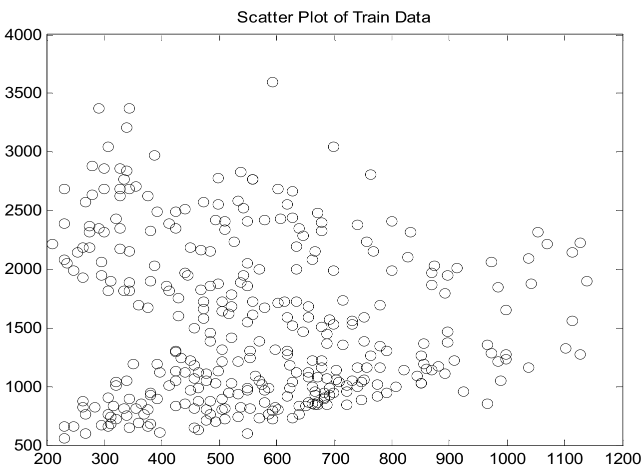
[31] S. Chikhi, and M. Batouche, "Probabilistic neural method combined with Radial-Bias functions applied to reservoir characterization in the Algerian Triassic province", Journal of Geophysics and Engineering, 1 (2004), Pp. 134–142.

[32] X. Deyi, W. Dave, Y. Tina, and R. San, "*Permeability estimation using a hybrid genetic programming and fuzzy/neural inference approach*", 2005 Society of Petroleum Engineers Annual Technical Conference and Exhibition held in Dallas, Texas, U.S.A., 9 - 12 October 2005.

[33] S. Abe, "*Fuzzy LP-SVMs for multiclass problems*", Proceedings of European Symposium on Artificial Neural Networks (ESANN'2004) Bruges, Belgium, 28-30 April 2004, d-side Publisher, ISBN 2-930307-04-8, pp. 429-434.

AUTHOR'S PROFILE

Fatai Adesina Anifowose was formerly a Research Assistant in the department of Information and Computer Science, King Fahd University of Petroleum and Minerals, Saudi Arabia. He now specializes in the application of Artificial Intelligence (AI) while working with the Center for Petroleum and Minerals at the Research Institute of the same university. He has been involved in various projects dealing with the prediction of porosity and permeability of oil and gas reservoirs using various AI techniques. He is recently interested in the hybridization of AI techniques for better performance.



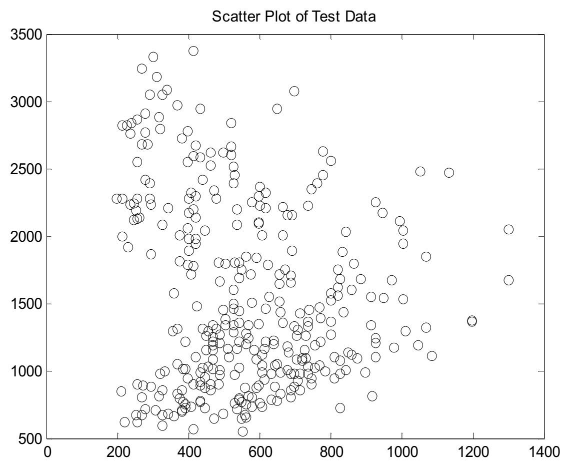


Figure 2. Scatter plot of training data with 338 observations and test data with 333 observations.

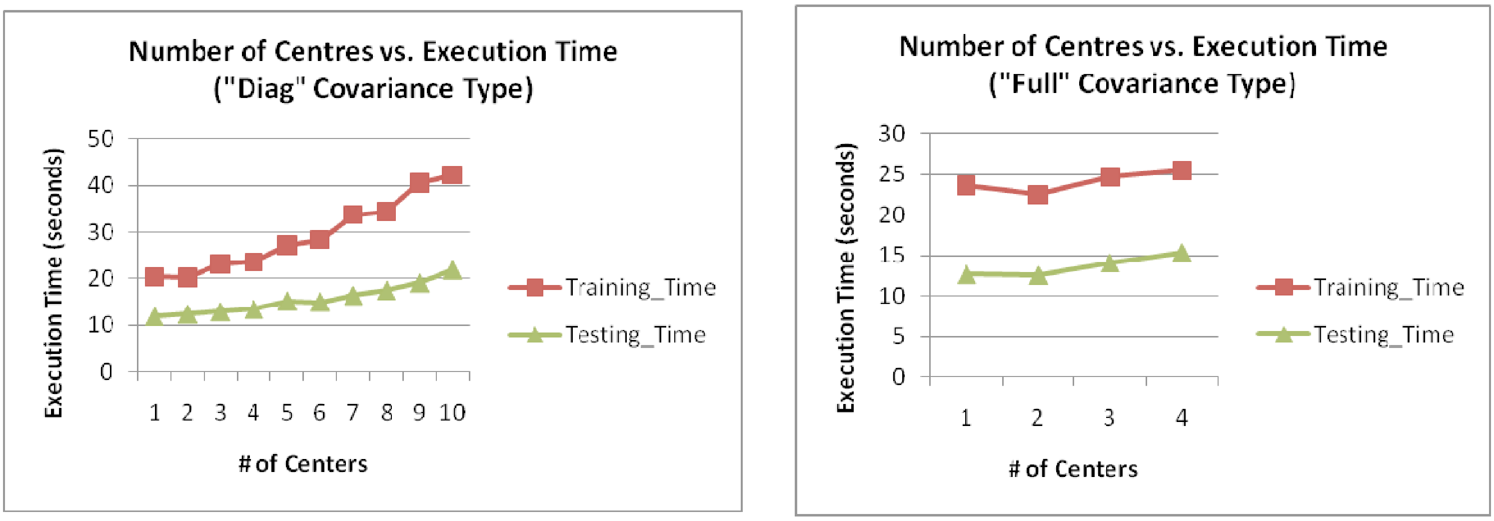


Figure 3. Relationship between the number of centers and execution time for GMM “diag” and "full" covariance types.

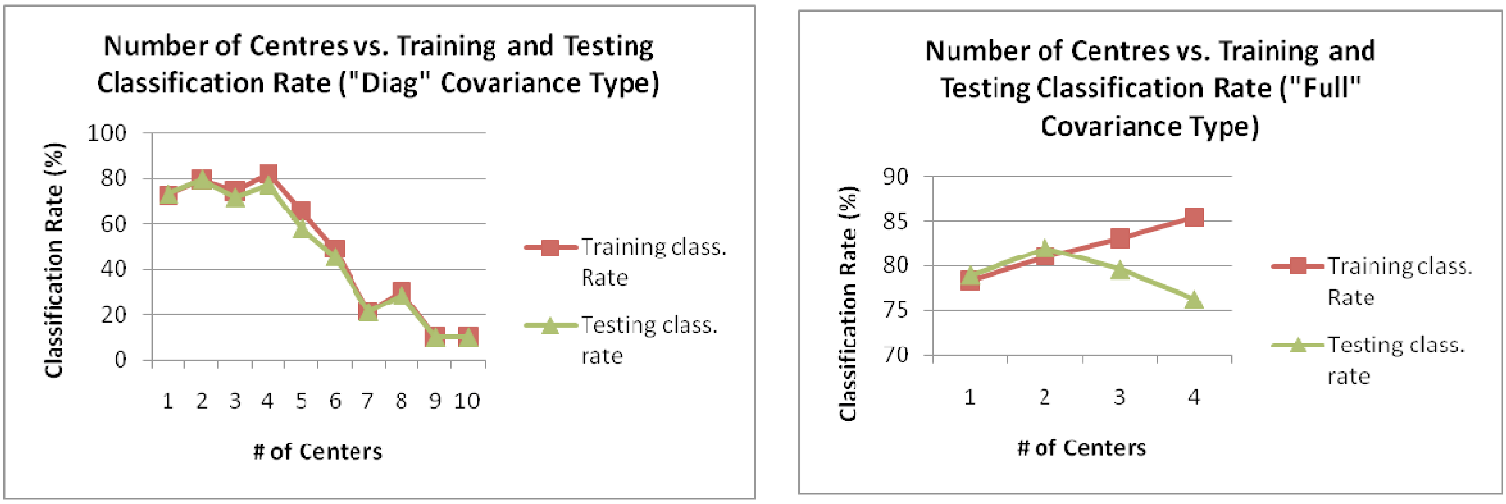


Figure 4. Relationship between the number of centers and recognition rate for GMM “diag” and "full" covariance types.

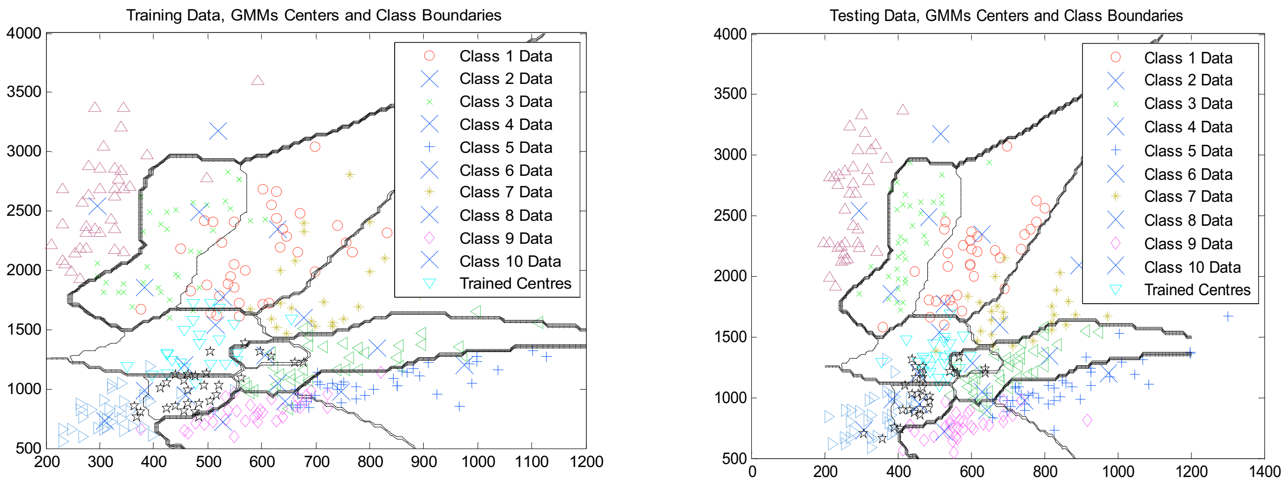


Figure 5. Class boundaries generated by the GMM Model for training and testing.

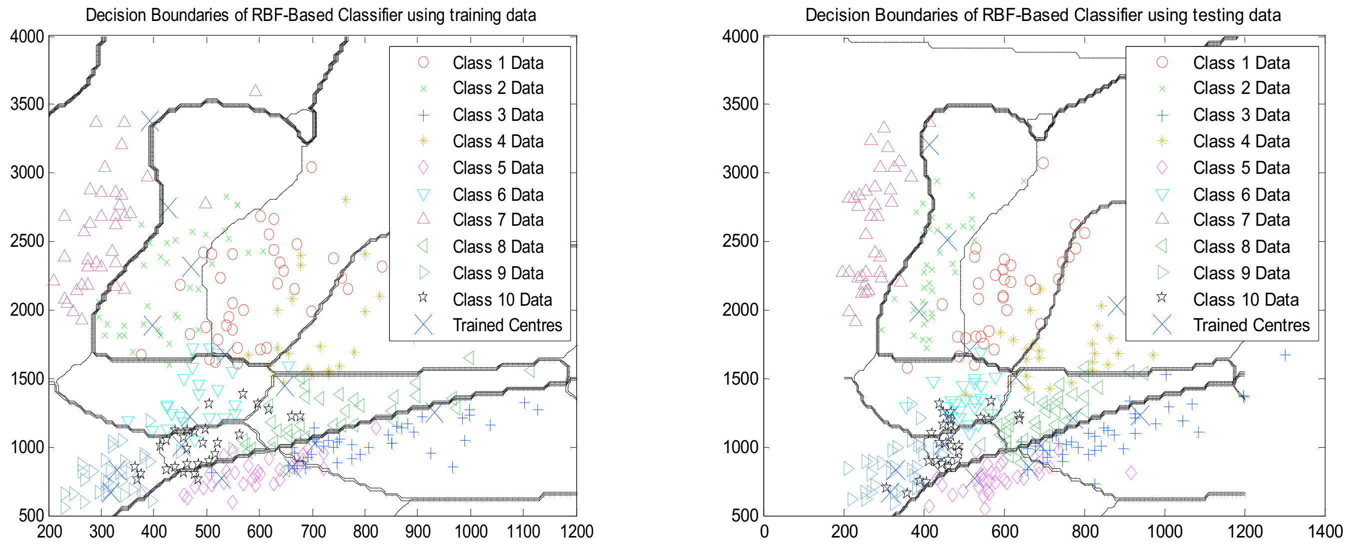


Figure 6. Decision boundaries of the RBF-based classifier using training and testing data.

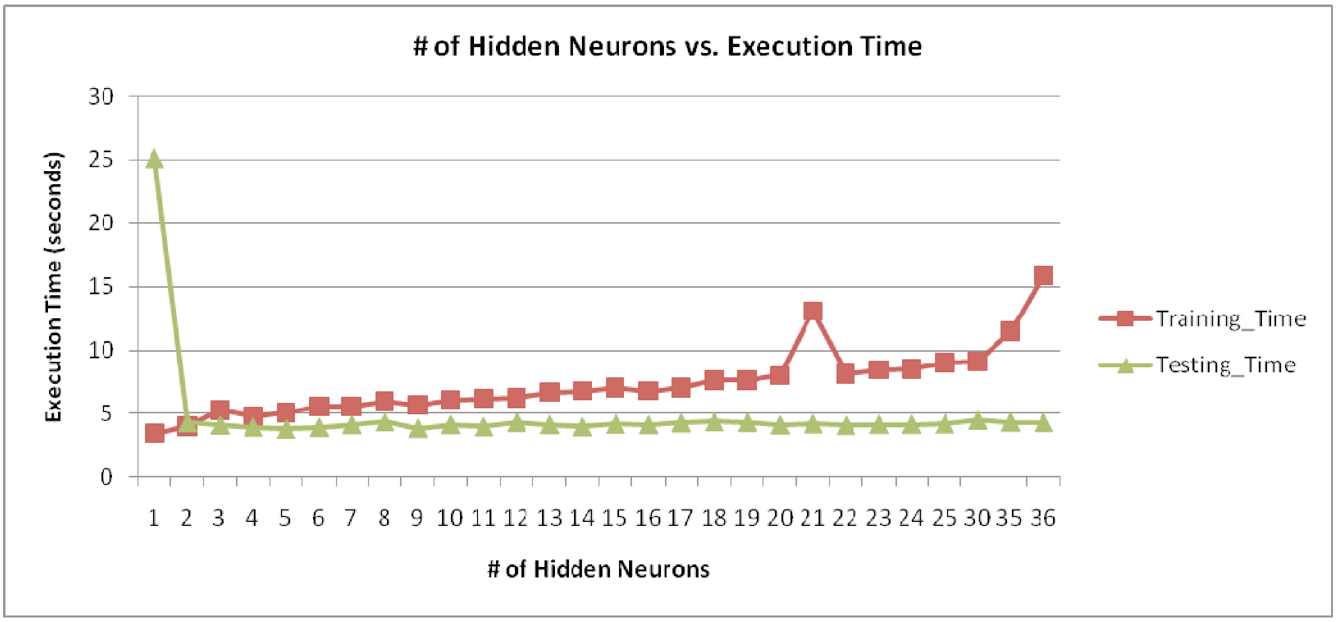


Figure 7. Relationship between the number of hidden neurons and the execution time.

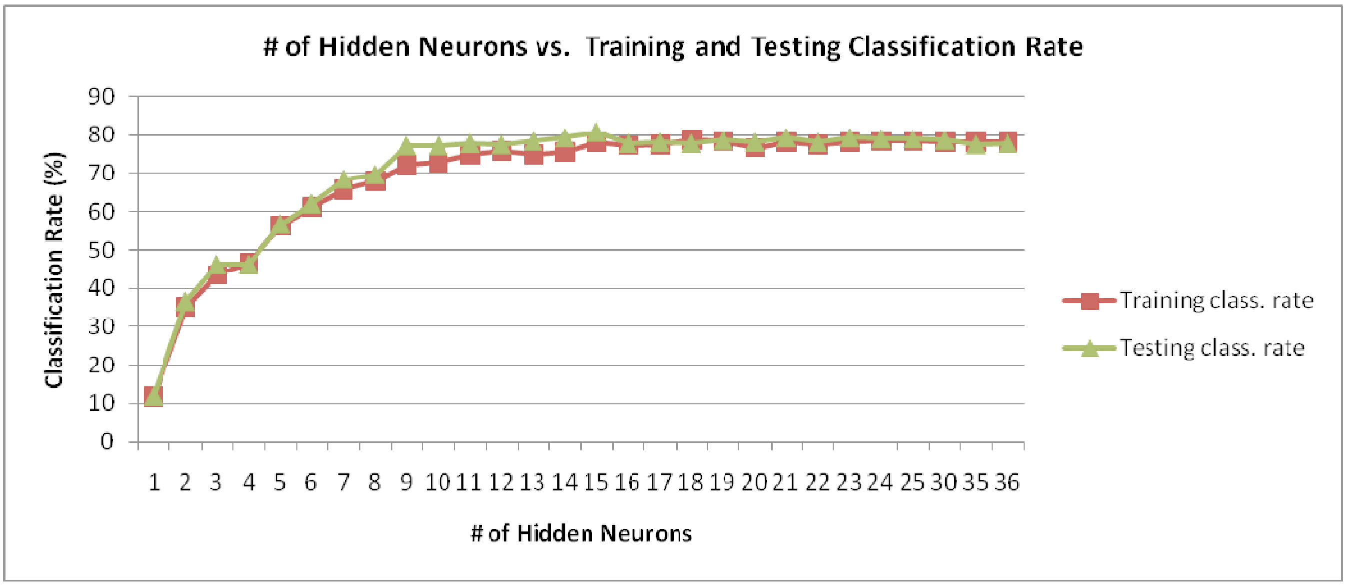


Figure 8. Relationship between the number of hidden neurons and recognition rate.

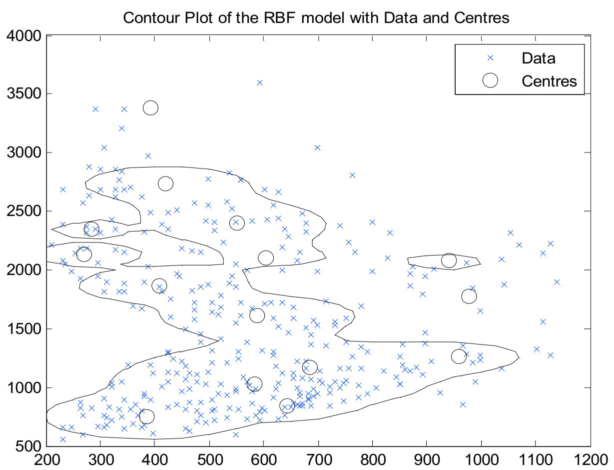


Figure 9. Contour plot of the RBF model showing the 15 hidden neurons.

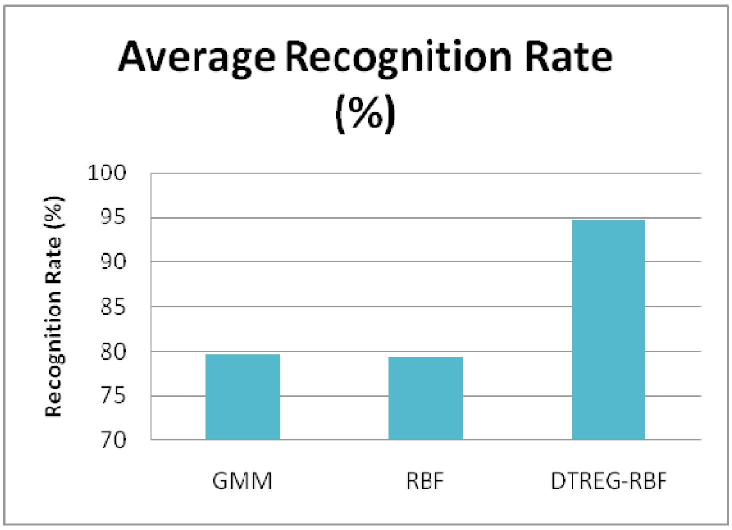


Figure 10. A comparison of GMM, RBF and DTREG RBF models by recognition rate.

Multiphase Scalable Grid Scheduler Based on Multi-QoS Using Min-Min Heuristic

Nawfal A. Mehdi, Ali Mamat, Hamidah Ibrahim, Shamala A/P K
Faculty of Computer Science and Information Technology
University Putra Malaysia,
Serdang, 43400,Selangor, Malaysia

***Abstract*—In scheduling, the main factor that affects searching speed and mapping performance is the number of resources or the size of search space. In grid computing, the scheduler performance plays an essential role in the overall performance. So, it is obvious the need for scalable scheduler that can manage the growing in resources (i.e. scalable). With the assumption that each resource has its own specifications and each job has its own requirements; then searching the whole search space (all the resources) can waste plenty of scheduling time. In this paper, we propose a two-phase scheduler that uses min-min algorithm to speed up the mapping time with almost the same efficiency. The scheduler is also based on the assumption that the resources in grid computing can be classified into clusters. The scheduler tries first to schedule the jobs to the suitable cluster (i.e. first phase) and then each cluster schedule the incoming jobs to the suitable resources (i.e. second phase). The scheduler is based on multidimensional QoS to enhance the mapping as much as it can. The simulation results show that the use of two-phase strategy can support the scalable scheduler.**

Keywords- Multi-phase; QoS; Grid Scheduling.

I. INTRODUCTION

With the development of the network technology, grid computing used to solve larger scale complex problems becomes a focus technology. The goal of schedulers is to utilize all available computational resources to overcome difficulties brought about by complicated tasks with enormous computing workloads.[1]

One of the nearest grid definition to our work is given by Ian Foster [2] "*The real and specific problem that underlies the Grid concept is coordinated resource sharing and problem solving in dynamic, multi-institutional Virtual Organizations (VO)*"[2]. We can conclude from Foster's definition: although the Grid has the characteristics of heterogeneity and dynamicity, these features are not flatly distributed in resources, but are rather distributed hierarchically and locally in many cases, due to the composition of the Grid resources. Current Grid resources are usually distributed in a clustered fashion[3]. The key technologies that affect the Grid efficiency involve Grid resource allocation, management and task scheduling algorithm.

Task scheduling is a challenging problem in grid computing environment [4] and has shown to be NP-complete in its general as well as in some restricted forms[5]. According to [6], a valid schedule is the assignment of tasks to specific time intervals of resources, such that no two tasks use any resource simultaneously, or such that the capacity of the resource is not exceeded by the tasks. The schedule of tasks is optimal if it minimizes a given optimality criterion (objective function).

Grid scheduler (GS) receives applications from grid users, selects feasible resources for these applications according to the acquired information from the Grid Information Service module, and finally generates application-to-resource mappings based on certain objective functions and predicted resource performance. Unlike their counterparts in traditional parallel and distributed systems, Grid schedulers usually cannot control Grid resources directly, but they work like brokers or agents[7]. One of the most issues in grid scheduling is the QoS; the quality of services (QoS) becomes a big concern of many Grid applications in such a non-dedicated dynamic environment. The meaning of QoS is highly dependent on particular applications, from hardware capacity to software existence. Usually, QoS is a constraint imposed on the scheduling process instead of the final objective function.[3]

This paper addresses the problem of resources growing in one search space and the ability of the main scheduler to control this growing by two phase mapping. The work in this paper is concerned with scheduling computing intensive independent task; each task requires multi QoS specification. Each task should be mapped to a cluster that can fulfill its requirement with a minimum completion time.

This work introduces the ability to schedule the tasks to a cluster to be scheduled later by the cluster's local scheduler. The main scheduler should have full information about the clusters starting from number of resources in each one to the common characteristics of the resources. Also, the main scheduler receives a set of tasks from the clients each one (i.e. task) with its QoS constraints to be mapped to the best fit cluster that can give the minimum execution time with the respect to its restrictions.

The remainder of this paper is organized as follows: in the next section II, we provide the related works. Section III, introduces task problem modeling and the new algorithm and its time complexity analysis. Section IV, shows the implementation and experiments results. Recommendations and future plan are given in section V.

II. RELATED WORKS

Over the years, task scheduling problem has become a well-recognized discipline in Grid computing and is identified as NP complete problem[8]. Many scheduling heuristics have been proposed to solve the mapping process in grid computing. Min-min heuristic depends on the minimum completion time, such that the task that has the minimum completion time is executed first. X. He *et al*.[9], proposed a QoS Guided Min-Min heuristic which can guarantee the QoS requirements of particular tasks and minimize the makespan at the same time. Wu, Shu and Zhang[10], proposed an algorithm that ordered tasks list by completion time, then segmenting the ordered list to be applied in Min-Min algorithm. They show in their results that, the algorithm can outperform the typical Min-Min. Another popular heuristic for independent scheduling is called Suffrage. The rationale behind Suffrage is that a task should be assigned to a certain host and if it does not go to that host, it will suffer the most. This algorithm has been studied by Maheswaran *et al* [11]. Muthuvelu *et al* [12] proposed a dynamic task grouping scheduling algorithm to deal with these cases. Once a set of fine grained tasks are received, the scheduler groups them according to their requirements for computation (measured in number of instructions) and the processing capability that a grid resource can provide in a certain time period. All tasks at same group are submitted to the same resource which can finish them all in the given time. Hence, the overhead for scheduling and job launching is reduced and resource utilization is increased. S´ebastien Noel *et al*[13], studied the use of a framework called YML for developing HPC applications on Grids, and proposed a multi-level scheduling architecture for it. K. Etminani and M. Naghibzadeh introduced a new scheduling algorithm based on two conventional scheduling algorithms, Min-Min and Max-Min, to use their cons and at the same time, cover their pros. It selects between the two algorithms based on the standard deviation of the expected completion time of tasks on resources. They evaluated their scheduling heuristic, the selective algorithm, within a grid simulator called GridSim. They also compared their approach to its two basic heuristics. F. M. Ciorba *et al* [15], studied the problem of scheduling loops with iteration dependencies for heterogeneous (dedicated and non-dedicated) clusters. The presence of iteration dependencies incurs an extra degree of difficulty and makes the development of such schemes quite a challenge. They extended three well known dynamic schemes (CSS, TSS and DTSS) by introducing synchronization points at certain intervals so that processors compute in pipelined fashion. Their scheme is called Dynamic Multi-Phase Scheduling (DMPS) and they applied it to loops with iteration dependencies. They implemented their new scheme on a network of heterogeneous computers and studied its performance. Through extensive testing on two real-life applications (the heat equation and the Floyd-Steinberg algorithm), they showed that the proposed method is efficient for parallelizing nested loops with dependencies on heterogeneous systems.

III. TASK SCHEDULING PROBLEM

This work is based on scheduling the tasks in two phases to reduce the search space for the scheduler. The proposed algorithm should already have the set of clusters that is available at that time. Each cluster should come with its specifications that is used to fit with user's QoS restrictions. Also, the algorithm takes a set of tasks, each one with its QoS restrictions.

A. Problem Modeling

We model the scheduling problem by $E_i=(J_i,C_j)$, where J_i is a job, C_j is a cluster and E_i is the mapping.

Jobs are defined in this work as:

- J is the set of M jobs such that $J=\{J_1,J=,\ldots,J_M\}$. Each job J_i has four QoS characteristics that are described in details in the next point.
- Q is the set of QoS dimensions that is attached with each job J_i such that $Q=\{L_i, S_i, SE_i, BW_i\}$,where
  - L_i is the length of the job J_i.
  - S_i is the maximum cost that can be paid by job J_i.
  - SF_i is the security value that represents the amount of security needed by J_i.
  - BW_i is the amount of network bandwidth that is needed by J_i.

Clusters are defined in this work as:

- C is the set of N clusters such that $C=\{C_1,C_2,\ldots.,C_N\}$. Each cluster C_i has four properties.
- P is the set of five properties attached with each cluster C_i such that, $P_i=\{SP_i,CB_i,CC_i,CS_i,Z_i\}$ where:
  - SP_j is the speed of cluster C_j.
  - CB_j is the bandwidth offered by cluster C_j.
  - CC_j is the cost/hour offered by cluster C_j.
  - CS_j is the security value that represents the amount of security offered by C_j.
  - Z_j is the size of cluster C_i (i.e. the number of resources)
- R_j is a set of size Z_j represent the resources' ready time for cluster C_j.

To model the servers in our work, we suppose:-

- RS is the set of W resources such that $RS=\{RS_1,RS_2,\ldots.,RS_W\}$. Each resource RS_i has four properties.
- PS is the set of five properties attached with each resource $\mathrm{RS_i}$ such that, $PS_i=\{SP_i,RN_i,RC_i,RS_i,RD_i\}$ where:
  - RSP_j is the speed of Resource RS_i.

- RB_i is the bandwidth offered by Resource RS_i.
- RC_i is the cost/hour offered by Resource RS_i.
- RS_i is the security flag that is set if the Resource RS_i offered security.
- RD_i is the ready time for resource RS_i.

In the cluster's class there is an R_j field that is responsible for holding the ready time for each host inside the cluster. This list should be always in ascending order to facilitate selecting the best cluster. Initialed to zero, this list is firstly. Each job's class has two fields, first one (*TCT*) is responsible for holding the best completion time offered by a cluster that its address is held in the second field cluster index (*Clr_ndx*).

In this work, ET_{ij} represents the expected completion time of task J_i on a host in cluster C_j. First(R_j) represents the best ready time for cluster C_j. CE_{ij} represents the expected completion time of task J_i on a host in cluster C_j. EC_{ij} represents the expected cost to execute job J_i in cluster C_j.

```
1.  While (J is not empty) do
2.    For each job Ji in J do
3.      Ji.TCT=Double.Max_value    //TCT= Temporary Completion Time
4.        For each cluster Cj do
5.          ETij=Li/SPj
6.          CTij=ETij+first(Ri)
7.          ECij=(Li /SPi)* CCi
8.          If (Ji.TCT>CTij) and (Ji.SF=Cj.CS) and (Ji.S<= ECij)
            and(Ji.BW<=Cj.CB) then
9.            Ji.TCT= CTij
10.                   Ji.Clr_ndx=j
11.             End if
12.           End For
13.         If (there is no match) then
14.           Print out Job Ji has no match
15.           Delete Ji from J
16.         End if
17.         Else
18.             If (Ji has minimum Completion time) then
19.                       Set Min_Clstr=j
20.                       Set Min_Job=i
21.             End if
22.       End For
23.     Map JMin_job to CMin_Clstr
24.     Delete JMin_job from J
25.     Update RMin_Clstr such that the set should stay sorted in ascending
  order.
26.   End While
```

Figure 1. Global grid Scheduler Algorithm

This algorithm computes the expected completion time for all tasks on all clusters using these equations:

$$ET_{ij} = SP_j / L_i \tag{1}$$

$$CT_{ij} = ET_{ij} + first(R_j) \tag{2}$$

Then it computes the expected cost using (3):

$$EC_{ij} = ET_{ij} * CC_j \tag{3}$$

This algorithm has loop J_i (line 4..line 12) that finds the best cluster that fulfill J_i QoS constraints and has the minimum completion time by using equations (1,2,3). After the loop J_i, an If condition (line 13) checks if the J_i got any host that can fulfill its constraints. If there is no such a host then delete this job (J_i) from the job list, otherwise check again (line 18) if this J_i has the minimum execution time and save its index if true. At line 23 we have J_{Min_job} that holds the index for the minimum completion time job, so we map it to its cluster C_{Min_Clstr}. Line 25 is responsible for updating the list *R* in such a way it stays in ascending order.

B. Algorithm Analysis

The time complexity of the proposed algorithm is:

$$f(n,m) = O(m \log m) \cdot O(n) \tag{4}$$

Where m is the number of jobs and n is the number of clusters. From above, we can see that this algorithm has a little effect by the increase in the number of servers inside the clusters because updating servers list required just *log(Zj)* where *Zj* is the number of servers inside cluster *Cj*.

In comparison with this algorithm, the time complexity for the old algorithm is:

$$f(n,m) = O(m \log m) \cdot O(w) \tag{5}$$

Where *w* is the number of servers in the cluster.
Therefore, it is quite clear the effect of increasing the number of servers on the proposed algorithm is not that much intense.

C. Quality of Service (QoS)

This work uses QoS restriction to find the suitable cluster that can execute user's tasks. Multi-dimensions QoS have been used so that the users should submit their tasks with many parameters. These parameters are:

- **Bandwidth**: The user should submit his task with the minimum amount of bandwidth needed to execute it. Bandwidth is set to zero in case it does not need any bandwidth.
- **Security**: These days, the most important issue in distributed system is the security and its type. In this work, we proposed a multi-type QoS security check. It means the algorithm can check for the user the suitable type that he needs to execute the task. Security parameter is an integer value, where each value represents a type or level of security.
- **Cost**: Budget cost is the amount of payment from a user to a resource for its service. Here the user should specify the maximum cost, which can be afforded.

IV. Implementation and Experiment Results

This algorithm is used in the first or higher level, while in the second level, the normal MM is used.

We use Java programming language in order to implement the simulator to test the proposed algorithm. The implementation consists of several classes, these are:

- **Create population**: This class is responsible for creating the set of tasks with its QoS restrictions, set of clusters with their specifications and a set of servers to be used as for old algorithm. The size of set of servers is equal to the number of clusters multiplied by the number of servers in each cluster. The number of tasks, clusters and servers/cluster are fixed, and the QoS restrictions and the clusters specifications are generated randomly.
- **New Min-Min**: this class is an implementation for the improved min-min that (2PMM) is responsible for mapping the tasks to the appropriate cluster.
- **Old Min-Min**: this class is an implementation for the old algorithm to be used for performance comparison.

Firstly, Create population class generates 1000 tasks in one list, N clusters each one with Z servers and list of W servers such that:

$$W = N * Z \qquad (6)$$

Secondly, *Old Min-Min* and *New Min-Min* start working to make the mapping and compute the performance metrics which is the makespan. Makespan can be define as the time spent from the beginning of the first job to the end of the last job.

Two experiments have been made to test the performance of 2PMM algorithm. Each experiment consists of six sizes (i.e. number of clusters and servers). The test for each size is made ten times and the average has been taken for the comparison.

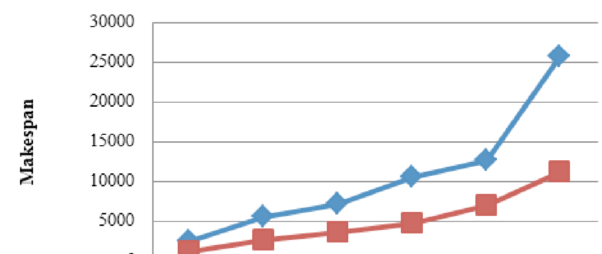
The first experiment compares the performance and cost in both old and new algorithms (figure 2). This figure shows the effect of increasing of servers on mapping time. The Y-axis in this figure represents the total execution time for the mapping process, while the x-axis represents the number of servers and it is written in form of equation (6)(i.e. 10*5=50 means, 10 clusters and 5 servers in each cluster as a test bed for the new algorithm and 50 servers as a test bed for the old algorithm). In this experiment, we fixed the number of tasks to 1000 and the number of clusters to 10 and changed the number of servers in each cluster. It is quite clear that the effect of increasing the number of servers (i.e. increasing the search space) on the execution time of the scheduler is not that much intense. Figure (4) shows the improvement mapping time between 2PMM and MM algorithms.

The second experiment (figure 3) shows the influence of increasing the number clusters on the new algorithm. As in figure (2), the y-axis represents the mapping execution time while the x-axis represents the number of clusters, number of servers for each cluster and total number of servers. We can see that the total execution time is directly affected by the number of clusters in its search space but its time is still far from the time needed in the old algorithm.

V. Conclusion and Future Work

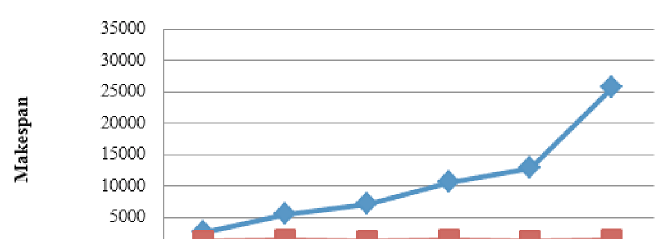
This paper investigates the job scheduling algorithm in grid environments as an optimization problem. The proposal is to minimize the scheduling time for urgent jobs, by mapping the jobs to the best cluster as the first phase and then reschedule to the best resource in the selected cluster.

The algorithm is developed based on Min-Min Algorithm to find the proper cluster that can execute the job with minimum execution time with respect to QoS job requirements. The improved algorithm is compared with the previous Min-Min algorithm. The results show a better performance in scheduling time point of view. It can map the jobs faster than the normal Min-Min. The future work will focus on clustering algorithms and study the effect of three phase clustering on the system.



	10*5	20*5	30*5	40*5	60*5	100*5
MM	2465.2	5454.3	7115.2	10466.7	12696.3	25646.4
2PMM	1149.6	2529.2	3619.9	4698	6921.3	11109

Figure 2. The effect of increasing the number of clusters with fixed number of servers on makespan



Cluster*Server	10*5	10*10	10*15	10*20	10*30	10*50
MM	2465.2	5454.3	7115.2	10466.7	12696.3	25646.4
2PMM	1076.1	1254.2	1130.9	1258.9	1026.1	1229.2

Figure 3. The effect of increasing the number of servers with fixed number of clusters on Makespan.

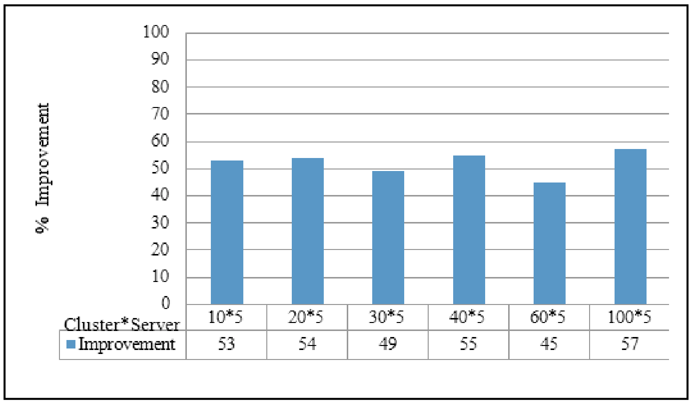


Figure 4. The percentage of makespan improvement when increasing the number of clusters with fixed number of servers.

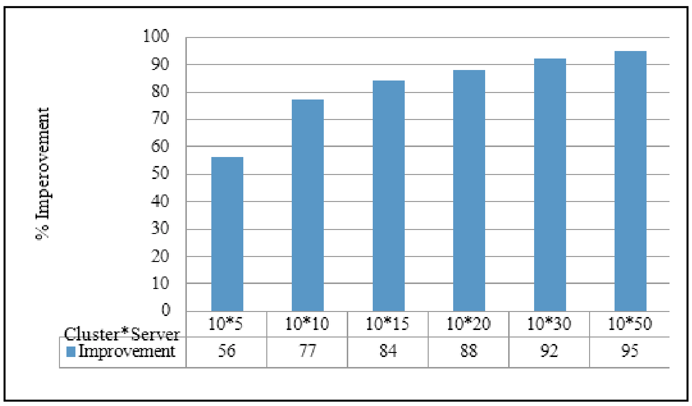


Figure 5. The percentage of makespan improvement when increasing the number of servers with fixed number of clusters

REFERENCES

[1] Ehsan Ullah Munir, Jianzhong Li, and Shengfei Shi, "QoS Sufferage Heuristic for Independent Task Scheduling in Grid," Information Technology, vol. 6, no. 7. pp.1166-1179, 2007.

[2] I. Foster, C. Kesselman, and S. Tuecke, "The Anatomy of the Grid: Enabling Scalable Virtual Organizations," International Journal of High Performance Computing Applications, vol. 15, no. 3. pp.200, 2001.

[3] F. Dong and S. G. Akl, "Scheduling Algorithms for Grid Computing: State of the Art and Open Problems," Queen's University School of Computing.January, 2006.

[4] I. Foster and C. Kesselman, The Grid: Blueprint for a New Computing Infrastructure: Morgan Kaufmann, 2004.

[5] H. El-Rewini, T. G. Lewis, and H. H. Ali, Task scheduling in parallel and distributed systems, 1994.

[6] P. Fibich, L. Matyska, and H. Rudovb, "Model of Grid Scheduling Problem," Exploring Planning and Scheduling for Web Services, Grid and Autonomic Computing. pp.05-03

[7] F. Berman, R. Wolski, H. Casanova et al., "Adaptive computing on the Grid using AppLeS," Parallel and Distributed Systems, IEEE Transactions on, vol. 14, no. 4. pp.369-382, 2003.

[8] O. Sinnen and I. NetLibrary,Task Scheduling for Parallel Systems: Wiley-Interscience, 2007.

[9] H. E. XiaoShan, S. U. N. XianHe, and G. von Laszewski, "QoS Guided Min-Min Heuristic for Grid Task Scheduling," Journal of Computer Science and Technology, vol. 18, no. 4, 2003.

[10] M. Y. Wu, W. Shu, and H. Zhang, "Segmented min-min: A static mapping algorithm for meta-tasks on heterogeneous computing systems." 9th IEEE Heterogeneous Computing Workshop (HCW 2000) , pp. 375-385. 2000.

[11] M. Maheswaran, S. Ali, H. J. Siegel et al., "Dynamic matching and scheduling of a class of independent tasks onto heterogeneous computing systems." 8th Heterogeneous Computing Workshop (HCWÆ99) , pp. 30-44. 1999.

[12] N. Muthuvelu, J. Liu, N. L. Soe et al., "A dynamic job grouping-based scheduling for deploying applications with fine-grained tasks on global grids." Proceedings of the 2005 Australasian workshop on Grid computing and e-research-Volume 44 , pp. 41-48. 2005. Australian Computer Society, Inc. Darlinghurst, Australia, Australia.

[13] S. Noel, O. Delannoy, N. Emad et al., "A Multi-level Scheduler for the Grid Computing YML Framework," LECTURE NOTES IN COMPUTER SCIENCE, vol. 4375. pp.87, 2007.

[14] K. Etminani and M. Naghibzadeh, "A Min-Min Max-Min selective algorihtm for grid task scheduling." Internet, 2007.ICI 2007.3rd IEEE/IFIP International Conference in Central Asia on , pp. 1-7. 2007.

[15] F. M. Ciorba, T. Andronikos, I. Riakiotakis et al., "Dynamic Multi Phase Scheduling for Heterogeneous Clusters." Proc.of the 20th IEEE IntÆl Par.& Dist.Proc.Symp.(IPDPSÆ06), Greece . 2006.

Loss Reduction in Distribution System Using Fuzzy Techniques

Sheeraz kirmani, Md. Farrukh Rahman, Chakresh Kumar
Department of Electrical and Electronics Engineering
Manav Rachna International University Faridabad India
Email: sheerazkirmani@gmail.com,farrukh24rahman@gmail.com,chakreshk@gmail.com

***Abstract*- In this paper, a novel approach using approximate reasoning is used to determine suitable candidate nodes in a distribution system for capacitor placement. Voltages and power loss reduction indices of distribution system nodes are modelled by furzy membership functions. A fuzzy expert system (FES) containing a set of heuristic rules is then used to determine the capacitor placement suitability of each node in the distribution system. Capacitors are placed on the nodes with the highest suitability. a new design methodology for determining the size, location, type and number of capacitors to be placed on a radial distribution system is presented. The objective is to minimize the peak power losses and the energy losses in the distribution system considering the capacitor cost. Test results have been presented along with the discussion of the algorithm.**

***Keywords - Capacitor placement, distribution systems, fuzzy expert system*.**

I. INTRODUCTION

Efficiency of power system depends on distribution system. Distribution system provides the final link between the high voltage transmission system and the consumers. A distribution circuit normally uses primary or main feeders and lateral distributors. The main feeder originates from the substation, and passes through the major load centers. Lateral distributors connect the individual load points to the main feeder with distribution transformers at their ends. Many distribution systems used in practice have a single circuit main feeder and are defined as radial distribution systems. Radial systems are popular because of their simple design and generally low cost [4].

Capacitor placement problem has been extensively discussed in technical literature especially since 1980's as the distribution system planning and operation started getting renewed focus. Since then, many solution techniques have been suggested identifying the problem as a complex problem of large scale mixed integer non-linear programming.

Analytical techniques [8]–[11], heuristics [12], [13], mathematical programming [6] and a host of other methods have been developed to solve the problem.

Artificial intelligent techniques have been tried in recent years in search of a superior solution tool. With rapid growth of computing power, new class of search techniques capable of handling large and complex problems has been developed during the last few decades. These techniques have also been explored for the solution of the capacitor placement problem. Among these techniques, evolutionary computing methods such as Genetic algorithm [14], [15] and Ant colony optimization [9] have been reported to produce superior results. Simulated annealing [10] and Tabu searches [11] had also been very successful. However, one common drawback of these techniques lies in the huge computing task involved in obtaining the solution On the other hand, there had always been the efforts of the system engineers to avoid applications of computation intensive complex solution processes and use simple, physically understandable logics to solve the problems, though such simplified solutions occasionally can not find the best one. Fuzzy based approaches [9]–[12] involve less computational burden.

The power loss in a distribution system is significantly high because of lower voltage and hence high current, compared to that in a high voltage transmission system [5]. The pressure of improving the overall efficiency of power delivery has forced the power utilities to reduce the loss, especially at the distribution level. In this paper a radial distribution system is taken because of its simplicity.

Fuzzy based solution methods use fuzzy membership functions to model the actual systems. Identification of proper membership function is the most challenging task in the development of fuzzy based solution techniques. Node voltage measures and power loss in the network branches have been utilized as indicators for deciding the location and also the size of the capacitors in fuzzy based capacitor placement methods.

II FRAME WORK OF APPROACH

For capacitor placement, general considerations are:

(1)the number and location;
(2) type(fixed or switched);
(3) the size;

When capacitors are placed power loss is reduced & also energy loss is reduced. Both these factors contribute in increasing the profit. Cost of capacitors decreases this profit. So profit is weighted against the cost of capacitor installation [1]. Whole saving can be given as follows [3].

$$S = K_P \Delta P + K_E \Delta E - K_C C \qquad \textbf{(1)}$$

Where-
K_P-Per unit cost of peak power loss reduction ($/KW)
K_E-Per unit cost of energy loss reduction ($/KWh)
K_C-Per unit cost of capacitor ($/KVar)
Δp-Peak power loss reduction (KW)
ΔE-Energy loss reduction (KWh)
C-Capacitor size (KVar)
S-Saving in money per year ($/year)

Then by optimising the profit 'S' due to capacitor placement actual capacitor size is determined i.e. by setting $\partial S / \partial C$=0, and then solving for C, the capacitor size. The above procedure is repeated until no additional savings from the installation of capacitors are achieved.

For each solution voltage constraint must be satisfied. Voltage (pu) should be between min (0.9) to max (1.1).i.e.

$$V_{\min} \leq V \leq V_{\max} \qquad (2)$$

In this paper shunt (fixed) capacitors are used. A simple 10 bus radial distribution system is taken as the test system. It has only main feeder & no branches. To determine the location & size of capacitors to be installed, a load flow program was executed on MATLAB. This gave the location of capacitor most suitable for capacitor placement. **Shunt capacitors** to be placed at the nodes of the system have been represented as **reactive power injections [3]**.

III. ALGORITHM ADOPTED FOR LOAD FLOW SOLUTION

A balanced three-phase radial distribution network is assumed and can be represented by its equivalent single line diagram [2]. Line shunt capacitance is negligible at the distribution voltage levels. The algorithm for capacitor location finding & sizing is as follows:

1. Perform load flow program to calculate bus voltages and segment losses.

2. Find the membership functions of voltage drops, power loss and suitability of capacitor node, and decision for the fuzzy sets of voltage drops, power loss and capacitor nodes.

3. Identify the node having highest suitability ranking.
4. Install a capacitor at optimal node (s). Select capacitor that has the lowest cost and size.

5. Check whether voltage constraint is satisfied. If yes, go to next step, otherwise, go to step- 9.
6. Compute the benefits due to reduction in peak power loss, energy loss and cost of capacitor banks and net savings.
7. Check whether net savings is greater than zero. If yes, go to next step, otherwise, go to step-9.

8. Increment size of capacitor bank and go to step-2.
9. Reject the installation.

Compensation of each bus' reactive power demand is done by placing capacitor. Calculation of power loss reduction & voltage were done thereafter .Highest power loss reduction was assigned '1' & lowest loss reduction was assigned '0'. All other power loss reductions were placed between 0 &1.Voltage is also given in pu values [6].

IV. CAPACITOR LOCATION FINDING USING FUZZY TECHNIQUES

For the capacitor allocation problem, rules are defined to determine the suitability of a node for capacitor installation. Such rules are expressed in the following form:

IF premise (antecedent), THEN conclusion (consequent)

For determining the suitability of capacitor placement at a particular node, a set of multiple-antecedent fuzzy rules have been established. The inputs to the rules are the voltage and power loss indices, and the output consequent is the suitability of capacitor placement. As given in table I.

The consequents of the rules are in the shaded part of the matrix. The fuzzy variables, power loss reduction, voltage, and capacitor placement suitability are described by the fuzzy terms *high, high-medium/normal, medium/normal, low-medium/normal* or *low*. These fuzzy variables described by linguistic terms are described by the fuzzy terms *high, high-medium/normal, medium/normal, low-medium/normal* or *low* [2].

These fuzzy variables described by linguistic terms are represented by membership functions. The membership functions are graphically shown in Fig.1,2 & 3.The membership functions for describing the voltage have been created based on Ontario Hydro Standards of acceptable operating voltage ranges for distribution systems [6]. The membership functions for the PLRI and CPSI indices are created to provide a ranking. Therefore, partitions of the membership functions for the power and suitability indices are equally spaced apart.

V. IMPLEMENTATION OF FUZZY ALGORITHM FOR CAPACITOR SIZING

A 10 bus radial distribution feeder with 23 KV rated voltage system is taken as the main system. 1st bus is source bus & other 9 buses are load bus.

TABLE I DECISION MATRIX FOR DETERMINING SUITABLE CAPACITOR LOCATION

AND		VI					
		V-Low	Low	Lo-Normal	Normal	Hi-Normal	High
PLRI	Low	Med	Lo-Med	Lo-Med	Low	Low	Low
	Lo-Med	Hi-Med	Med	Lo-Med	Lo-Med	Low	Low
	Med	High	Hi-Med	Med	Lo-Med	Low	Low
	Hi-Med	High	Hi-Med	Hi-Med	Med	Lo-Med	Low
	High	High	High	Hi-Med	Med	Lo-Med	Lo-Med

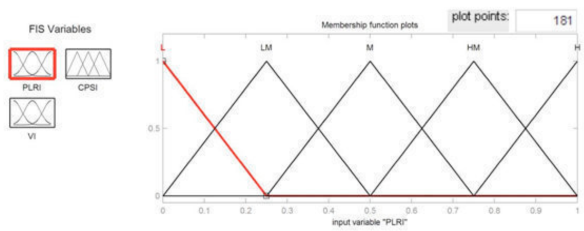


Figure1-Input 1(PLRI) membership function

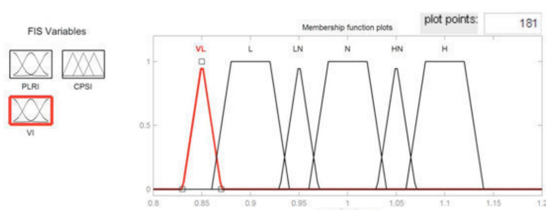
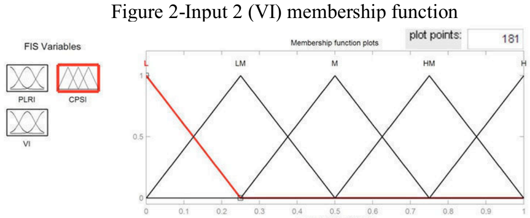


Figure 2-Input 2 (VI) membership function



S

Figure 3-Output membership (CPSI) function

Rule base

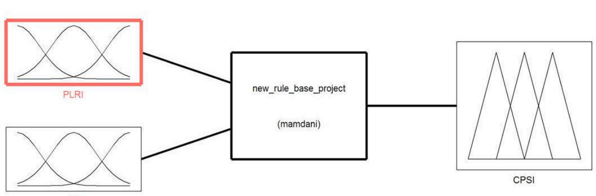


Figure 4-Rule base

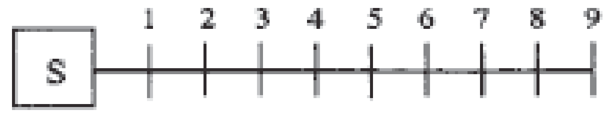


Figure 5- 10 bus radial distribution feeder

Its line data & Bus data is given in table 2 & 3

TABLE 2 LOAD DATA OF TEST SYSTEM

Bus #	1	2	3	4	5	6	7	8	9
P(Kw)	1840	980	1790	1598	1610	780	1150	980	1640
Q(Kvar)	460	340	446	1840	600	110	60	130	200

TABLE 3 BUS DATA OF TEST SYSTEM

From Bus# i	To Bus # i+1	$R_{i,i+1}$ (Ω)	$X_{i,i+1}$ (Ω)
0	1	.1233	.4127
1	2	.0140	.6051
2	3	.7463	1.2050
3	4	.6984	.6084
4	5	1.9831	1.7276
5	6	.9053	.7886
6	7	2.0552	1.1640
7	8	4.7953	2.7160
8	9	5.3434	3.0264

First bus is source bus. All the other 9 load buses were fully compensated by placing capacitors. Then power loss reduction in the entire system is calculated by load flow program using MATLAB. Both the power loss reduction index (PLRI) & voltage sensitivity index (VI) is scaled in pu values. Based on these two values capacitor placement suitability index (CPSI) for each bus is determined by using fuzzy toolbox in MATLAB. As shown in table 4.The bus which is in urgent need of balancing will give maximum CPSI. Buses which are already balanced will give lesser values. Bus which gives highest values of CPSI is first considered for capacitor placement. Then value of capacitor to be place is determined.

TABLE 4 –BUS LOCATION FINDING FOR CAPACITOR PLACEMENT

Bus	PLRI(Input 1)	VI(Input 2)	CPSI(Output)
1	0	0.993	0.08
2	0.031	0.983	0.14
3	0.176	0.960	0.25
4	1	0.953	0.75
5	0.49	0.918	0.73
6	0.084	0.903	0.34
7	0.039	0.884	0.30
8	0.144	0.855	0.64
9	0.246	0.837	0.74

Bus 4 has highest CPSI, so its selected for capacitor placement. Now value of capacitor is to be found. So equation (1) was used for saving calculation.

$S=K_P\Delta p+K_E\Delta E-K_C C$

Where
K_P=$120/KW
K_E=$0.3/KWh
K_C=$5/KVar
Load factor= 0.56 (Main feeder) [3].

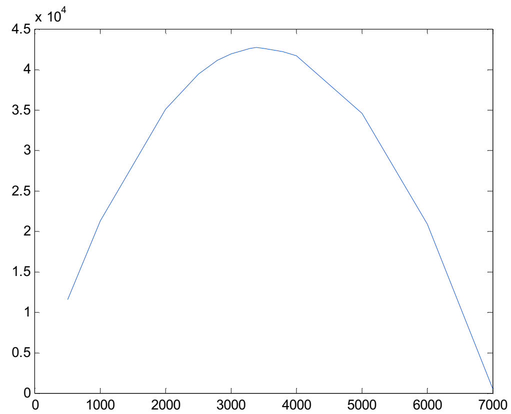


Figure 4-Curve of C Vs S for bus 4.

From load flow program on MATLAB relevant data is obtained, and a graph between C & S for bus 4 is plotted. **S is max for C=3400KVar**.So capacitor of this value is installed on bus 4. After bus 4 same process is repeated. First location is determined by fuzzy techniques, then saving is calculated for different capacitor values. C-S graphs are plotted for other buses. capacitor corresponding to maximum saving is the required capacitor.

VI Result & discussion

Table 5,6 & 7 shows results after placement of capacitors.

A. Savings: As power & energy loss is reduced and power factor improves, so there is a net benefit in installing the capacitors.

TABLE 5 CAPACITOR LOCATION, VALUE AND SAVING

S. No.	Capacitor location (Bus No)	Capacitor value (KVar)	Saving($)
1	4	3400	42650
2	5	400	1038
3	9	400	7036
Total saving			50,724/-

B.Voltage stabilisation: There is a considerable improvement in voltage profile after the compensation of system. It satisfies the voltage constraint.

TABLE 6 VOLTAGE IMPROVEMENT

	Before compensation	After compensation
Minimum voltage(pu)	0.85	0.91
Maximum voltage(pu)	0.990	0.996

C .Power & Energy loss reduction-As a result of capacitor placement reactive power is compensated as a result power factor of the system improves. So both energy loss & power loss reduces. Data is obtained from load flow programme on MATLAB.

TABLE 7 POWER AND ENERGY LOSS REDUCTION

	Before compensation	After compensation
Power loss (KW)	861.4	751.9
Energy loss (KWh)	236866	214594.9

VII. CONCLUSION

An approach incorporating the use of fuzzy sets theory has been presented in this project to determine the optimal number, locations and ratings of capacitors to place in a distribution system. In choosing the ideal locations for capacitor placement, a compromise of the reactive losses and the voltage sensitivity is determined. Application of this method to a sample test system has shown its effectiveness in peak power and energy loss reductions, and improvement in voltage regulation. The same procedure with some additional considerations can be successfully applied to complex systems having sub feeders or system with more buses... In addition, this algorithm can easily he adapted for capacitor allocation in distribution system planning, expansion or operation.

REFERENCES

[1] S.K.Bhattacharya, and S.K.Goswami, "Improved Fuzzy Based Capacitor Placement Method for Radial Distribution System". *IEEE Trans. Power Apparatus and Systems*, vol. 108, no. 4, pp. 741–944, Apr. 2008.

[2] S. M. Kannan, Member, IEEE, A. Rathina Grace Monica, and S. Mary Raja Slochanal, "Fuzzy Logic Based Optimal Capacitor Placement on Radial Distribution Feeders *IEEE Trans. on Power Apparatus and Systems, vol. 100, pp.1105-1118,2008.*

[3] H.N.Ng, N.M.A. Salama and .Y. Chikhani, "Capacitor allocation by Approximate reasoning fuzzy capacitor placement", IEEE Trans. Power Delivery, Vol. 15, issue. 1, P.393-398, January 2000.

[4] Hawary, "Load Flow Solution of Radial Distribution Feeders: a new contribution", International journal of Electrical Power and Energy systems,24 (9),P.701707,Nov 2002.

[5] C.L.Wadhwa, Electrical power systems, New age international (P) limited, New Delhi, India, 2005.

[6] M. M. A. Salama, A. Y. Chikhani, R. Hackam, and E. A. A. Mansour, "Control of reactive power in distribution systems with an end-load and fixed load condition," *IEEE Trans. Power*

Apparatus and Systems, vol. 104, no. 4, pp. 941–947, Apr.1995.

[7] C. Chin, W. M. Lin, "Capacitor Placements for Distribution Systems with Fuzzy Algorithm", *Proceedings of the 1994 Region 10 Ninth Annual International Conference,* 1994, pp-1025 - 1029.

[8] Y. G. Bae, "Analytical method of capacitor allocation on distribution primary feeders", *IEEE Trans. Power Apparatus and Systems,* vol. 97, no. 11,pp. 1232–1238, July/Aug. 1978.

[9] J.J. Grainger, and S. H. L.ee, "Optimum Size and Location of Shunt Capacitors for Reduction of Losses on Distribution Feeders", *IEEE Trans. on Power Apparatus and Systems, vol. 100, pp.1105-1118, March 1981*

[10] J. J. Grainger, S. H. Lee, "Capacity Release by Shunt Capacitor Placement on Distribution Feeders: A New Voltage Dependent Model", *IEEE Trans. on Power Apparatus and Systems, pp. 1236-1244, August 1982.*

[11] M.H. Haque, "Capacitor Placement in Radial Distribution Systems for loss Reduction", *IEE Proceedings, Generation, Transmission and Distrbution, Vol. 146 issue:5, sp. 1999, pp. 501-505.*

[12] S.F.Mekhamer, M.E.El-Hawary, S.A.Soliman, et.al, "New Heuristic Strategies for Reactive Power Compensation of Radial Distribution Feeders*", IEEE Transaction on Power System, Vol.17, No.43, Oct. 2002, pp. 1128-1135.*

[13] Hogan, P.M.; Rettkowski, J.D.; Bala, J.L., Jr."Optimal capacitor placement using branch and bound", *Power Symposium, 2005, Proceedings of the 37th Annual North American, 23-25 Oct. 2005, pp. 84-89.*

[14] G. Boone and H. D. Chiang, "Optimal capacitor placement in distribution systems by genetic algorithm", Electrical *Power & Energy Systems, vol. 15, no. 3, pp. 155–162, 1993.*

[15] S. Sundhararajan and A. Pawa," Optimal Selection of Capacitor for Radial Distribution Systems using a Genetic Algorithm", *IEEE Transaction on Power System, Vol.9, No. 3 Aug. 1994, pp. 1499-1507.*

A threat risk modeling framework for Geospatial Weather Information System (GWIS): a DREAD based study

K. Ram Mohan Rao[#], Durgesh Pant[*]

[#]Geoinformatics Division, Indian Institute of Remote Sensing
Kalidas Road, Dehradun, India
rammohan@iirs.gov.in

[*]Department of Computer Science, Kumaun University
Nainital, India
durgesh.pant@gmail.com

Abstract **— Over the years, the focus has been on protecting network, host, database and standard applications from internal and external threats. The Rapid Application Development (RAD) process makes the web application extremely short and makes it difficult to eliminate the vulnerabilities. Here we study web application risk assessment technique called threat risk modeling to improve the security of the application. We implement our proposed mechanism the application risk assessment using Microsoft's threat risk DREAD model to evaluate the application security risk against vulnerability parameters. The study led to quantifying different levels of risk for Geospatial Weather Information System (GWIS) using DREAD model.**

Keywords **— Rapid Application Development, Risk rating, Security assessment.**

I. Introduction

There has been tremendous success of World Wide Web (WWW). Today most of the applications are developed using web technologies in different areas *viz.,* banking, ecommerce, education, government, entertainment, webmail and training. Many companies are depending on their web sites for the publicity and business and some of the companies came into business like online shopping through the possibilities of WWW only. Many of customers also find convenient to get benefit from these services of web application rather than conventional or manual methods. The technology of web also enormously developed with modern technologies to build more reliable and cost effective web applications. The technology is now in a position to cope up with various issues like interoperability, multiple platforms and to connect with different database technologies.

Despite the importance of web applications with improved technologies, hacking techniques also gained momentum in cashing the vulnerabilities of the applications. Web Application Security Consortium gave report on web hacking statistics [1]. These statistics clearly states that the number is gradually increasing from year to year, even with the added security feature technology in web application development tools.

II. Security Challenges

Web applications are increasingly becoming high value target for attackers. 71% of the reported application vulnerabilities have affected the web technologies such as web servers, application servers and web browsers [2]. In 2007, a survey was conducted by the Cenzic and Executive alliance on the state of web application security level [3]. Some of the interesting key findings are, there is lack of confidence in the current state of web application security. Around 50% of the people are not confident about their application security, although most of them are happy about their application technology. 83% of the CEOs are aware of the web security, but most of them and other senior management are not sure about the financial implications of the unsecured web applications.

The above findings evidently show that, organizations are still not matured enough to take care of the application security issues against the ever growing threats. Therefore, it becomes imperative than ever to assess the web application security concerns. In the past, organization relied more on gateway defenses, Secure Socket Layer (SSL), network and host security to keep the data secured. Unfortunately, majority of the web attacks are application attacks and the mentioned technologies are generally unable to cope up with the security needs against the application attacks [4]. The gateway firewall and antivirus programs though offer protection at network and host level, but not at the application level [5]. Firewall may not detect malicious input sent to a web application. Indeed, firewalls are great at blocking ports, but not complete solution. Some firewall applications examine communications and can provide very advanced indication still. Typical firewall helps to restrict traffic to HTTP, but the HTTP traffic can contain commands that exploit application vulnerabilities. Firewalls are only an integral part of security, but they are not a complete solution [6]. The same holds true for Secure Socket Layer (SSL), which is good at encrypting traffic over the network. However, it does not validate the application's input or protect from a poorly defined port policy.

The Software Unlimited Organization [7] listed the top 10 firewall limitations. Web servers are becoming popular attack targets. Between 1998 and 2000, around 50 new attacks exploit the Microsoft's widely utilized web server Internet Information Server (IIS) and published these reports in the public domain [8]. Of these attacks 55% allowed an intruder to read sensitive information such as ASP source files, configuration files and finally the data records as well. These growing numbers of attacks target the databases which reside behind the web server. By exploiting the vulnerabilities in the web server it is possible to run SQL commands for gaining the access of database server. Hence protecting the web server is becoming huge concern in the web application security domain.

A. *Web application concerns*

Today's client/server technology has progressed beyond the traditional two tiered concept to three-tier architectures. Application architectures have three logical tiers called presentation services, process services, and data services. As with all these technologies, three tier gives the opportunity to reap these benefits, but a number of challenges to implementing three tier architecture exist. This is because of the number of services that need to be managed, and because the tools are still skeletons for the applications. Furthermore, three tier systems are inherently more complicated because of the multiple technologies involved in the design and development of the application. From pure security point of view, lack of security in any one of the technology will result the total system vulnerable.

Web application must be secured in depth, because they are dependent on hardware, the operating system, web server, database, scripting language and application code. So web applications have numerous entry points that can put database at risk. Hackers generally look into the different fundamental areas of application to break the security. The general types of attacks are IP access, port access, and application access. Hackers get the IP address of the server and do the telnet to exploit the server. There are so many tools for extracting the passwords of the logins. Applications are normally configured to listen on a predefined port for incoming requests. These vulnerable ports are also major sources for the attacks on the application. Web applications include the series of web servers, file servers and database servers etc. Each of these servers attracts potential point of entry to break the application security. But there are so many other areas where the application is vulnerable to the attacks. The major challenges associated with the web application are their most critical vulnerabilities that are often the results of insecure information flow, failure of encryption, database vulnerabilities etc [9]. They are inherent in web application codes, and independent of the technologies in which they are deployed [10]. Attacker may exploit these vulnerabilities at anytime. Almost every week, the media reports on new computer crimes, latest attack techniques, application vulnerabilities, system break-ins, malicious code attacks, and ever growing cyber crime threat. Web Application Security Consortium (WASC) has listed the top 10 web application vulnerabilities for the year 2007 out of reported 24 classes of attacks. Application vulnerabilities, network vulnerabilities, viruses, trojans etc. are some of the external threats. But there are many other internal threats other than external threats posed by rogue administrators, bad employees, some casual employees and social engineering. The solution to the web application security is more than technology. It is all about practices, precautions and countermeasures. That is why security is not a path, its destination. Security is about risk management and effective countermeasures [11].

B. *Security assessment*

Traditionally, security assessment has been considered sa sub function of network management, and has been identified as one of the functional areas of the open system interconnection, management framework. As defined in the OSI management framework, security assessment is concerned not with the actual provision and use of encryption or authentication techniques themselves but rather with their management, including reports concerning attempts to breach system security. Two important aspects are identified (i) managing the security environment of a network including detection of security violations and maintaining security audits, and (ii) performing the network management task in a secure way [12]. Sloman et al, 1994 defines security assessment as the support for specification of authorization policy, translation of this policy into information which can be used by security mechanisms to control access, management of key distribution, monitoring and logging of security activities [13]. Meier et al, 2004 defines security assessment involves holistic approach, applying security at three layers: the network layer, host layer, and the application layer [14]. Additionally, applications must be designed and built using secure design and development guidelines following good security principles. Russ et. al., 2007 concludes security assessment is an organizational level process that focuses on the nontechnical security functions within an organization [15]. In the assessment, it examines the security policies, procedures, architectures, and organizational structure that are in place to support the organization. Although there is no hands on testing (such as scans) in an assessment, it is a very hands on process, with the customer working to gain an understanding of critical information, critical systems, and how the organiation wants to foucs the future of security.

Application security is the use of software, hardware and procedural methods to protect applications from external threats. Security measures built into application and sound application security procedures minimize the likelihood of the attack. Security is becoming an increasingly important concern during development as applications are more frequently accessible over networks. As a result, applications are becoming vulnerable to a wide variety of threats. Application security can be enhanced by rigorously by implementing a security framework known as threat modelling. It is the process of defining enterprise assets, identifying what each application does with respect to these assets, creating security profile for each application, identifying and prioritizing potential threats.

III. General Threat Modeling Principles

Threat is a specific scenario or a sequence of actions that exploits a set of vulnerabilities and may cause damage to one or more of the system's assets. Threat modeling is an iterative process that starts in the early phases of analysis, design, coding & testing and continues throughout the application development life cycle. It systematically identifies and rates the threats that are most likely to effect the web application. By identifying and rating the possible threats with detailed understanding of application architecture the appropriate countermeasures can be implemented against all possible threats in a logical order. Fig. 1 shows the threat modeling process, which is an iterative process

Threat modeling is an essential process for securing web application. It allows organizations to determine the correct controls and product effective countermeasures against all vulnerabilities in the application. Fig. 2 shows the interrelation between a threat and assets, vulnerabilities and countermeasure entities. The threat described in the figure may cause damages to any of application assets and even may exploit all possible vulnerabilities in the system. A successful attack exploits all vulnerabilities in the application and may take over the total control of application. It is probably because of weak design principles, weak coding practices, and configuration mistakes of the applications. Well defined countermeasures can be implemented to the application to mitigate attacks as shown in fig. 2.

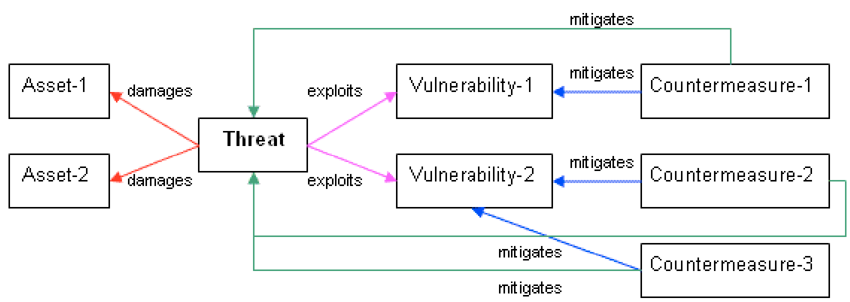


Fig. 2 Interrelation between threat, asset, vulnerability and countermeasure [17]

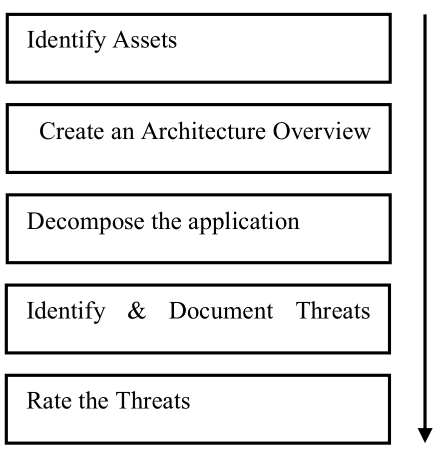


Fig. 1 Threat modeling process [16]

A. *Performing threat risk modelling*

Application development team needs to understand the organization security policy and the overall objectives of the application. Asset is information, capability, an advantage, a feature, a financial or a technical resource that should be defined from any damage, loss or disruption. The damage to an asset may affect the normal functionality of the system as well as the individuals or organizations involved with in the systems. Normally, in the web application technology assets are database, application and web servers.

It is always a difficult task to build a secure web application without knowledge of possible threats. The purpose of the threat modeling is to analyze the application design with solid understanding of application architecture.

The next step is documenting the known treats by keeping an intelligent attacker in mind to shape the application design to meet security objectives, reduce the risks arising during development and deployment. While designing web application, it is essential to design threat risk assessed controls to make application assets more hack resilient at the design time rather than the deployment stage. But it is not possible to document all the possible threats a web application faces as the application development is dynamic process in nature. So the option would be conducting a brain storming session with development people, testers, architecture designers, and professionals etc. to identify the maximum threats at the design time itself. Then the process of documenting the threats in a hierarchical mode that defines core set of attributes to capture for each threat. It is important to rate the threats to prioritize the most frequently occurring possible threats, and which can cause maximum risk to the application. The rating methods depend on different parameters and generally calculated with probability of occurrence and the damage potential that threat could cause.

A. *Threat risk models*

Over the last five years, threat risk modeling became important mitigation development in the web application security environment [18]. Different process models exist for identifying, documenting and rating the threats such as Microsoft Framework, OWASP model, Trike, CVSS, AS 4360 and OCTAVE model [19]. It is up to the security specialist to choose the model according to the suitability of risk assessing method and the technology being used in the application. It is always best practice to adopt one of the risk models to reduce the business risk to the application. This study adopts the basic Microsoft Threat Modeling methodology for implementing threat risk modeling both at design and implementation stages.

IV. Geospatial Weather Information System: A Threat modeling approach

Geospatial Weather Information System (GWIS) is a web based tool for capturing, storing, retrieving and visualization of the weather climatic data. The GWIS contains historical climatic data for nearly hundreds of land stations country wide. The database is provided with both climatic daily and monthly data. Daily data has been nearly for 150 ground stations country wide and covering temperature, rainfall, humidity details. The climatic monthly data has for wide range of land stations around 3000 countrywide. Daily data is being captured from different sources after then arranged in GWIS format for storing in the database. The source for monthly data is Global Historical Climatology Network (GHCN). It is used operationally by National Climatic Data Centre (NCDC) to monitor long-term trends in temperature and precipitation. The mission of GWIS is to integrate the weather related information from different available sources and organize the data in structured GWIS format. The application tool is designed to cater the research needs of various application scientists working on different themes.

Microsoft provides a thereat-modeling methodology for .NET technologies. The process starts from identifying threats, defining architecture overview, decomposing the application, identifying the threats, document the threats and rating the threats. More emphasis has been given to the detailed architecture design describing composition and structure of the application including the sub systems addressing the technologies being used in the web application. As the Microsoft always emphasizes on holistic approach methodology, it again adopts holistic approach in identifying the threats [20].

A. *Identifying threats*

Threats are generally point to network, host and application layers. Identifying network threats is mainly concerned with understanding the network topology, the flow of data packets and the connecting network devices such as router, firewall, and switch. The most frequently occurring network threats are IP Spoofing, Session hijacking, open port policies, open protocols and any weak authenticated network device.Host threats mainly concerned with the security settings of operating system. Possible host vulnerabilities are unpatched servers which can be exploited by viruses, systems with nonessential ports, weak authentication, social engineering etc. Application threat is a big area compared to any other domain of web application. Since the web application includes combination of multiple technologies, there is always a chance for the technology gap between any two. Hence it is always important to evaluate the application vulnerability categories. The major application vulnerability categories are authorization, input validation, cryptography, configuration management, and exception handling. The mentioned areas are normal known threats in the web application environment. But there may be many more number of unknown threats in specific area. However, there are some other approaches to document potential threats using attack trees and attack patterns.

B. *Attack trees and Attack pattern*

As web application often includes the client / server technology with dynamic process of application development, it is very difficult to document all the possible threats. Attack Trees and Attack Patterns are special tools that most of security professionals use for identifying potential threats in the application. They refine information about the attacks by identifying the compromise of enterprise security or survivability as the root of the tree. Each tree represents an event that could significantly harm the asset. Each path through an attack tree represents a unique attack of the asset. Typically threat tree imparts lot more information in shorter time for the reader but takes longer to construct, and attack pattern is much easier way to write but takes longer for the impact of the threats to become obvious. Attack trees provide a formal way of describing the security of systems, based on varying attacks. It represents attacks against a system in a tree structure, with the goal as the root node and different ways of achieving that goal as leaf nodes. Fig. 3 and 4 represents attack tree and attack pattern of GWIS respectively. Attack trees are represented in a tree structure by decomposing a node of an attack tree either as,

- a set of attack sub-goals, all of which must be achieved for the attack to succeed, that are represented as an AND-decomposition.
- a set of attack sub-goals, any one of which must be achieved for the attack to succeed, they are represented as an OR-decomposition.

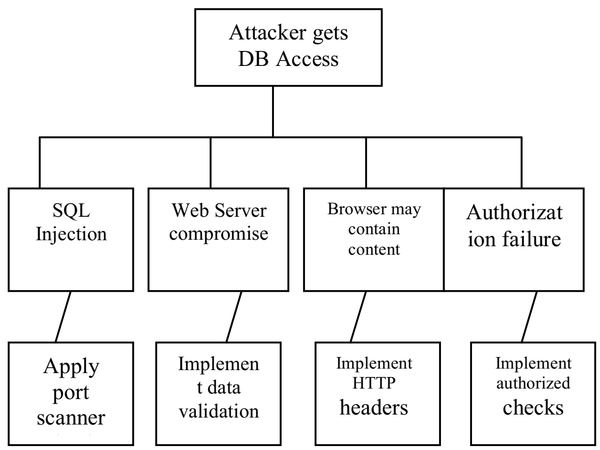


Fig. 3 Attack tree representation of GWIS

Attack patterns are generic representations of commonly occurring attacks that can occur in a variety of contexts. The pattern defines the goal of the attack as well as the conditions that must exist for the attack to occur, the steps that are required to perform the attack, and the results of the attack [21].

Threat #1:The attacker will learn the structure of the SQL query, and then use this knowledge to thwart the query by injecting data that changes the query syntax into performing differently than indented.

1.1 SQL Vulnerabilities
1.1.1 Block SQL Injection, Blind SQL Injection, Cross Site Scripting, HTTP response splitting etc.

1.1.1.1 By verifying that user input does not contain hazardous characters, it is possible to prevent malicious users from causing your application to execute unintended operations, such as launch arbitrary SQL queries, embed javascript code to be executed on the client side, run various operating system commands

Fig. 4 Attack paattern representation of GWIS

B. Document the threat

Documenting the possible known threats of GWIS application gives the great edge to deal with the vulnerabilities. Sometimes it is very difficult to document the unknown threats. But documenting the known threats with the help of third party vulnerability assessment tools will give great knowledge to the developer / administrator to reduce the risks. GWIS application has been scanned thoroughly to perform vulnerability testing to find out the vulnerabilities in the application. For this type of application assessment, single type of vulnerability scanner is not sufficient for scanning the application. So larger sites may require multiple vulnerability scanners to support the assessment needs. The reason is the specific tools are effective in some of the areas and may not be good at other functional areas. For this reason, the GWIS application has been scanned with multiple scanners namely AppScan, CENZIC, and Nessus tools. The consolidated list of vulnerabilities observed is shown in Table 1.

TABLE I
VULNERABILITIES BY PATTERNS

Vulnerability Patterns	**No.of Instances**
Blind SQL Injection	2
Login page SQL Injection	2
Unencrypted login request	1
Application Error	1
Inadequate account lockout	1
Permanent cookie contains sensitive session information	1
Session information not updated	1
Unencrypted password parameter	3
Unencrypted viewstate parameter	7

The vulnerabilities are documented in the threat list as per the Microsoft threat template. Threat list generates the application threat document with the details of threat target, attack techniques, risk and possible countermeasures that are required to address the threat.

C. Rating the Risk

The rating process measures the probability of threat against the damage that could result to an application. This process will generate the list of priority with the overall rating of threats. This allows addressing the most risk generated threat first on priority with proper countermeasures to mitigate the risk. The risk can be calculated from a simple formula [22].

Risk = Probability x Damage potential

Where risk posed by a particular threat is equal to probability of threat occurring multiplied by the damage potential. With this formula risk can be categorized into high, medium and low by calculating in the scale of 1-10. Most of the security professionals may not agree up on the simple rating system in calculating the risk of the application because of the equal distribution of the assets. To resolve this issue, Microsoft came up with a modeling formula called DREAD which is used to calculate risk. On the basis of these parameters, values can be calculated for the given threats and then can be categorized as high risk, medium risk and low risk

D. Rating risk with DREAD approach

DREAD methodology is used to calculate the risk. For each threat the risk rating is calculated by assessing damage potential, reproducibility of attack, exploitability of hte vulnerability, discoverbility of vulnerability and finally total risk points of the application.

D: Damage potential – The loss if the vulnerability is exploited
R: Reproducibility - How easy is it to reproduce the attack
E: Exploitability - How easy to attack the assets
A: Affected users - Average affected users in enterprise
D: Discoverability – How easy to find out the vulnerabilities
T: Total - Total calculated risk points

The threat is rated with high value, if it poses significant risk to the application and needs to be addressed immediately. Table 2 shows the risk rating value of GWIS application using DREAD approach.The scoring system does not consider more than one vulnerability, if the application has more than one number of similar types of vulnerability. For example, GWIS consists of two instances of blind SQL injections and seven unencrypted view state parameters. But finally the scoring has given only for one blind SQL injection, and one unencrypted view state parameter, as the type of vulnerability is same. But when particular type of vulnerability is addressed, total number of instances is taken care. This is because of the reason that, each vulnerability provides equal chance of opportunity for exploiting the application. Once the risk rating is obtained, the threat is documented and with full information of threat target, risk rating, attack technique and necessary countermeasure as shown in Table 3. This template is quite useful for the administrators and application designers for understanding the risk they are dealing with.

V. Experimental results

The GWIS application has been scanned thoroughly for the vulnerabilities across the presentation, business, and database layers of GWIS. Nine vulnerability patterns are found including total 20 instances.

The DREAD scores are calculated against each vulnerability of the application, and the final scores are derived as per the risk catagories. In order to experiment with DREAD model, the study has been chosen the GWIS application to implement security assessment. During the assessment phase, the application flaws are completely assessed with variety of tools for finding out vulnerabilities of the application. The found vulnerabilities are billed with DREAD factors. Fig. 5 shows the DREAD severity gauze of the GWIS. The exploitability factor is maximum for the application, which shows that the vulnerabilities present in the applicaiton are easy to exploit. The damanage potential also is more if the vulnerability is exploited by the attacker. Hence fromthe business point of view the risk is medium. Affected user of hte applicaion, discoverability of the applicaiton is medium when applicaiton is explited. But the reproducibility of hte attack is very less for GWIS. So from the technical point of view the risk is less. Now these DREAD scores are combined together to get final severity risk rating for the GWIS.

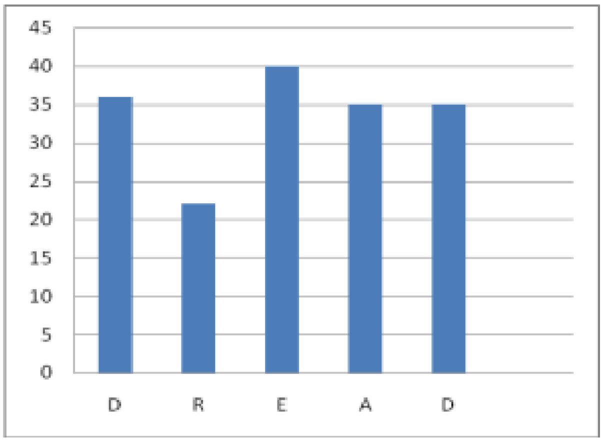


Fig. 5 DREAD severtity gauze

Threat	D	R	E	A	D	T	Average	Rating
Blind SQL Injection	9	6	8	9	6	38	7.60	High
Login page SQL Injection	9	6	8	9	6	38	7.60	High
Unencrypted login request	6	4	6	5	5	26	5.2	Medium
Application Error	2	1	3	2	3	11	2.2	Low
Inadequate account lockout	2	1	3	2	3	11	2.2	Low
Permanent cookie contains sesnsitive session information	2	1	3	2	3	11	2.2	Low
Session information not updated	2	1	3	2	3	11	2.2	Low
Unencrypted password parameter	2	1	3	2	3	11	2.2	Low
Unencrypted viewstate parameter	2	1	3	2	3	11	2.2	Low

TABLE III
THREAT RISK DOCUMENTATION TEMPLATE OF GWIS

Threat Description	***Blind SQL Injection***	***Login page SQL Injection***	***Unencrypted login request***	***Application Error***	***Inadequate account lockout***	***Permanent cookie contains sensitive session information***	***Session information not updated***	***Unencrypted password parameter***	***Unencrypted viewstate parameter***
Threat Target	Data access component	Data access component	Data access component	Application and Data access component	Application and Data access component	Application	Application	Application	Application
Risk	High	High	Medium	Medium	Medium	Medium	Medium	Medium	Low
Attack technique	The attacker will learn the structure of the SQL query, and then use this knowledge to thwart the query by injecting data that changes the query syntax.	In this case an attacker will inject malicious data, which when incorporate into an SQL query, changes the original syntax .	Information sent to server as a clear text, may be stolen and used later for identify theft user impersonation.	The attacker can gain useful information from the application's response to the request.	Attacker gains the access to the application by sending large number of possible user account by hit and trail method.	During the application test, sensitive information such as user credential or session information are stored in a permanent cookie on client computer.	Session fixation is attack technique that forces user session ID to an explicit value.	Input parameter of the type " password" is sent unencrypted to the server.	The ASP contains property called view state an is sent to client and back as hidden variables.
Countermeasures	By verifying that user input does not contain hazardous characters,	Do not use the special character	Make sure that sensitive information such as username password , etc. is always sent encrypted to the server.	Check incoming request for the presence of all expected parameters and values.	Fix the number of login accounts to be attempted. Make sure that if number of logi–n account exceeds, the account is locked .	Make sure that sensitive information such as user credentials session tokens will always be stored in a non-permanent cookies.	Always generate a new session to a user by strong userid/password authentication. Prevent user ability to manipulate session id.	Make sure that sensitive information such as username password , data id, lat, long, sid, location id etc. is always sent encypted to the server.	If a property is not persisted in view state, it's a good practice to return its default value on post back.

As shown in Table 3, probability of occurring the threat in GWIS is MEDIUM, and the damage potential is MEDIUM and hence severity level is medium. SQL injection attracts high sever scores, unencrypted login request carries medium severity scores. Rest of the vulnerabilities are of low sever levels. So from pure business point of view, the risk factor is LOW. But on the whole, probability and damage potential levels of GWIS are MEDIUM and MEDIUM respectively. Therefore the overall severity of the risk is MEDIUM. To minimize the risk levels of the GWIS, it is crucial to fix the most sever risk generating vulnerabilities first such as, blind SQL injection and login page SQL injection vulnerabilities in the GWIS. Similarly the other vulnerabilities also should be fixed to further reduce the risk of GWIS.

VI .CONCLUSIONS

Web based application should be addressed by threat modeling process to identify the potential threats, attacks, vulnerabilities and countermeasures. It is basically a software engineering approach to see the application is meeting the company's security objective and to mitigate the risk at maximum level. It helps to identify the vulnerabilities in the application context. The paper has discussed Microsoft's framework DREAD approach to evaluate the risk of GWIS, and the remediation levels for the vulnerabilities. The output of threat modeling process is standard document of the security aspects about the architecture of application and list of rated threats. This document helps as reference to the designers, developers and testers to make secure design choices, writing code to mitigate the risks and to write the test cases against the vulnerable areas identified in the document.

REFERENCES

[1] WASC, 2008, Web Application SecurityConsortium, Web Hacking Statistics. Accessed from http://www.webappsec.org/projects/whid/statistics.shtml. Accessed on 14.06.2007.

[2] K Mandeep, 2008, Cenzic Application Security Trends Report - Q4, 2007, Cenzic Inc. Whitepaper. Accessed from http://www.Cenzic.com Accessed on 12.02.2008.

[3] Cenzic, 2008, The voice of IT leadership on Web security, Findings of survey conducted by CENZIC and Executive alliance, October 2007. Whitepaper. Accessed from http://www.Cenzic.com Accessed on 12.07.2006.

[4] IBM, 2008, Web application security management, "Understanding the web application security", Whitepaper. Accessed from http://www.ibm.com Accessed on 12.07.2006.

[5] M Curphey, D Endler, W Hau, S Taylor, T Smith, A Russel , M McKenna, R Parke, K McLaughlin, N Tranter, A Klien, D Groves, By-Gad, S Huseby, M Eizner, R Mcnamara, 2002, A guide to building secure web applications, The open security web application project, V.1.1.1. Whitepaper, Available from http://www.first.org/cvss/cvss-guide.html Accessed on 12.04.2007.

[6] J.D Meier, A Mackman, S Vasireddy, M Dunner, S Escamilla, Murukan A, , Improving web application security : Threats and Countermeasures. Microsoft Corporation, 2003, pp.3,.

[7] SoftwareUnlimited, 2005, Firewall Limitations, Irvine, CA. Whitepaper. Accessed fromhttp://www.softwareunlimited.com/securityfirewalltop10.htm. Accessed on 06.20.2005.

[8] 2001, "Understanding the IIS vulnerabilities: fix them, SANS Institute,Whitepaper.Accessed from http://www.sans.org/training/category.php?c=SEC, Accessed on 06.20.2005.

[9] F Ricca and P Tonella, 2000, Web site analysis: structure and evolution, in:roceedings of the IEEE international Conference on Software maintenance, San Jose, California.

[10] D Scott and R Sharp, Developing secure web applications, IEEE nternet Computing, 2002, vol.6.

[11] J.D Meier, A Mackman, S Vasireddy, M Dunner, S Escamilla, A Murukan, Improving web application security : Threats and Countermeasures. Microsoft Corporation, 2003,pp.3.

[12] A Langsford, , *OSI Management Model and Standards*. Chapter 4 in Network and Distributed Systems Management (Sloman, 1994ed), 1994, pp. 69-93.

[13] M. S Sloman, (1994b). *Policy Driven Management for Distributed Systems*. Journal of Network and Systems Management, vol. 2(4), pp. 333-360, December 1994.

[14] J.D Meier, , A Mackman, S Vasireddy, M Dunner, S Escamilla, A Murukan, 2003, Improving web application security : Threats and Countermeasures. Microsoft Corporation, pp.4.

[15] R Russ, D Ted, Greg Miles, Ed Fuller Greg, Security Assessment: Case studies for implementing the NSA IAM, 2007Syngress, Syngress Media, Inc,

[16] G Ygor, 2005, Practical Threat Analysis for the Software Industry, PTA Whitepaper.Available from http://www.securitydocs.com/library/2848.

[17] J.D Meier, A Mackman, S Vasireddy, M Dunner, S Escamilla, A Murukan, Improving web application security: Threats and Countermeasures. Microsoft Corporation, pp 45-49.

[18] A Wiseman, A Adrew, V.D Stock, C Mark, S Ray, 2003,.OWASP : Open Web Application Security Project, A Guide to build secure Web Applications and Web Services. Available from http://www.owasp.org. Accessed on 22.07.2006.

[19] K.Ram Mohan Rao, D. Pant,2006, A Multi-model approach to threat risk modeling of web based application, ICT 2007 Conference on "Communication and computational techniques: Current and future trends" organized at Dehradun February 10-11, 2007.

[20] J.D Meier, A Mackman, S Vasireddy, M Dunner, S Escamilla, A Murukan, Improving web application security : Threats and Countermeasures. Microsoft Corporation, pp.6-7.

[21] S Andrew, 2003, Using Vulnerability Assessment Tools to develop and Octave risk profile. Available from http://www.giac.org. Accessed on 23.08.2007.

[22] J.D Meier, A Mackman, S Vasireddy, M Dunner, S Escamilla, A Murukan, Improving web application security: Threats and Countermeasures. Microsoft Corporation, pp.63.

A Council-based Distributed Key Management Scheme for MANETs

Abdelmajid HAJAMI[#], Mohammed ELKOUTBI[#]

[#]SI2M, ENSIAS, Université Mohammed V Souissi
B.P. 715, ENSIAS Rabat Morocco

abdelmajid_hajami@yahoo.fr

elkoutbi@ensias.ma

***Abstract*—Mobile ad hoc networks (MANETs) have been proposed as an extremely flexible technology for establishing wireless communications. In comparison with fixed networks, some new security issues have arisen with the introduction of MANETs. Secure routing, in particular, is an important and complicated issue. Clustering is commonly used in order to limit the amount of secure routing information. In this work, we propose an enhanced solution for ad hoc key management based on a cauterized architecture. This solution uses clusters as a framework to manage cryptographic keys in a distributed way. This paper sheds light on the key management algorithm for the OLSR protocol standard. Our algorithm takes into account the node mobility and engenders major improvements regarding the number of elected cluster heads to create a PKI council. Our objective is to distribute the certification authority functions for a reduced and less mobile cluster heads that will serve for keys exchange.**

***Keywords*- Key Management; MANET; Clustering.**

I. INTRODUCTION *(HEADING 1)*

In Mobile Ad hoc Networks, devices may have different configurations, and must cooperate to ensure the existence of such networks. MANET devices are free to move in the network, re-enter and leave at will, which shows the spontaneous nature of this type of networks. In addition, these networks do not support the existence of any supervisory or management authority, which provides equipments the same roles in the functioning of the network.

To ensure communication between network devices, MANETs use the radio link. This allows a malicious node to infiltrate easily to disrupt the network. To prevent such behavior, a cryptographic authentification system should be established. However, the authentification system should include a trusted entity that will manage the cryptographic keys.

Effective management of keys, or digital certificates holding the keys, is one of the key factors for the successful wide-spread deployment of cryptographic keys. PKI (Public Key Infrastructure), an infrastructure for managing digital certificates, was introduced for this purpose. The most important component of PKI is the CA (Certificate Authority), the trusted entity in the system that vouches for the validity of digital certificates. The success of PKI depends on the availability of the CA to the nodes in the network since a node must correspond with the CA to get a certificate, check the status of another node's certificate, acquire another node's digital certificate, and so on.

However, connectivity, which was assumed to be good in previous wired PKI solutions, is no longer stable in ad hoc networks. Unfortunately, maintaining connectivity is one of the main challenges in ad hoc networks, since the inherent infrastructurelessness of ad hoc networks makes it hard to guarantee any kind of connectivity. Another serious problem inherent in ad hoc networks is the physical vulnerability of the nodes themselves. Considering that most ad hoc networks will be deployed with mobile nodes, the possibility of the nodes being captured or compromised is higher in wired networks than in those with stationary hosts. With an infrastructure-based solution, mobile nodes may store all sensitive information in the infrastructure and maintain minimal information in the device. Since there is no stable entity in an ad hoc network, the vulnerability of its nodes is even higher.

Our proposed solutions to provide PKI for ad hoc networks deal with the physical vulnerability of the nodes by employing the distribution of CA functionality across multiple nodes and using cryptography threshold. This approach also increases the availability of the CA.

In this work, we will present a solution for managing cryptographic keys based on a clustered architecture for securing the OLSR routing protocol.

Our solution describes how to build key- management infrastructure on a clustered architecture in which a set of nodes in the network are selected using a specific criterion to represent the other nodes in the network [1][2]. These elected nodes, which are cluster-heads of the network, will form what we call the council of the PKI.

This paper is organized as follows: in Part II, we'll present an overview of the OLSR standard protocol. Part III will present an overview of the key management in ad hoc networks.

In Part IV we'll give an overview of the clustering solution that we have adopted. In Part V, we'll discuss in more detail our key management proposal in which we will show the results obtained from the performed simulations. Finally, in Part VI of this paper, we would put in test the robustness of our PKI solution by applying two types of attacks on our proposed architecture.

II. THE OLSR PROTOCOL

The optimized link state routing (OLSR) protocol [3] is a proactive routing protocol that employs an efficient link state packet forwarding a mechanism called multipoint relaying. Optimizations are done in two ways: by reducing the size of the control packets and also by reducing the number of links that are used for forwarding the link state packages. The reduction in the size of link state packets is made by declaring only a subset of the links in the link state updates. The subset neighbors that are designated for link state updates are assigned the responsibility of packet forwarding are called multipoint relays. The optimization by the use of multipoint relaying facilitates periodic link state updates. The link state update mechanism does not generate any other control packet when a link breaks or when a link is newly added. The link state update optimization achieves higher efficiency when operating in highly dense networks. The set consisting of nodes that are multipoint relays is referred to as MPRset. Each given node in the network pinpoints an MPRset that processes and forwards every link state packet that this node originates. Each node maintains a subset of neighbors called MPR selectors, which is nothing than the set of neighbors that have selected the node as a multipoint relay. A node forwards packets that are received from nodes belonging to its MPRSelector set. The members of both MPRset and MPRSelectors keep changing over time. The members of the MPRset of a node are selected in such a manner that every node in the node's two-hop neighborhood has a bidirectional link with the node.

The selection of nodes that constitute the MPRset significantly affects the performance of OLSR. In order to decide on the membership of the nodes in the MPRset, a node periodically sends Hello messages that contain the list of neighbors with which the node has a bidirectional link. The nodes that receive this Hello packet update their own two-hop topology table.

The selection of multipoint relays is also indicated in the Hello packet. A data structure called neighbor table is used to store the list of neighbors, the two-hop neighbors, and the status of neighbor nodes. The neighbor nodes can be in one of the three possible link status states, that is, unidirectional, bidirectional, and multipoint relay.

III. OVERVIEWING THE KEY MANAGEMENT IN MANETS

The aim of this section is to show some solutions for key management in ad hoc networks. The major problem in providing security services in such infrastructure with few networks is how to manage the cryptographic keys that are needed. In order to design practical and sufficient key management systems it is necessary to understand the characteristics of ad hoc networks and why traditional key management systems cannot be used [4].

A. Key Management

As in any distributed system, in ad hoc networks the security is based on the use of a proper key management system. As ad hoc networks significantly vary from each other in many aspects, an environment-specific and efficient key management system is needed.

The security in networking depends, in many cases, on proper key management. Key management consists of various services, each of which is vital for the security of the networking systems. The services must account for these issues: Trust model, Cryptosystems, Key creation, Key storage and Key distribution [5].

The key management service must ensure that the generated keys are securely distributed to their owners. Any key that has to be kept secret must be distributed so that confidentiality, authenticity and integrity are not violated. For instance whenever symmetric keys are applied, both or all of the parties involved must receive the key securely. In public-key cryptography the key distribution mechanism must guarantee that private keys are delivered only to authorized parties. The distribution of public keys need not preserve confidentiality, but the integrity and authenticity of the keys must still be ensured. We showed several solutions for key management in ad hoc networks.

1) Partially Distributed Certificate Authority

This solution proposed by Zhou and Hass [6] uses a (k, n) threshold scheme to distribute the services of the certificate authority to a set of specialized server nodes. Each of these nodes is capable of generating a partial certificate using their share of the certificate signing key skCA, but only by combining k such partial certificates can a valid certificate be obtained. The solution is suitable for planned, long-term ad hoc networks. Since it is based on public key encryption it requires that all the nodes are capable of performing the necessary computations. Finally it is assumed that subsets of the nodes are willing or able to take on the specialized server role.

This solution requires that a server and organizational/administrative infrastructure is available and therefore is only applicable to a subset of ad hoc network applications.

Viewed from a functional standpoint the solution has a number of faults or weaknesses of which the lack of a certificate revocation mechanism is the most critical. Any certificate-based solution should, considering the risk of compromise in ad hoc networks, provide such a mechanism.

Also the solution requires that the server nodes store all of the issued certificates. This requires a synchronization mechanism that propagates any new certificates to all the servers. It must also handle the case when the network has been segmented and later re-joined.

2) Fully Distributed Certificate Authority

This solution is first described by Luo and Lu in [7]. It uses a (k, n) threshold scheme to distribute an RSA signing certificate key to all nodes in the network. It also uses verifiable and proactive secret sharing mechanisms to protect against denial of service attacks and compromise of the signing certificate key. This solution is aimed towards planned, long-term ad hoc networks with nodes capable of public key encryption. However, since the service is distributed among all the nodes when they join the network, there is no need in electing or choosing any specialized server nodes.

Similar to the partially distributed Certificate Authority (CA) this solution requires an organizational/administrative infrastructure to provide the registration and initialization services. The main benefit of this solution is its availability and that , unlike the other certificate based solution proposed, provides a certificate revocation mechanism.

Since all nodes are part of the CA service, it is sufficient that a requesting node has k one-hop neighbors for the CA service to be available. The amount of the network traffic width is also limited.

The cost of achieving this high availability is a set of rather complex maintenance protocols, e.g. the share initialization and the share update protocols. A larger number of shares is also exposed to compromise since each node has its own share as compared to only the specialized server nodes in the partially distributed solution. The k parameter therefore need to be larger since an attacker may be able to compromise a larger number of shares between each share update. This in turn affects the availability of the service. The solution must also provide for a synchronization mechanism in the case of network segmentations.

The proposed certificate revocation method assumes that each node is capable of monitoring the behavior of all its one-hop neighbors. This assumption, however, may be too strong in certain ad hoc networks.

3) Self Issued Certificates

This solution is proposed by Hubaux [8] and provides a public key management solution similar to PGP (Pretty Good Privacy) in the sense that certificates are issued by the users themselves without the involvement of any certification authority. Unlike the public key based solutions, this one is intended to function in spontaneous ad hoc networks where the nodes do not have any prior relationship. Nevertheless, due to this, it requires an initial phase during which its effectiveness is limited and therefore it is unsuitable for short-term networks. Since it is based on public key encryption it requires that the nodes have sufficient computational capacity.

The main benefit of this solution is that it does not require any form of infrastructure neither routing, server or organizational/administrative. However it lacks a certificate revocation mechanism. Also like PGP it has problems during its initial stages before the number of certificates issued reaches a critical amount. This solution also assumes the PGP terminology being called trusted introducers or even meta-introducers. A trusted introducer is a user that is trusted to introduce other users, i.e. to issue certificates to other users. A meta-introducer is a trusted introducer that is trusted to introduce other trusted introducers. [9]

4) Secure Pebblenets

This solution proposed by Basagni [10] provides a distributed key management system based on symmetric encryption. The solution provides group authentification, message integrity and confidentiality.

This solution is suitable for planned and distributed, long-term ad hoc networks consisting of low performance nodes that are unable to perform public key encryption

This solution based on symmetric cryptography requires an organizational/administrative infrastructure that initializes the network nodes with the shared group identity key and additional other parameters. The main weakness of this solution is that it requires that the nodes maintain a tamper-resistant storage. Such a requirement excludes the use of standard networking devices since these typically don't include any tamper-resistant memory. If the group identity key is compromised then all the network nodes need to be re-initialized with a new group identity key.

Finally since only group authentification is supported this solution is not applicable in applications where the communication is peer-to-peer.

5) Demonstrative Identification

This solution proposed by Balfanz [11] presents a mechanism for trust relationships in local ad hoc networks where the network nodes have no prior relationship with each other.

Examples of such local ad hoc networks could be a group of people at a meeting wishing to setup a temporary network or a PDA wishing to temporarily connect to a printer. Since the solution does not require that the nodes have any prior relationship, it is suitable for spontaneous, localized ad hoc networks. It is unsuitable for distributed ad hoc networks since it requires that the nodes be in a close proximity of each other during the initial bootstrapping. It allows the participating nodes to have diverse capabilities, i.e. some are limited to symmetric encryption while others are capable of public key encryption.

All previous solutions have required either an organizational/administrative infrastructure or some sort of social interaction as in the solution based on self issued certificates. The use of demonstrative identification however allows the formation of a secure ad hoc network in a purely self-configured way. As an example two users need only to point their PDAs towards each other. The PDAs then automatically exchange the authentification information required to secure the following communications.

A possible down-side is that the networking devices must be equipped with some sort of location-limited channel. However since the majority of portable devices, e.g. PDAs and laptops are equipped with an infrared interface this should not be a problem. Also this solution is only applicable for localized ad hoc networks.

IV. THE CLUSTERING SOLUTION

The network can be considered as a set of areas (or clusters). Each cluster is formed around a representative called Cluster Head. Cluster Heads are selected according to a well defined criterion.

A cluster is designated by an identifier that relates to its representative (i.e. its cluster head). Each node in the network carries the cluster identifier to which it belongs.

Our proposal presents a simple, light and quiet solution [1][2]. First, our proposal does not add any new control message and the network is not overloaded or slowed at all. No changes are made to standard control messages. Our solution works transparently with the OLSR standard protocol. Clusters are formed around the nodes with the densest environment; in other words, the node that has the largest number of symmetric neighbors is selected as the cluster head.

In this way, we are sure that the cluster is represented by the node that covers the largest number of nodes in the cluster.

V. KEY MANAGEMENT SCHEME

As in any distributed system, in ad hoc networks the security is based on the use of a proper key management system. As ad hoc networks significantly vary from each other in many aspects, an environment-specific and efficient key management system is needed.

The security in networking depends, in many cases, on proper key management. Key management consists of various services, each of which is vital for the security of the networking systems. The services must provide solutions to be able to answer the following questions: Trust model, Cryptosystems, Key creation, Key storage and Key distribution [12].

VI. THE PROPOSED SOLUTION

As described previously, the approaches presented in the literature tried to solve key management problem in ad hoc networks, but these solutions still carry many limits (administrator availability and congestion, dependence of nodes on the administrator and so on). To solve the problem of key management, three solutions are possible. The first is to distribute the functions of PKI on all nodes in the network. But given the dynamics of the network, it is difficult to ensure that all members be available. The second solution is to designate a fixed set of nodes as permanent members of the PKI; these nodes are free to move in the network area. The final solution is based on a clustered architecture in which the cluster-heads form the members of the PKI as will be described later. In our work, we perform a comparative study between the second and final solution. In this section, we are going to describe the approach that we propose for key management in ad hoc networks, which is based on both the clustering technique and the partially distributed PKI solution which is inspired from Threshold Secret Sharing Scheme.

- (k, n) Threshold Secret Sharing Scheme

In secret sharing scheme, a secret is shared among a group of users called shareholders. The secret is shared in such a way that no single user can deduce the secret from his share alone and in order to construct the secret, one need to combine a sufficient number of shares. Adi Shamir [13] proposed a classical (k,n) secret sharing algorithm based on polynomial interpolation. The scheme describes how a secret S can be divided in to n partial shares (S1,S2,...,Sn) where a minimum of k out of n are partial shares are needed to generate a secret S. The threshold value k is a balancing point between fault tolerance and service availability. Asmuth and Bloom [14], Brickell [15], and Karin-Greene-Hellman [16] have enhanced this work. Also, work has been done in the issues related to verifiable secret sharing [17] and verifiable secret redistribution [18].

A. *Description of the scheme*

In this section, and once clusters are formed and heads are designated, as described in [1][2], we would expose the scheme in which we'd gather the cluster heads services of cluster heads in a single service called Council. Each Council node will have equal functionality and utilize the (k,n) threshold scheme for performing the cluster head functionality. The main function of this Council will be key management. A certificate will be validated by participation of at least k nodes out of n Council member. The key management cluster head function will now be able to work even when more than one (but limited to min {k,n-k+1})cluster head is compromised.

In our scheme, we propose a novel architecture that we call 'Council'- based clusters. The scheme uses a collaborative approach to perform as Council-based clusters throughout the network, making it as extremely efficient as possible. Once the Council- based clusters are formed, each Council member can apply (k,n) threshold scheme in a way that a minimum of k cluster heads out of n need to participate together to perform any CA function. For example, for key distribution functionality, Every Council member (each serving as CA) will serve his cluster members. By having multi-cluster heads, the network will be able to work even when more than one (but limited to min {k,n-k+1}) cluster heads are compromised.

- Key Management Scheme on Council- Based Cluster

Key management is an important aspect of ad hoc network security. To ensure security using public key cryptography scheme, each node carries a public-private key pair and a certificate issued by the CA. As discussed earlier, one of the cluster head functionalities can be to function as the CA. A CA certifies that a public key belongs to a particular entity.

Having a single centralized CA is not suitable for highly vulnerable ad hoc networks. Using our scheme, the process of certification can be distributed among all Council nodes within each cluster. We divided our study into two major parties. In the first part, the council is composed of members designated in advance and do not change during the lifetime of the network. This is what we called fixed membership architecture. And in the second part, council members are formed by the heads of clusters; that is what we called clustered architecture. Council issues a certificate to a member node‘s public key by digitally signing it with the private key of the cluster. In order to validate the certificate of a node, at least k Council members out of the n need to work together and combine their contributions. Since at least k among n contributions are required to validate the node's certificate,the system will work even if more than one, but limited to min (k,n-k+1),Council members are compromised.

- Why Limited to Min (k, n-k+1) Compromised the Cluster Heads

In the above section we have mentioned that the cluster head functionality will be able to work even when more than one but limited to min {k,n-k+1}cluster heads are compromised. Let us discuss why our (k,n)threshold scheme is limited to min {k,n-k+1}. In (k,n)secret sharing scheme, a minimum of k cluster heads out of n need to participate together to perform any cluster head functionality. If k or more cluster heads are compromised they will be able to combine their secret share together to perform any compromised cluster head functionality. Thus the total number of compromised nodes cannot exceed k-1. What is more is that in order to perform cluster head service the operation will require at least k non-compromising cluster heads; the system will not if the number of compromised cluster heads are equal to or greater than n-k+1. In general our (k, n) secret sharing scheme will work for any T compromised cluster heads where 1< T < min {k,n-k+1}. For ex. in (5, 12) secret scheme, the system will not work for 5 or more compromised cluster heads as minimum of 5 compromised cluster heads can participate together to perform any cluster head functionality. The (7,12)scheme will not work if 6 or more cluster heads are compromised, as a minimum of 7 cluster heads are required for making the decision.

- Finding (k, n)

We have also addressed the problem of choosing a suitable (k,n) pair on Council based clusters. Not being uniformly distributed, the whole network makes the choice of (k,n) difficult. We find the value of n in an adaptive fashion depending on the availability in the networks. In short, the number of Council members per cluster will give us the value of n. The threshold value k is a balancing point between fault tolerance and service availability. Let's discuss the special cases of choosing k:

• k =1: The secret is shared by n nodes and anyone of them can get the secret using just 1 share. This scheme is similar to a single cluster head and hence vulnerable to a single point of failure.

• k =n: The secret is shared by n nodes and all these nodes need to participate together with their shares in order to get the secret. This scheme provides maximum security but requires accessibility to all the nodes. For highly secure networks like military applications, we will choose k =n and apply (n,n) threshold secret share concept on Council.

• 1<k <n: We chose such a k in a way that there should be a balance between security and availability.

- Scheme steps

The scheme can be explained by the following steps:
1. startup scenario: when starting the network, at least k nodes among members, must share in face-to-face a physical starting key. This key will serve as a session key that will be changed immediately after the start of the network. In this way, any unwanted intrusion will be rejected. Nodes that create the network for the first time are permanent members of the council of PKI. They have a maximum trust metric, and take care of authenticating other nodes that join the network later.
2. After starting the network, if a node arrives for the first time, it must contact in face-to-face one of the permanent members to have a physical certificate. This certificate contains a session key that will enable the new node to connect to the network.
During network operation, each PKI council member records all delivered certificate and broadcast it to the rest of the council. Each network node that is not part of the council must register all obtained certificates.

If a node leaves and joins the network, or if it changes the cluster due to a move, it must be authenticated by one of the council members (as used architecture: fixed membership architecture or clustered architecture) by presenting its first physical certificate. Based on this certificate, the council member broadcasts a request for certificate validation by the other council members. If the authenticator member receives at least k positive responses among n, the node that wants to authenticate will be accepted and the certificate will be delivered.

B. Performance analysis

To see the pertinence of this approach and to measure the effect that will cause the implementation of our algorithm in an OLSR network, we performed several simulations with a variable number of nodes and different nodes velocity.

We used NS2 [20] as network simulator with the following parameters:

Parameter	value
Simulation area	1000 x 1000
Radio range	250 m
Number of nodes	From 10 to 100 by step of 10
Velocity of nodes	From 0 m/s to 40 m/s by step of 5
Simulation time	300 s

We separated our operations into two phases. In the first phase, we decided to measure the impact of our PKI solution on network performance. Parameters that we evaluated are:

the end to end delay, the average through put, the packet delivery fraction (PDF), normalized routing load generated (NRL), the number of collisions in the network, routing loops in the network, the rate of traffic and the number of non-route in the network. In this phase, we made a comparison for three different architectures: an OLSR network without a PKI, an OLSR network with permanent PKI members and an OLSR network with cluster -based PKI.

In the second phase, we defined a set of performance metrics to evaluate the effectiveness of our key management solution. These parameters are: the delivery delay of a certificate representing the time elapsed since the request to delivery of a certificate, the CDF (Certificate Delivery Fraction) which represents the percentage of certificates issued and finally the response time of the PKI which represents the time elapsed between the start of the network and the delivery of the first certificate. We also measure the influence of the threshold parameters (k and n) in order to observe the behavior of the performance of the PKI.

The first phase results are as follows.

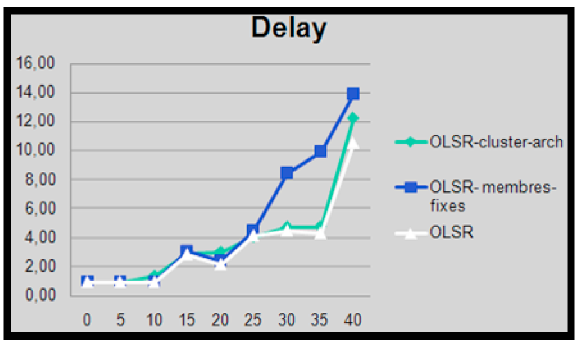


Figure 1- end to end delay in term of node's velocity

Figure 1 shows the end to end delay depending on the speed of the nodes in the network. For low speeds, we notice that the three architectures behave in the same way, while for speeds above 30 m/s, the architecture with permanent members creates an additional delay due to the nodes high mobility. Change of the number of nodes in the network has no effect on the delay parameter.

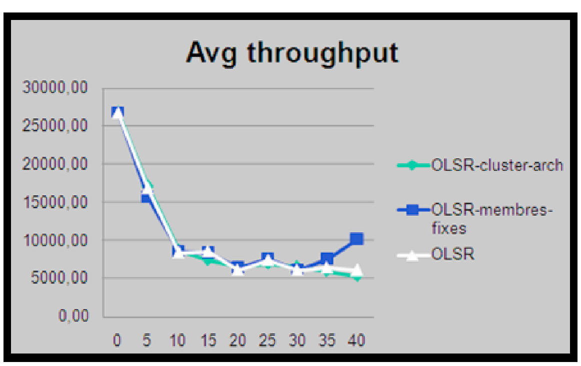


Figure 2- average throughput in term of node's velocity

In Figure 2, we note that in general, the flow remains the same for the three architectures

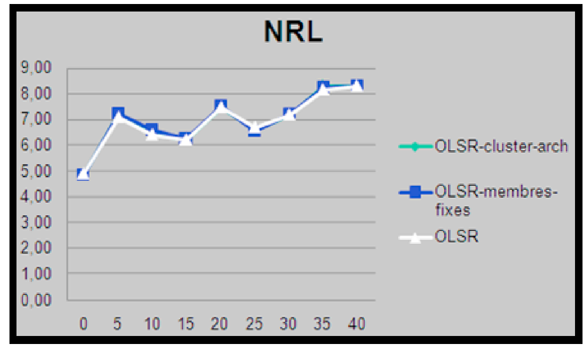


Figure 3- Normalized Routing Load in term of node's velocity

Figure 3 also shows that the NRL is unaffected and it remains the same for the different architectures, both in term of velocity or in term of the number of nodes

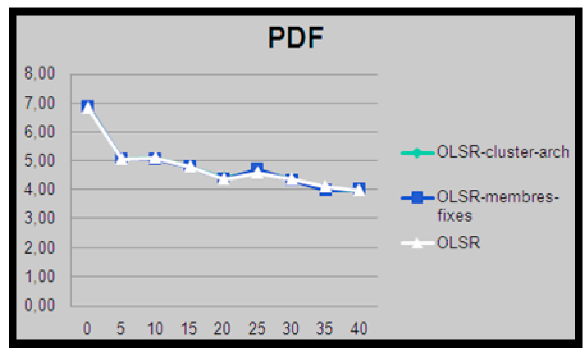


Figure 4- Packet Delivery Fraction in term of node's velocity

In figure 4 also we show that the PDF parameter is unaffected and it remains the same for the different architectures, both in term of velocity or in term of number of nodes

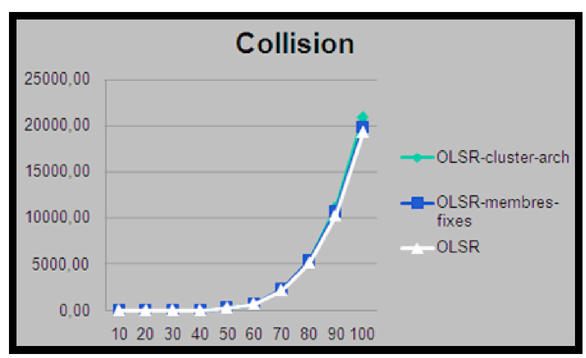


Figure 5- Collision based on the number of nodes

Similarly, in Figure 5 we note that collisions in the network remain the same even if the PKI architecture changes.

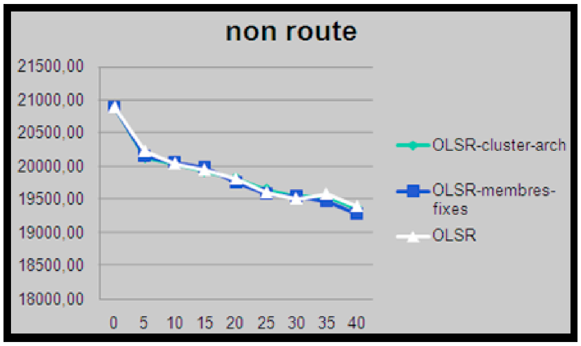


Figure 6- average non-routes in term of node's velocity

Like the other parameters, figure 6 shows that the number of non-routes is always the same.

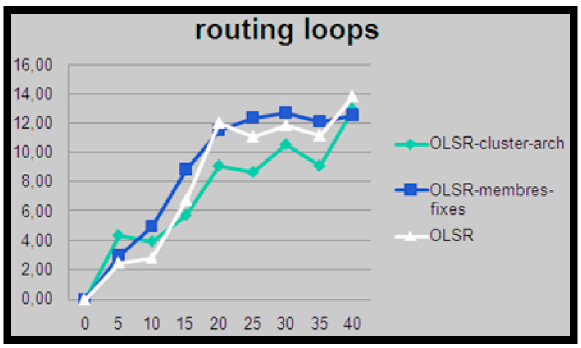


Figure 7- average routing loops in term of node's velocity

In Figure 7 we note a slight decrease in the number of routing loops in the network architecture for clustering

Generally, we can conclude that the operation of key management does not have bad effects on network performance. In the next section, we'll present the measurements of the second phase of this stage of our work. These measures represent some performance indices that we used to assess the reliability of our key management solution.

We measured the influence of the threshold parameters (k and n) in order to observe the behavior of the performance of the PKI. We recall that k represents the threshold contributions to validate a certificate, and n is the maximum of members of the PKI council.

The second phase results are as follows.

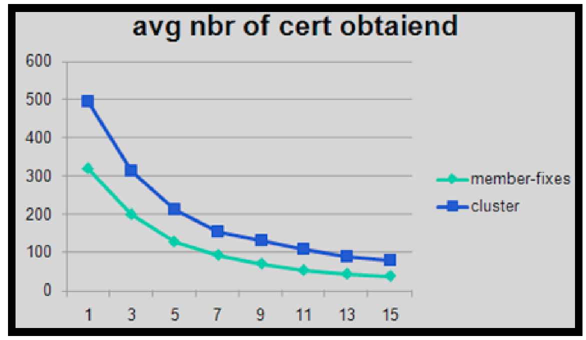


Figure 8- number of certificates obtained in term of k

Figure 8 shows the effect of parameter k on the number of certificates issued during the simulation for both fixed membership architecture, and clusterized architecture. For both architectures, we notice that as the threshold k increases, the number of licenses issued decrease. But the clusters based solution delivers more certificates than the fixed-member architecture.

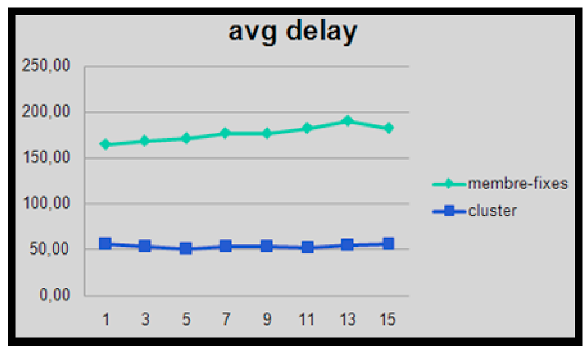


Figure 9- certificates delivery delay in term of k

Figure 9 shows the average delay to deliver a certificate. We note that the parameter k has no influence on the delivery of a certificate. However, the delay of the clustered architecture is 3.5 times lower than that of the fixed-member architecture.

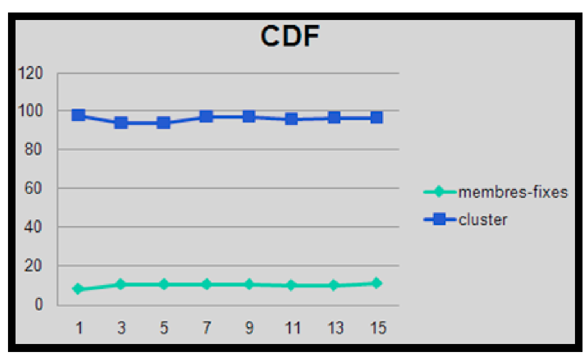


Figure 10- Certificates delivery fraction in term of k.

Figure 10 shows the certificates delivery fraction which represents the percentage of issued certificates by emitted certificates. We note that the parameter k has no influence on the CDF, and the cluster based architecture gives the better result than the fixed-member one.

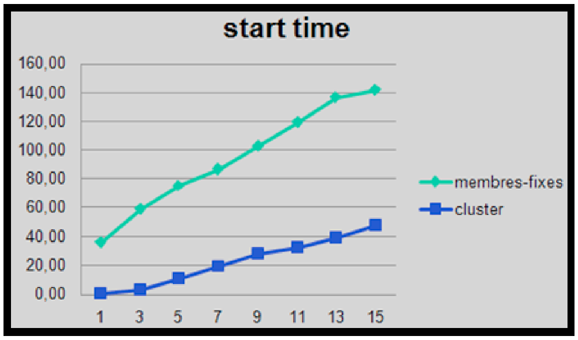


Figure 11- response time in term of k.

In figure 11 we show the PKI response time in term of threshold k. Response time (or start time as shown in the figure) represents the time elapsed between the beginning of the simulation and the delivery of the first certificate. For both architectures, the response time grows when the number k grows. But response time for cluster-based architecture is less than the fixed-member one.

In the next section, we'll present the results of the effect provoked by the parameter n (which represents the number of council members) on the performance of the PKI.

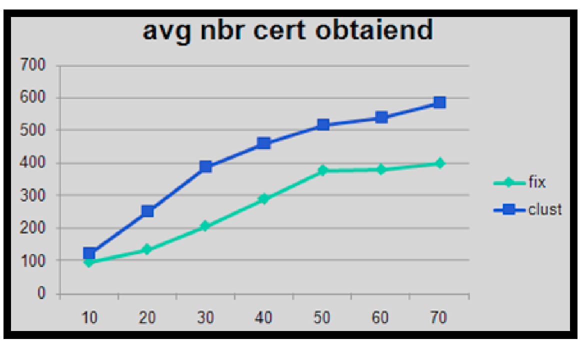


Figure 12- nbr of certificates obtained in term of n.

Figure 12 shows the effect of parameter n on the number of certificates issued during the simulation. Gradually as the number of members of council increases, the number of certificates issued increases.

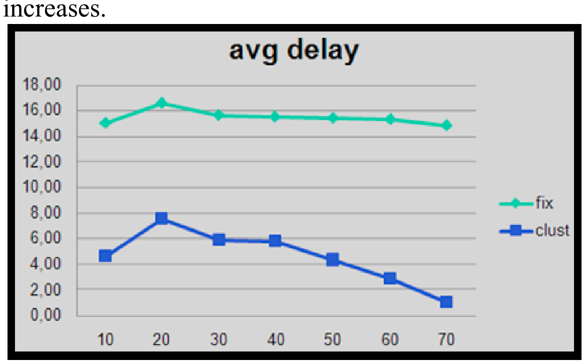


Figure 13- certificates delivery delay in term of n

Figure 13 shows the average delay to deliver a certificate. We note that the parameter n has less influence on the delivery of a certificate. However, the delay of the clustered architecture is around 2.6 times lower than that of the fixed-member architecture.

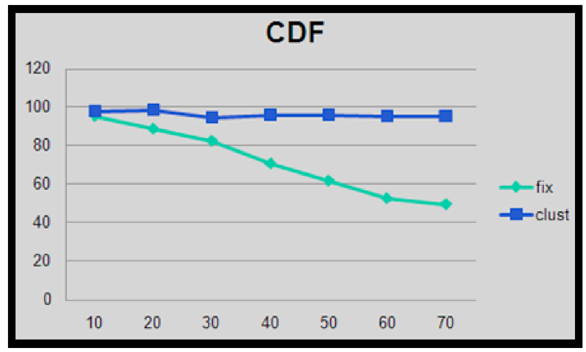


Figure 14- Certificates delivery fraction in term of n.

As shown in figure 14, in term of CDF, the parameter n has no effect on the cluster based architecture when it decreases the CDF in the case of the fixed-member architecture.

VII. Robustness Of The PKI

To measure the robustness of our PKI, we have tested to possible attacks. So we simulate two different types of attacks. The first type consists of a black-hole attack [21], in which the attacker node absorbs the traffic and do not let it pass from one network area to another. This way the network will be decomposed into inaccessible areas. The second type of attack consists of a grey-hole attack [21], in which the attacker modifies the control traffic by false data before to rebroadcast it from one side of the network to another. In this way, the network topology information becomes wrong, causing a loss of data packets.

We perform three simulations. The first simulation concerns an OLSR network without PKI in which attackers make an attack of black-hole that we have appointed "attack 1 OLSR" on the graphs that follow. The second simulation concern an OLSR network without a PKI in which attackers make an attack of gray-hole that we have appointed "attack 2 OLSR" on graphs. Third model of simulation concerns an OLSR network with PKI in which the PKI structure reacts with any type of attack by ignoring the malicious packets.

We measure some performance parameters of the network to observe the behavior of the proposed architecture against the attacks that we simulate. In what follows we present the results we have achieved.

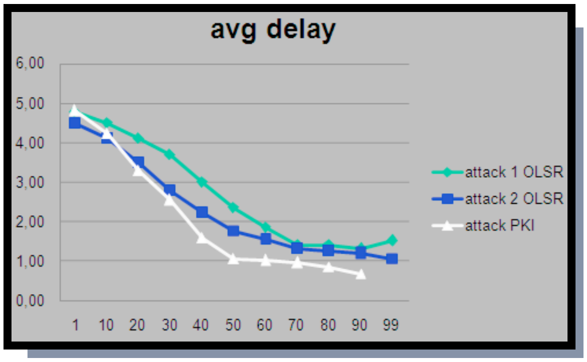


Figure 15- end to end delay in term of percentage of bad nodes.

Figure 15 shows the average end to end delay in the network. We note that the delay is improved in the case of a network with PKI, so it remains high in the absence of PKI.

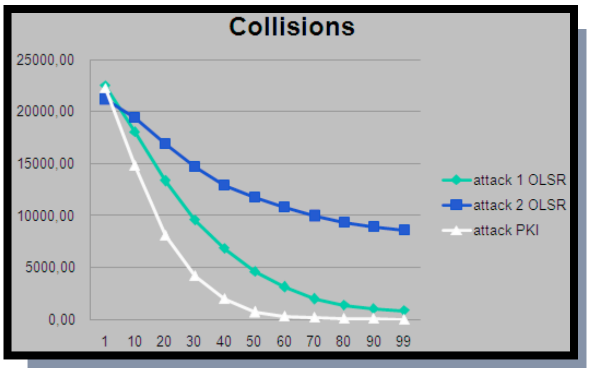


Figure 16 - number of collisions in term of percentage of bad nodes.

In figure 16 we show the number of collisions in the network. We note that the architecture with PKI generates a less amount of collisions than the case without PKI.

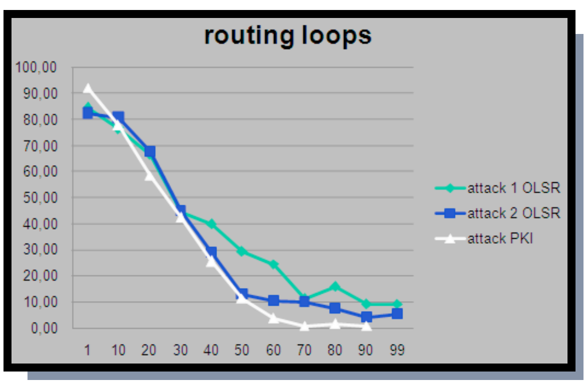


Figure 17- number of routing loops in term of percentage of bad nodes.

In figure 17, we show the number of routing loops in the network. We note that in an OLSR network with our PKI, routing loops are lower than in a network directly exposed to this kind of attacks.

Generally, we conclude that the PKI architecture that we proposed has a good robustness to different types of attacks that we simulate, and allows to optimize network resources in term of delay and bandwidth by eliminating unwanted traffic.

VIII. Conclusion And Perspectives

In a PKI (Public Key Infrastructure), the most important component is the CA (Certificate Authority), which is the trusted entity in the system that vouches for the validity of digital certificates. The success of PKI depends on the availability of the CA to the nodes in the network since a node must correspond with the CA to get a certificate, check the status of another node's certificate, acquire another node's digital certificate, and so on.

However, connectivity, which was assumed to be good in previous wired PKI solutions, is no longer stable in ad hoc networks. Unfortunately, maintaining connectivity is one of the main challenges in ad hoc networks, in view of the fact that the inherent infrastructurelessness of ad hoc networks makes it hard to guarantee any kind of connectivity. In this work, we presented a solution of cryptographic key management for ad hoc networks, based on a clustered architecture. In fact, the difficulty of finding a stable and permanent entity to ensure the function of CA requires distributing this role on all nodes in the network. But the problem of availability of all nodes simultaneously may cause the unavailability of AC services. For this, inspired by the threshold cryptography we have proposed creating a council of PKI that is composed of a subset among all nodes of the network.

Now to choose the council members we have proposed two solutions. The first solution is to designate a set of nodes that will make the council of PKI. These members are chosen randomly, and they remain the same as the network exists. That is what we mentioned in our article by fixed-members architecture.

The second solution is to organize the network in the form of clusters and each cluster will be represented by its cluster-head. All cluster heads of the network form the council of PKI. That is what we mean by Cluster-based Architecture.

We compared these two architectures and we figured out a set of measures that show that the clustered architecture provides a better result and is well suited to the dynamic environment of mobile ad hoc networks.

As perspective to this work, we plan to develop some aspects, focusing on the choice of the threshold parameter values (or k), and the council members number (or n).

References

[1]. A. Hajami, K. Oudidi, M. Elkoutbl, 'A Distributed Key Management Scheme based on Multi hop Clustering Algorithm for MANETs', IJCSNS International Journal of Computer Science and Network Security , Vol. 10 No. 2 pp. 39-49

[2]. A. Hajami, K. Oudidi, M. Elkoutbl, 'An enhanced algorithm for MANET clustering based on Multi hops and Network Density' NOTERE 2010, 31 Mai, 2 Juin 2010 Tozeur, Tunisie

[3]. T. CLAUSEN ET P. JACQUET. *Optimized Link State Routing Protocol (OLSR)*.http://www.ietf.org/rfc/rfc3626.txt, 2003, RFC 3626

[4]. Del Valle Torres Gerardo, Gómez Cárdenas Roberto. Proceedings of ISADS 2005. 2005 International Symposium on Autonomous Decentralized Systems (IEEE Cat. No. 05EX1007)

[5]. Kârpijoki Vesa, Security in Ad Hoc Networks., Telecommunications Software and Multimedia Laboratory 2002

[6]. L. Zhou and Z. J. Haas, .Securing Ad Hoc Networks., IEEE Networks, Volume 13, Issue 6 1999

[7]. H. Luo and S. Lu, .Ubiquitous and Robust Authentication Services for Ad Hoc Wireless Networks., Technical Report 200030, UCLA Computer Science Department 2000

[8]. J-P. Hubaux, L. Buttyán and S. Capkun, .The Quest for Security in Mobile Ad Hoc Networks.,ACM 2001

[9]. S. Garfinkel, PGP: Pretty Good Privacy, O.Reilly & Associates 1995,ISBN 1-56592-098-8

[10]. S. Basagni, K. Herrin, E. Rosti and Danilo Bruschi, .Secure Pebblenets., ACM 2001

[11]. D. Balfanz, D. K. Smetters, P. Stewart and H. Chi Wong, .Talking To Strangers: Authentication in Ad-Hoc Wireless Networks., Internet Society, Conference Proceeding of NDSS Conference 2002

[12]. Kärpijoki Vesa, .Security in Ad Hoc Networks., Telecommunications Software and Multimedia Laboratory 2002

[13]. A.Shamir, “How to Share a secret”, Communication of the ACM, Vol. 22, pp. 612-613, November 1979

[14]. C. Asmuth anf J. Bloom, “A Modular Approach to Key Safeguarding”, IEEE Trans. On Information Theory, IT-29, pp. 208-211, 1983

[15]. E. F Brickell, “Some Ideal Secret Sharing Schemes”, Journal of Combinatorial Mathematics and Combinatorial Computing, No. 6, pp. 105-113, 1989.

[16]. E. D. Karnin, J. W. Greene, and M. E. Hellman, “On Secret Sharing Systems”, IEEE Trans. On Information Theory, IT-29, pp. 35- 41, 1983

[17]. T. P. Pederson, “Non-Interactive and Information-Theoretic Secure Verifiable Secret Sharing “, Lecture Notes in Computer Science, pp. 129-140, 1992

[18]. Y. Desmedt and S. Jajodia, “Redistribution Secret Shares to New Access Structures and its Applications”, Technical Repport ISSE TR-97-01, George Mason University, Fairfax, VA, July, 1997.

[19]. Vivek Shah, “Parallel Cluster Formation for Secured Communication in Wireless Ad hoc Networks”, work submitted as part of requirement for the degree of Master of Science in Computer Science. University of Cincinnati. 2004

[20]. A. UC Berkeley and USC ISI, “The network simulator NS-2”, Part of the VINT project, Available from www.isi.edu/nsnam/ns, 1998.

[21]. Yih-Chun Hu, Adrian Perrig, and David B. Johnson. Ariadne: A secure on-demand routing protocol for ad hoc networks. In Proceedings of the 8th Annual ACM International Conference on Mobile Computing and Networking (MobiCom ’02), September 2002.

[22]. A. Patwardhan, J. Parker, A. Joshi, M. Iorga, and T. Karygiannis. Secure Routing and Intrusion Detection in Ad Hoc Networks. Third IEEE International Conference on Pervasive Computing and Communications, Kauaii Island, Hawaii, March, pages 8–12, 2005.

Improved Spectrogram Analysis for ECG Signal in Emergency Medical Applications

A.K.M Fazlul Haque
Department of Computer Science and Engineering,
Jahangirnagar University,
Dhaka, Bangladesh.
Email: akm_haque@yahoo.com

Md. Hanif Ali
Department of Computer Science and Engineering
Jahangirnagar University,
Dhaka, Bangladesh.
Email: hanif_ju03@yahoo.com

M Adnan Kiber
Department of Applied Physics, Electronics and Communication Engineering, University of Dhaka,
Bangladesh.
Email: ma_kiber@yahoo.com

***Abstract*- This paper presents the spectrogram effect of biomedical signal, especially for ECG. Simulation module developed for the spectrogram implementation. Spectrogram based on ECG signal and power spectral density together with off-line evaluation has been observed. ECG contains very important clinical information about the cardiac activities of heart. The features of small variations in ECG signal with time-varying morphological characteristics needs to be extracted by signal processing method because there are not visible of graphical ECG signal. Small variations of simulated normal and noise corrupted ECG signal have been extracted using spectrogram. The spectrogram found to be more precise over conventional FFT in finding the small abnormalities in ECG signal. These form time-frequency representations for processing time-varying signals. By using the presented method, it is ensure that high resolution time-varying spectrum estimation with no lag error can be produced. Other benefits of the method are the straightforward procedure for evaluating the statistics of the spectrum estimation.**

***Keywords*- Spectrogram, ECG, PSD, Periodogram, Time-varying signal, FFT.**

I. INTRODUCTION

Electrocardiogram (ECG) is the electrical manifestation of the heart muscle activity. Electric impulse originating at sino-atrial node (SA) has an intrinsic rate that is regulated by the sympathetic and parasympathetic branches of Autonomic Nervous system (ANS) [1]. Nerve impulses arriving from the sympathetic branch tend to increase the mean heart rate while impulses from the parasympathetic branch mediated by vagus nerve have the opposite effect. These nerve impulses do not occur with exact regularity as they can be modulated by central and peripheral oscillators, causing variations in beat-to-beat interval which is termed as Heart Rate Variability (HRV) [1] [3] [6].

Bioelectrical signals are typically very small in amplitude (mV) and an amplifier is required to accurately depending on the hardware and software used, the biological amplifier serves to amplify the signal. It is also known that the frequency of heart signals is very low, approximately 5 to 10 Hz.

Spectrogram returns the time-dependent Fourier transform for a sequence, or displays this information as a spectrogram. The time-dependent Fourier transform is the discrete-time Fourier transforms for a sequence computed using a sliding window. This form of the Fourier transform, also known as the short-time Fourier transform (STFT), has numerous applications in speech, sonar, and radar processing. The spectrogram of a sequence is the magnitude of the time-dependent Fourier transform versus time [2] [4].
The estimation procedure was initiated with low pass filtering the ECG data which was originally sampled at 360Hz, and down sampled to 4Hz. The spectrogram (time frequency energy distribution) for signal S(t) was estimated as

$$P(t,f) = \frac{1}{2\pi} \left| \int e^{-j2\pi f\tau} S(\tau)\, h(\tau - t)\, d\tau \right|^2 \quad \ldots(1)$$

where $h(\tau\text{-}t)$ is a window function which slides along S(t). This corresponds to the maximum of P(t,f) in a given frequency range [13]. As the HRV does not vary abruptly, a frequency range of around 0.2 Hz was used for detection of the dominating frequency apart from the harmonics. From the spectrogram the mean frequency for a short interval of every 6 seconds was calculated. A band-pass filter whose center frequency (mean value from spectrogram) was varied according to the varying dominating frequencies was designed. Time-frequency estimation of power spectral density (PSD) is a common step in the analysis of nonstationary signals. The spectrogram is arguably the most popular technique, though the scaleogram and Wigner-Ville distribution are also common [1] [5] [7]. The spectrogram estimates the PSD by applying the modified periodogram to windowed signal segments separated by a fixed interval [2]. The user-specified length of the window controls the trade-off between time and frequency resolution of the image. In this paper, time-varying resolution has been estimated with no lag error statistically. The Matlab Simulink process is used to verify the process. FFT methods have been used in a large number of biomedical applications. There is some works on precise detection of ECG using FFT [14-21]. Karel *et al*. proposed the performance criteria to measure the quality of a wavelet, based on the principle of maximization of variance [14]. Mahmoodabadi *et al*. developed and evaluated an electrocardiogram (ECG) feature extraction system based on the multi-resolution wavelet transform [15]. David *et al*. presented a method to reduce the baseline wandering of an electrocardiogram signal [16]. Shantha *et al*. discussed the design of good wavelet for cardiac signal from the perspective of orthogonal filter banks [19]. Nikolaev and Gotchev proposed a two-stage algorithm for electrocardiographic (EGG) signal denoising with Wiener filtering in the translation-invariant wavelet domain [20].

II. METHODOLOGY

ECG signal is generated by writing a function. This function generates a wave similar to a sine function which representative of a true ECG waveform.

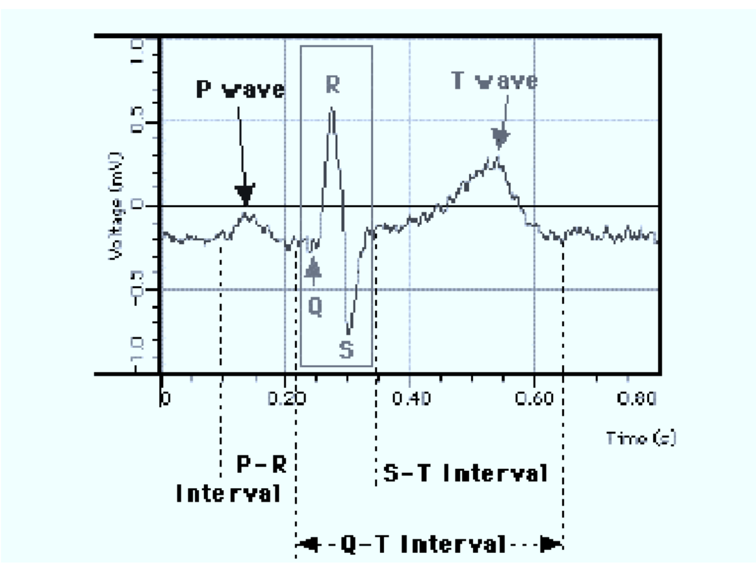


Figure **1.** A general representation of *ECG signal*

An ECG signal is the superposition of action potentials that occur throughout the heart. A typical ECG signal for one heartbeat is shown in Figure 1.

An ECG signal is characterized by the P wave, the QRS complex wave and the T wave, with each wave created by specific heart activities. "The P wave is produced by atrial depolarization, the QRS complex primarily by ventricular depolarization and the T wave by ventricular repolarization" [9] [10]. Some ECG signals also contain a small amplitude U wave following the T wave; U waves are common in slow heart rates but a prominent U wave may reflect a heart abnormality.

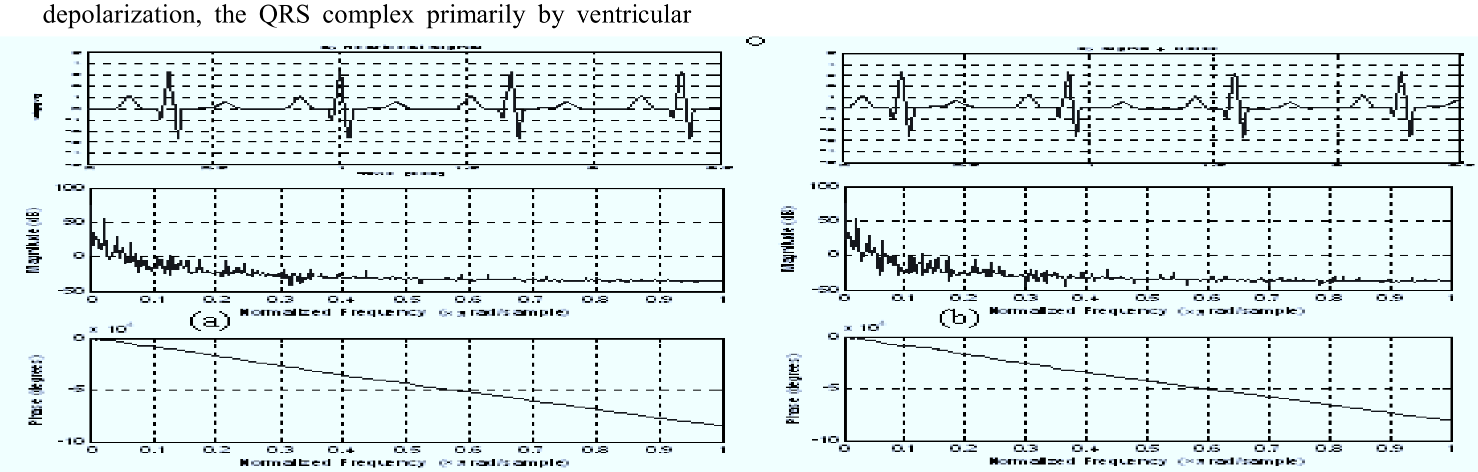
An ECG signal can also be broken down into three main intervals: the P-R interval, the Q-T interval and the S-T interval. The P-R interval is mainly caused by the depolarization of the atrium and slow conductance of the associated impulse to the ventricle by the atrioventricular (AV) node. The Q-T interval is defined by the depolarization and repolarization of the ventricle. The S-T interval corresponds to the "average duration of the plateau regions of individual ventricular cells [11] [12].

III. ANALYSIS AND DISCUSSION

The simulated standard ECG signals as well as the simulated noise corrupted signal have been implemented using FFT and spectrogram for proper feature extraction. From the human body, sudden pain of any parts may occur the continuous sinusoidal signal with very low frequency with approximately 0.5/1 Hz cause the small abnormalities of the cardiac activities of heart. Signals have been generated with different parameters using the following steps:

Step 1: Generation of standard ECG pattern having amplitude of 3.5mV and pulse repetition rate of 75 per minute. This signal is shown in Figure-2(a).
Step 2: Generation of a noisy signal having frequency of 0.5/1 and amplitude of 0.1 mV which is 2.85 percent of the standard ECG signal. This signal is shown in Figure-2(b).
Step 3: Generate the response of the signal using FFT method.
Step 4: Generate the response of the signal using wavelet method.
The algorithm is based on ECG which is represented by workspace block converted the captured file into 2-dimentional data file. Signal is characterized by the frequencies, amplitudes and phases of the sine-wave components. It is applied in every 10 ms. frame using the simulink block system's rules and regulations. From the principle of FFT, it does not have possibilities to detect the little change of the signals which is show in below.

Figure 2. Response using FFT of simulated normal (a) and noise corrupted (b) ECG signals.



From the principle of spectrogram, it has the possibilities to detect the little change of the signals with appropriate power density spectrum. The block diagram of the spectrogram procedure is given in figure 3.

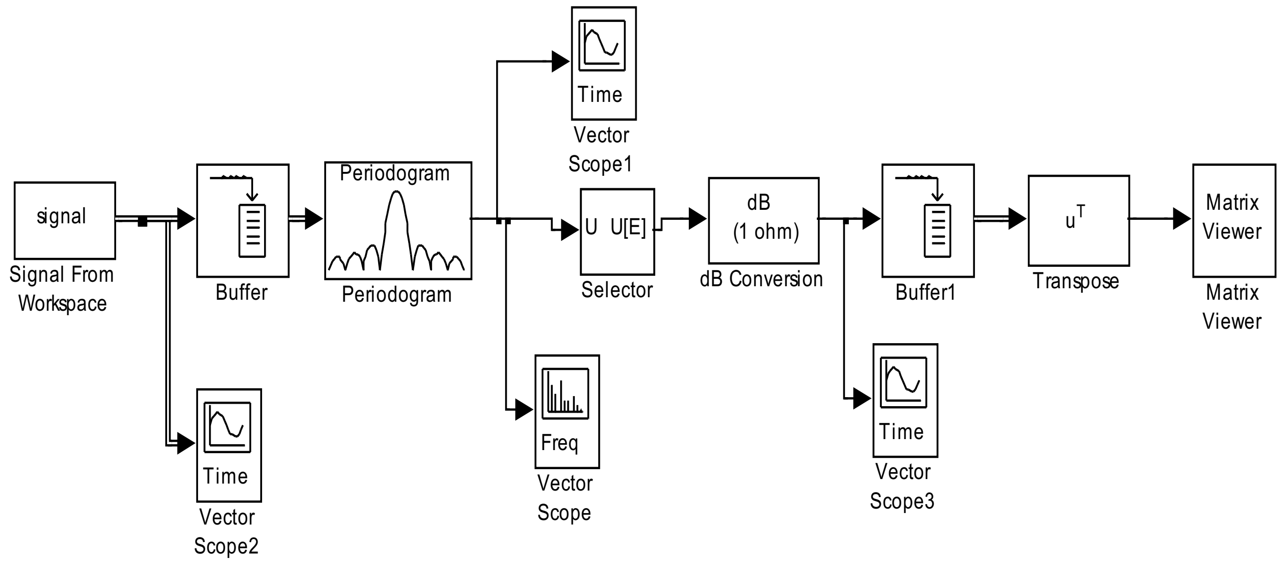


Figure 3. Diagram of proposed simulation method of Spectrogram

The brief discussions of the various block of simulated method of spectrogram have been introduced in the following.

Workspace (Figure 3.)- Simulink allows us to import input signal and initial state data from the MATLAB workspace and export output signal and state data to the MATLAB workspace during simulation. This capability allows us to use standard or custom MATLAB functions to generate a simulated system's input signals and to graph, analyze, or otherwise postprocess the system's outputs STFT reconstruction (Fig. 2) - The frequencies, phases and amplitudes are combined to form a sine-wave representation. The final reconstructed STFT is constructed from the sine waves by a convolution procedure.

To use this format, it selects input in the load from workspace pane and selects the array option from the format list on the data import/export pane. Selecting this option causes simulink to evaluate the expression next to the input check box and use the result as the input to the model. The expression must evaluate to a real (noncomplex) matrix of data type double. The first column of the matrix must be a vector of times in ascending order. The remaining columns specify input values. In particular, each column represents the input for a different import block signal (in sequential order) and each row is the input value for the corresponding time point. Simulink linearly interpolates or extrapolates input values as necessary if the interpolate data option is selected for the corresponding import. The total number of columns of the input matrix must equal n + 1, where n is the total number of signals entering the model's imports.

Vector Scope (Figure 3.) - The vector scope block is a comprehensive display tool similar to a digital oscilloscope. The block can display time-domain, frequency-domain, or user-defined signals. We can use the Vector Scope block to plot consecutive time samples from a frame-based vector, or to plot vectors containing data such as filter coefficients or spectral magnitudes. To compute and plot the periodogram of a signal with a single block, use the Spectrum Scope block. The input to the Vector Scope block can be any real-valued M-by-N matrix, column or row vector, or 1-D (unoriented) vector, where 1-D vectors are treated as column vectors. Regardless of the input frame status, the block treats each column of an M-by-N input as an independent channel of data with M consecutive samples. The block plots each sample of each input channel sequentially across the horizontal axis of the plot conversion to time domain (Fig. 2) - This is the final stage in which the time domain signal frame is computed [13].

Buffer(Figure 3.)- Convert scalar samples to a frame output at a lower sample rate. We can also convert a frame to a smaller or large size with optional overlap [13].

Periodogram(Figure 3.)- Nonparametric spectral estimation using the periodogram method. In this block, the power spectral density is estimated and viewed in both time time-domain and frequency domain.

dB Conversion(Figure 3.)- Converts inputs of Watts or Volts to decibels. Voltage inputs are first converted to power relative to the specified load resistance, where P = (V^2/R). When converting to dBm, the power is scaled to units of miliWatts [13].

Matrix Viewer(Figure 3.)- Compute the matrix transpose. Vector input signals are treated as [Mx1]

matrices. The output is always a matrix. The input and output data types may be any real signed 16-bit fixed point data type [13].

The method tried to produce the spectrogram and power spectral density in a relatively short period of time based on ECG. By using the presented method, high resolution time-varying spectrum has been estimated with no lag error. The straightforward procedures for evaluating the statistics of the spectrum estimation are as follows.

Step 1: Generation of ECG pattern having amplitude of 3.5 mV. Figure 4shows the generation of ECG signal and it has been generated in offline and the time domain data imported from the workspace.

Step 2: power spectrum density of different time varying in time domain has been showed in Fig. 5.
Step 3: Fig. 6 shows the power spectrum density in frequency domain.
Step 4: Fig.7 shows the dB conversion of power spectral density.
Step 5: Different time varying spectrogram of ECG are shown in Fig. 8

Time-frequency evaluation of power spectral density (PSD) is a frequent step in the analysis of nonstationary signals in which the information is extracted (Fig.6). The spectrogram is possibly the most popular technique, though the scaleogram and Wigner-Ville distribution are also common [1]. The spectrogram estimates the PSD by applying the modified periodogram to windowed signal segments separated by a fixed interval [2]. The user-specified length of the window controls the trade-off between time and frequency resolution of the image.

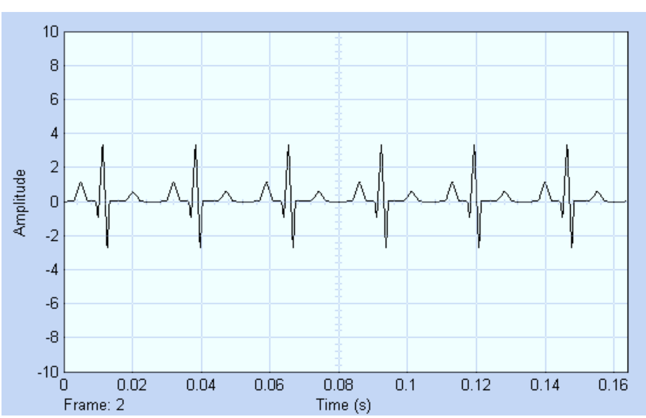


Figure 4. ECG signal generated from workspace

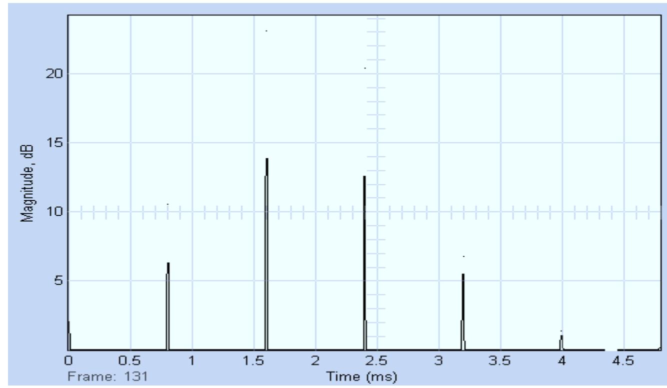


Figure 5. Power Spectral Density Estimation in time-domain.

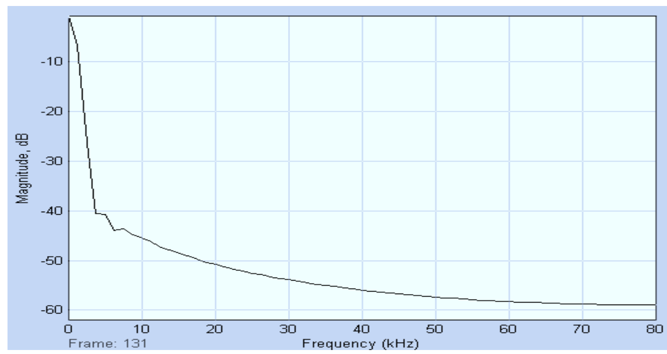


Figure 6. Power Spectral Density Estimation in frequency-domain.

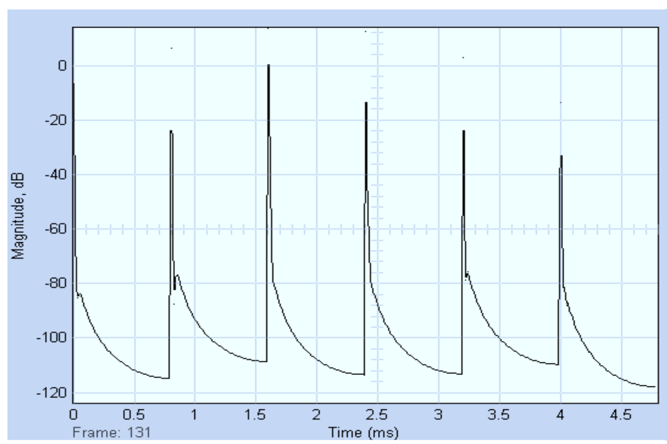


Figure 7. dB conversion of PSD.

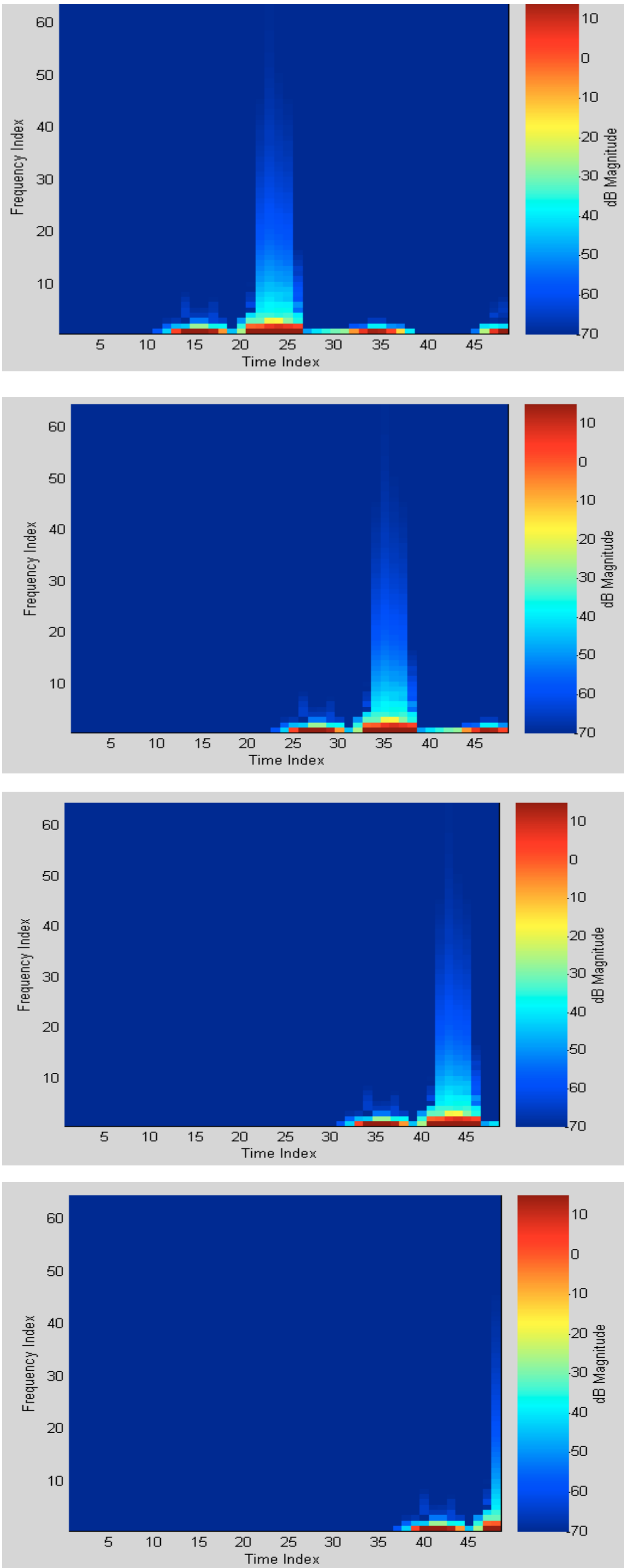


Figure 8. Time - varying spectrogram.

We can have the decision from the above point of view that high resolution time-varying spectrums' lag error have been eliminated in terms of time-varying spectrum density and the peak amplitude is quite visible in spectrogram where the time-domain peak amplitude appears. In the following graph (Fig.9) we can see the time versus lag error has been reduced exponentially.

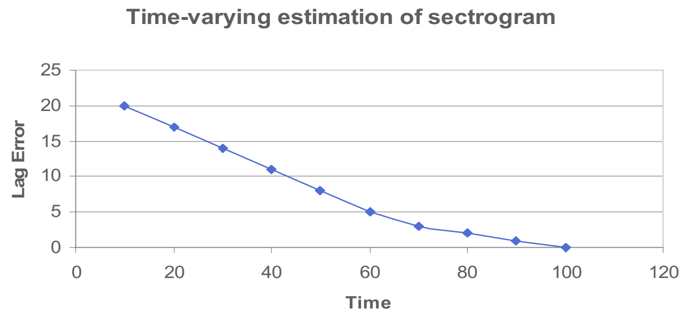


Figure 9. Statistics of Time-varying spectrogram

IV. CONCLUSIONS

In conclusion, the effect of biomedical signal, especially for ECG, has been experimented using the spectrogram implementation. A method of estimating the harmonic power spectral density that combines the power of the fundamental and harmonic components is also described. Spectrogram based on ECG signal and Power spectral density together with off-line evaluation has been observed. The status of the signal is performed by evaluating the statistics of the signal in time-domain and frequency domain. Further the power spectral density is considered by using the characteristic of periodogram. The gain and losses are also considered in dB. It accounts for variations in the power distribution among harmonic frequencies. It achieves better frequency resolution by leveraging the relatively better resolution at the harmonic frequencies. The presented method, high resolution time-varying spectrum has been estimated. Finally, lag error have been eliminated in terms of time-varying spectrum density and the peak amplitude is quite visible in spectrogram where the time-domain peak amplitude appears.

V. REFERENCES

[1] R. Aston, Principles of biomedical instrumentation and measurement. Columbus: Merrill Pub. Co., 1990.[2] Monitoring Patients via Radio Signals. About Wireless Medical Telemetry Service (WMTS).

[3] A.K.M Fazlul Haque, M. Shamim Kaiser, S. K. Aditya, "Background Noise Cancellation Using Adaptive LMS Algorithm in Medical Applications" Journal of Bangladesh Electronic Society, Volume 6, Issue 2, 2007.

[4] Parbhakar S. Naidu, "Sensor Array Signal Processing", New York, CRC Press Inc, 2004.

[5] Phil Mars, J R Chen and Raghu Nambiar, "Learning Algorithms: Theory and Applications in Signal Processing, Control and Communications", New York, CRC press,1998.
[6] John G Webster, "Medical Instrumentation Application and Design", 3rd Edition, Toronto, 2000.
[7] Rajendra Madhav Palnitkar, "A Review of Adaptive Neural Networks", IEEE SoutheastCon, 2004.
[8] J. Lee, Y.-H. Chen, and Y.-W. Chiu, "Convergence Investigation of the Generalized Multi delay Adaptive Filter, Journal of Signal and Image Processing", ACTA, pp #449, 2004.
[9] D. Kim, "Classification of Ultrasonic NDE Signals using the LMS Algorithm and SAFT based Imaging", Journal of Signal and Image Processing, ACTA, pp# 479, 2005.
[10] S. Qian, Introduction to Time_Frequency and Wavelet Transforms, New Jersey: Prentice Hall, 2002.
[11] M. H. Hayes, Statistical Digital Signal Processing and Modeling, New York: John Wiley & Sons, 1996, pp. 408-412.
[12] Sesar Navarro, Colin Turner, Omar Escalona, " A Method for the ECG Inverse Problem in the Frequency Domain" Proceeding of the 18th IEEE Symposium on Computer-Based Medical Systems, IEEE 2005.
[14] Karel J.M.H., Peeters R.L.M., Westra R.L., Moermans K.M.S., Haddad S.A.P. and Serdijn W.A., Optimal discrete wavelet design for cardiac signal processing, Proceedings of the 2005 IEEE, Engineering in Medicine and Biology 27th Annual Conference.
[15] Mahamoodabadi S.Z., Ahmedian A., Abolhasani M.D., ECG feature extraction using daubechies wavelet, Proceedings of the fifth IASTED International Conference VISUALIZATION, IMAGING and IMAGE PROCESSING, September 7-9, 2005,
Benidorm, Spain.
[16] David Cuesta-Frau, Daniel Novak, Vladimir Eck, Juan C. Pérez-Cortés, Electrocardiogram Baseline Removal Using Wavelet Approximations, Computer-Based Medical Systems, (CBMS 2002), Proceedings of the 15th IEEE Symposium 2002.
[17] Gilbert Strang, Truong Nguyen, (1997), Wavelets and Filter Banks, Wellesley-Cambridge Press, 1997.
[18] Soman K.P., Ramachandran K.I., Insight into wavelets from theory to practice. Prentice-Hall of India, 2004
[19] R.Shantha Selva Kumari, S. Bharathi, V. Sadasivam, Design of Optimal Discrete Wavelet for ECG Signal using orthogonal filter bank, International Conference on Computational Intelligence and Multimedia Applications, IEEE 2007.
[20] N. Nikolaev, Z. Nikolov, A. Gotchev, K. Egiazarian, Wavelet Domain Wiener Filtering For Ecg Denoising Using Improved Signal Estimate, IEEE2000.
[21] Oktem R., Yaroslavsky L., Egiazarian K. "Signal and Image De-noising in Transform Domain and Wavelet Shrinkage: *A* Comparative Study". Proceedings of EUSIPCO-98, pp. 2269-2272, Island of Rhodes, Greece, Sept.1998.

High Quality Integrated Data Reconstruction for Medical Applications

A.K.M Fazlul Haque
Department of Computer Science and Engineering,
Jahangirnagar University,
Dhaka, Bangladesh.
Email: akm_haque@yahoo.com

Md. Hanif Ali
Department of Computer Science and Engineering
Jahangirnagar University,
Dhaka, Bangladesh.
Email: hanif_ju03@yahoo.com

M Adnan Kiber
Department of Applied Physics, Electronics and Communication Engineering, University of Dhaka,
Bangladesh.
Email: ma_kiber@yahoo.com

Abstract **-- In this paper, the implementation of a high quality integrated data reconstruction model and algorithm has been proposed, especially for medical applications. Patients' Information acquired at the sending end and reconstructed at the receiving end by using a technique that would be high quality for the signal reconstruction process. A method is proposed in which the reconstruction of data like ECG, audio and other patients' vital parameters that are acquired in the time-domain and operated in the frequency-domain. Further the data will be reconstructed in the time-domain from the frequency domain where high quality data is required. In this particular case, high quality ensures the distortion less and noiseless recovered baseband signal. This would usually require the application of Fast Fourier Transform (FFT) and Inverse Fast Fourier Transform (IFFT) return the data to the spatial domain. The simulation is performed using Matlab. The Composite baseband signal has been generated by developing a program as well as by acquiring to the workspace. The feature of the method is that it can achieve high-quality integrated data reconstruction and can be associated easily with spatial domain.**

Keywords: **FFT, IFFT, ECG, Baseband, Reconstruction, Noise, FDA tool.**

I. INTRODUCTION

High quality integrated data reconstruction model and algorithm are used, within an electronic system, to extract the desired time-domain signal from the frequency-domain signal acquired from the human body (in offline), especially Electrocardiogram (ECG), audio and other vital parameters [1]. Each stage automatically generates a template of a source from the candidate events in the initialization period, and thereafter performs classification of the remaining candidate events based on a template matching technique. Matlab simulation results on offline demonstrate the effectiveness of the proposed method.

In recent literature [2, 3, 4], the perception of conveying vital information reconstruction used by medical practitioners has had some concentration. Li, Mueller and Ernst [2] emphasized on the methods for efficient, high quality volume resampling in the frequency domain. It was described in the use of frequency-domain filters for the accurate resampling of images and volumes at arbitrary magnification factor. It was also investigated the Freq approach in relation to higher-quality filters.

Chazan, Hoory, Cohen, and Zibulski [3] illustrated speech reconstruction from Mel frequency cepstral coefficients and pitch frequency. They presented a novel low complexity, frequency domain algorithm for reconstruction of speech from the Mel frequency cepstral coefficient. The construction technique was based on the sinusoidal speech representation.

Kong [4] focused on GPS modeling in Frequency Domain. The author presented a frequency domain modeling approach to model GPS errors and increase GPS positioning shaping filter. The application of approach was mainly used for vehicle navigation system.

In this paper, most of the information contained in the baseband signal is found below 100 Hz. High frequencies random noise may corrupt the reconstructed time domain baseband signal. To remedy the situation, Filter Design Tool (FDA) has been used to eliminate the high frequency component. There is no distortion appearing in that particular spatial domain instead of attenuation amplified by the gain of the signal.

Bioelectrical signals are typically very small in amplitude (mV) and an amplifier is required to accurately depending on the hardware and software used, the biological amplifier serves to amplify the signal. It is also known that the frequency of heart signals is very low, approximately 5 to 10 H_Z [5, 6].

II. MATERIALS AND METHODS

The reconstruction process would be in order of data collection, Discrete Fourier transform (DFT), Fast Fourier Transform (FFT), Inverse Fourier Transform (IFFT), and finally, noise cancellation. These topics are discussed below:

The discrete Fourier transform, or DFT, is the primary tool of digital signal processing. The foundation of the Signal Processing Toolbox is the fast Fourier transform

(FFT), a method for computing the DFT with reduced execution time [7, 8]. Many of the toolbox functions (including z-domain frequency response, spectrum and cepstrum analysis, and some filter design and implementation functions) incorporate the FFT. MATLAB provides the functions FFT and IFFT to compute the discrete Fourier transform and it's inverse, respectively. For the input sequence x and its transformed version X (the discrete-time Fourier transform at equally spaced frequencies around the unit circle), the two functions implement the relationships.

$$X(k+1) = \sum_{n=0}^{N-1} x(n+1) W_N^{kn}$$

$$x(n+1) = \frac{1}{N} \sum_{k=0}^{N-1} X(k+1) W_N^{-kn}$$

In these equations, the series subscripts begin with 1 instead of 0 because of the MATLAB vector indexing scheme, and

$$W_N = e^{-j\left(\frac{2\pi}{N}\right)}$$

Working with the Fourier transform on a computer usually involves a form of the transform known as the discrete Fourier transform (DFT). There are two principal reasons for using this form: The input and output of the DFT are both discrete, which makes it convenient for computer manipulations. There is a fast algorithm for computing the DFT known as the fast Fourier transform (FFT).

The DFT is usually defined for a discrete function f (m, n) that is nonzero only over the finite region 0<= m <= M and 0<= n <= N. The two-dimensional M-by-N DFT and inverse M-by-N DFT relationships are given by

The values F (p , q) are the DFT coefficients of f (m , n).

$$F(p,q) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m,n) e^{-j(2\pi/M)pm} e^{-j(2\pi/N)qn} \qquad \begin{array}{l} p = 0,1,\ldots,M-1 \\ q = 0,1,\ldots,N-1 \end{array}$$

$$f(m,n) = \frac{1}{MN} \sum_{p=0}^{M-1} \sum_{q=0}^{N-1} F(p,q) e^{j(2\pi/M)pm} e^{j(2\pi/N)qn} \qquad \begin{array}{l} m = 0,1,\ldots,M-1 \\ n = 0,1,\ldots,N-1 \end{array}$$

The zero-frequency coefficient, F (0, 0), is often called the "DC component." DC is an electrical engineering term that stands for direct current. (Note that matrix indices in MATLAB always start at 1 rather than 0; therefore, the matrix elements f(1,1) and F(1,1) correspond to the mathematical quantities f (0, 0) and F (0 , 0) , respectively.)

The DFT coefficients F (p , q) are samples of the Fourier transform $F(\omega_1, \omega_2)$.

$$F(p,q) = F(\omega_1, \omega_2)\Big|_{\substack{\omega_1 = 2\pi p/M \\ \omega_2 = 2\pi q/N}} \qquad \begin{array}{l} p = 0,1,\ldots,M-1 \\ q = 0,1,\ldots,N-1 \end{array}$$

III. DATA COLLECTION

Data collection process can be acquired by real time or off-line. In real time, data, especially ECG are acquired by data acquisition tools. Other vital parameters like audio, video can be acquired by existing audio-video editing software [9]. If the existing software does not support to acquire the above parameters, it can be considered for developing the software to perform the acquisition properly. The array of electrodes used for the capture of Body Surface Potentials (BSP) comprises a flexible plastic anterior and posterior electrode harness and a portable recording unit. The anterior harness contains 64 electrodes, including 3 proximal bipolar limb leads (Mason-Likar position), and a posterior harness with 16 electrodes. This lead configuration enables recording of 77 unipolar ECG signals with respect to the Wilson central terminal. Recordings are taken over 4198 ms, the sampling rate is 1 kHz and the potentials are recorded using a 12-bit resolution to represent a 5 mV range. The procedure for attaching the harnesses to a patient and taking measurements of electrical potentials lasts about 3-4 minutes [10].

In this regard, data have been taken off-line generated by the software Matlab. It is worth mentioning that different data carries the different information that would be recovered by the method given in the next part of this paper.

IV. PROPOSED METHOD

The algorithm is based on using a sinusoidal model, where audio signal and ECG are represented by workspace block converted the captured file, especially .wav file into 2-dimentional data file. signals are characterized by the frequencies, amplitudes and phases of the sine-wave components. The complete reconstruction scheme is presented in Figure. 1. It is applied in every 10 ms. frame using the simulink block system's rules and regulations.

The algorithm is comprised of the following stages:

1. Sum block - Add or subtract inputs. Specify one of the following:

a) String containing + or – for each input port.

b) Scalar >=1. A value >1 sums all inputs; 1 sums elements of a single input vector.

2. FFT - Outputs the complex Fast Fourier Transform (FFT) of real or complex input by computing radix-2 decimation-in-time (DIT) or decimation-in-frequency (DIF), depending on block options. Uses half-length and double-signal algorithm for real inputs where possible. Computes the FFT

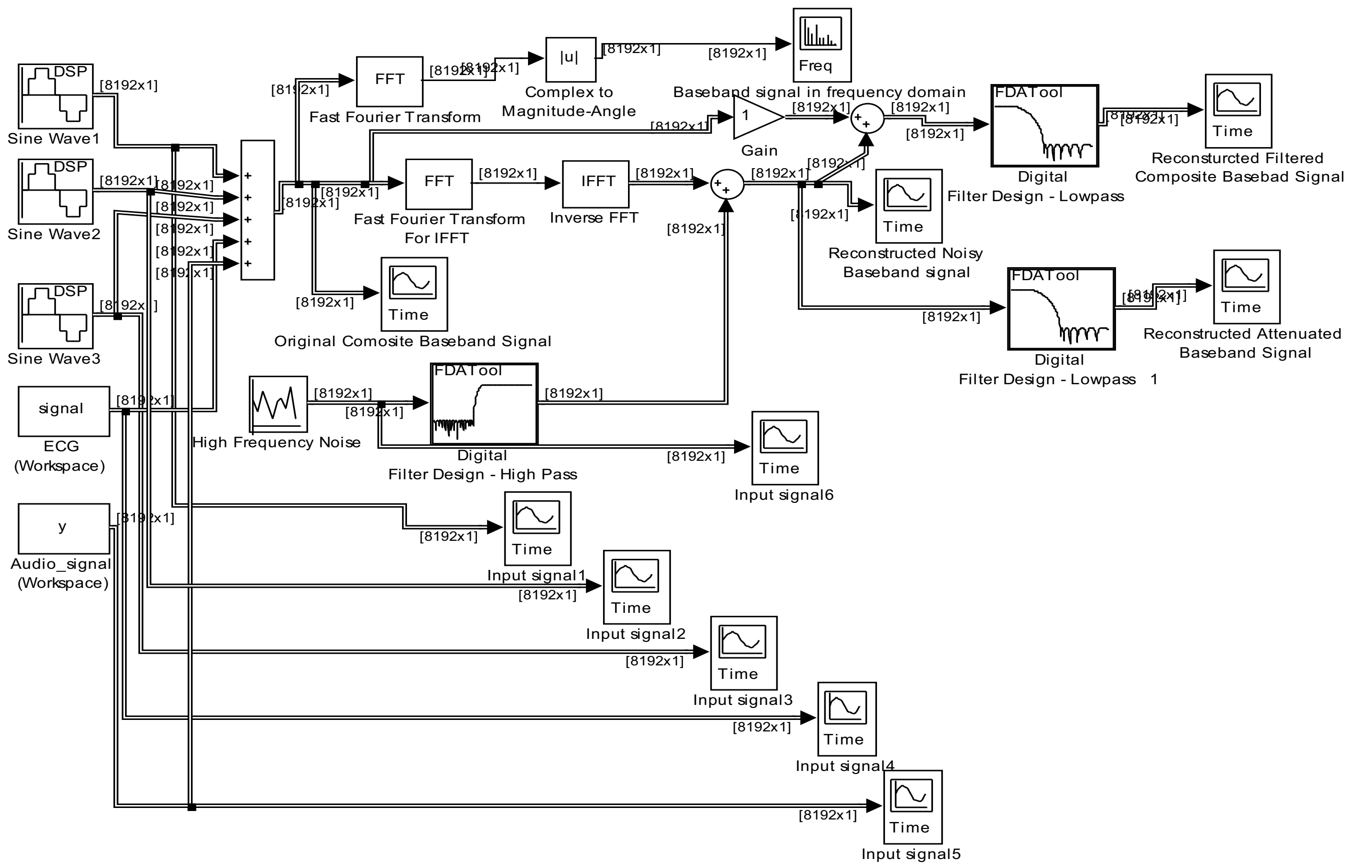


Figure 1. Block Diagram of proposed simulation method

along the vector dimension for sample-based vector inputs, which must have a power-of-2 length. Computes the FFT along each column for all other inputs, where columns must be a power-of-2 length.

3. IFFT - Outputs the Inverse Fast Fourier Transform (FFT) of real or complex input by computing radix-2 decimation-in-time (DIT) or decimation-in-frequency (DIF), depending on block options.

Outputs are real if we select 'Input is conjugate symmetric' option; otherwise, outputs are complex. Computes the IFFT along the vector dimension for sample-based vector inputs, which must have a power-of-2 length. Computes the IFFT along each column for all other inputs, where columns must be a power-of-2 length.

4. FDA Tool - The block applies the specified filter to each channel of a discrete-time input signal, and outputs the result. The outputs of the block numerically match the outputs of the Digital Filter block, the filter function in MATLAB, and the filter function in the Filter Design Toolbox. The sampling frequency, Fs, which specified in the FDA tool's Graphical User Interface (GUI), should be identical to the sampling frequency of the digital filter design block's input block. When the sampling frequencies of these blocks do not match, the digital filter design block returns a warning message and inherits the sampling frequency of the input block.

The block accepts inputs that are sample-based or frame-based vectors and matrices. The block filters each input channel independently over time, where

- Each column of a frame-based vector or matrix is an independent channel.
- Each element of a sample-based vector or matrix is an independent channel.

5. Workspace - Simulink allows us to import input signal and initial state data from the MATLAB workspace and export output signal and state data to the MATLAB workspace during simulation. This capability allows us to use standard or custom MATLAB functions to generate a simulated system's input signals and to graph, analyze, or otherwise postprocess the system's outputs STFT reconstruction (Fig. 1) - The frequencies, phases and

amplitudes are combined to form a sine-wave representation. The final reconstructed STFT is constructed from the sine waves by a convolution procedure.

To use this format, it selects input in the load from workspace pane and selects the Array option from the Format list on the data import/export pane. Selecting this option causes Simulink to evaluate the expression next to the Input check box and use the result as the input to the model. The expression must evaluate to a real (noncomplex) matrix of data type double. The first column of the matrix must be a vector of times in ascending order. The remaining columns specify input values. In particular, each column represents the input for a different Import block signal (in sequential order) and each row is the input value for the corresponding time point. Simulink linearly interpolates or extrapolates input values as necessary if the Interpolate data option is selected for the corresponding import. The total number of columns of the input matrix must equal n + 1, where n is the total number of signals entering the model's imports.

6. Vector Scope - The vector scope block is a comprehensive display tool similar to a digital oscilloscope. The block can display time-domain, frequency-domain, or user-defined signals. We can use the Vector Scope block to plot consecutive time samples from a frame-based vector, or to plot vectors containing data such as filter coefficients or spectral magnitudes. To compute and plot the periodogram of a signal with a single block, use the Spectrum Scope block. The input to the Vector Scope block can be any real-valued M-by-N matrix, column or row vector, or 1-D (unoriented) vector, where 1-D vectors are treated as column vectors. Regardless of the input frame status, the block treats each column of an M-by-N input as an independent channel of data with M consecutive samples. The block plots each sample of each input channel sequentially across the horizontal axis of the plot conversion to time domain (Fig. 1) - This is the final stage in which the time domain signal frame is computed.

7. Sine wave - The sine wave block may generate a multichannel real or complex sinusoidal signal, with independent amplitude, frequency, and phase in each output channel. A real sinusoidal signal is generated when the output complexity parameter is set to Real, and is defined by an expression of the type

$$y = A \sin (2\pi ft + \varphi)$$

where A in the amplitude parameter, f in hertz in the frequency parameter, and in radians in the phase offset parameter. A complex exponential signal is generated when the output complexity parameter is set to complex, and is defined by an expression of the type

$$y = Ae^{j(2\pi ft+\phi)} = A\{\cos(2\pi ft+\phi)+j\sin(2\pi ft+\phi)\}$$

8. Random source - Output a random signal with uniform or Gaussian (normal) distribution. Set output repeatability to nonrepeatable (block randomly selects initial seed every time simulation starts), repeatable (block randomly selects initial seed once and uses it every time simulation starts), or specify seed (block uses specified initial seed every time simulation, starts producing repeatable output).

V. ANALYSIS AND DISCUSSION

The first stage of the implementation is to pass the simulated integrated baseband signal corrupt by high frequency random noise. The method tries to produce the filtered simulated distortion less and noiseless integrated baseband signal in a relatively short period of time. Increasing the filter order above one slows down the convergence rate but makes the results more precise. The recovered signal closely resembles the original simulated signal minus the noise and hence indicates that the implementation of the algorithm functions correctly and efficiently. The second stage is to recover the original composite baseband attenuated signal using FDA tool. Finally, the original composite banseband signal has been recovered where noise, distortion and attenuation have been eliminated.

High frequency noise is used, as the signal corrupt by noise as well as the reference noise. With this setup, it is observed that the output signal has been attenuated a little bit but not distorted due to the affect of distortion. But the attenuation would be disappeared by amplifying the gain of the signal. All involved steps of the simulation are discussed below.

Step 1: Generation of different sinusoidal signals having amplitude of 1 mV, ECG pattern having amplitude of 3.5mV and audio signal having amplitude of 0-1 mV. Figure 2, 3, 4, 5, and 6 show the generation of input signal. Here ECG signal has been generated in offline and the time domain data imported from the workspace. The data of the audio signal may take from stored file or real time. In this particular case, it has been taken from the stored file and the time domain data also imported from the workspace.

Step 2: The composite baseband signal for the different input signals (Amplitude of 4.5 mV) has been showed in Fig. 7.

Step 3: Fig. 8 shows the original composite baseband signal in frequency domain.

Step 4: Generation of high frequency random noise signal amplitude of 3.5 mV is shown in Fig. 9.

Step 5: The noisy IFFT signal (recovered baseband signal + noise) is shown in Fig. 10.

Step 6: Fig. 11 shows the reconstructed filtered attenuated composite baseband signal using FDA tool.

Step 7: Fig. 12 shows the original reconstructed filtered composite baseband signal using FDA tool.

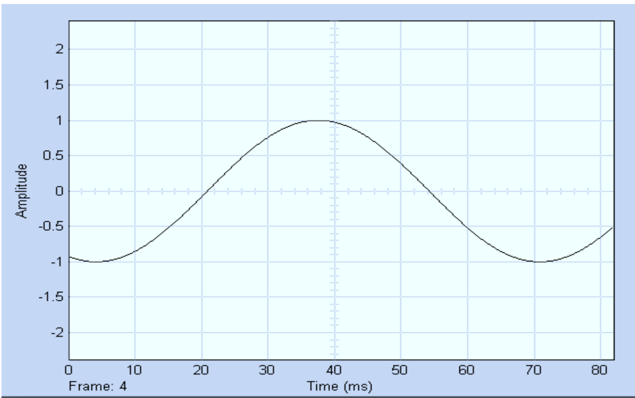


Figure 2. Sinusoidal signal with frequency of 15 Hz

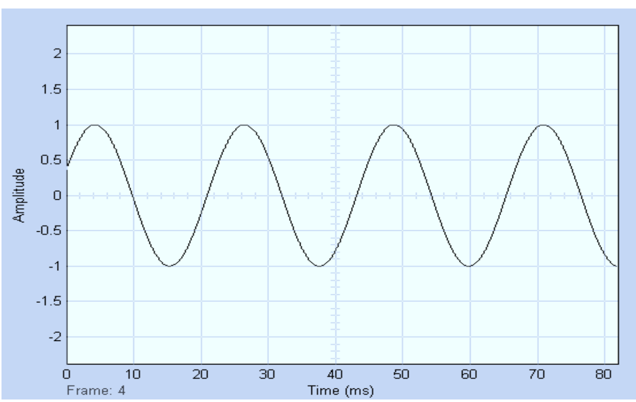


Figure 3. Sinusoidal signal with frequency of 45 Hz

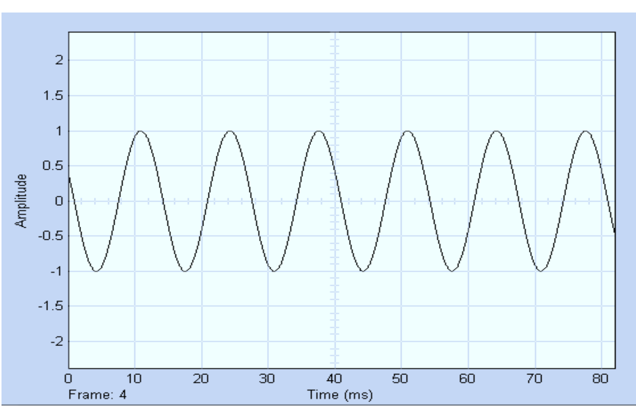


Figure 4. Sinusoidal signal with frequency of 75 Hz

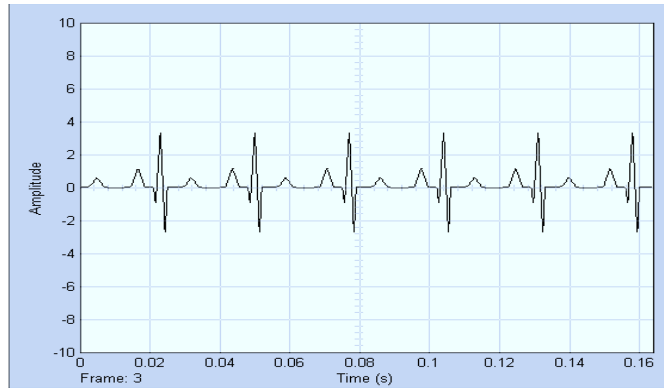


Figure 5. ECG Signal Generated from workspace

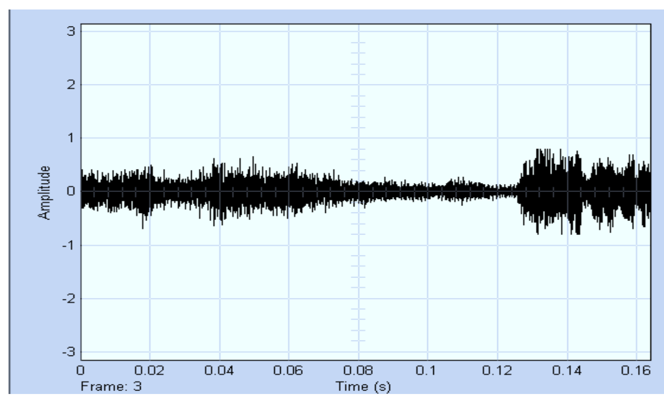


Figure 6. Audio Signal Generated from workspace.

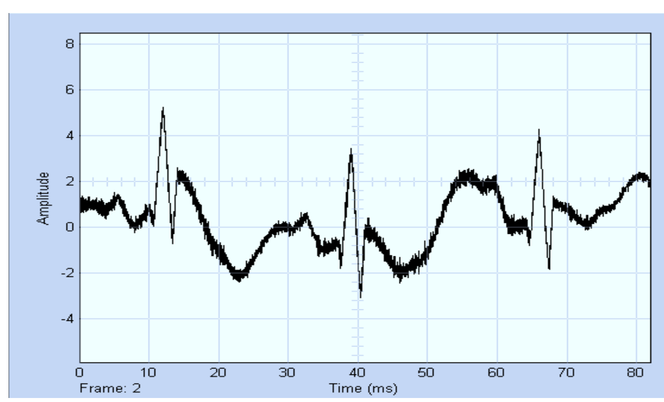


Figure 7. Original Composite Baseband Signal.

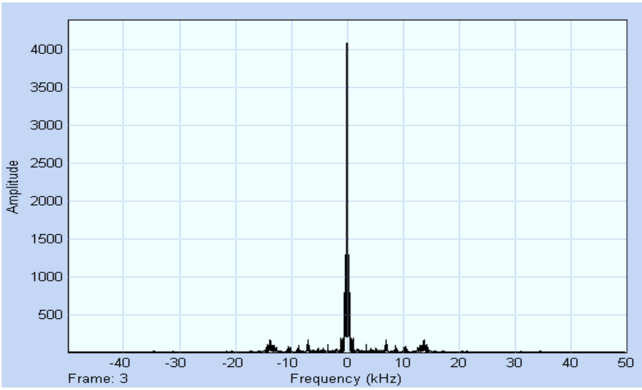


Figure 8. Original Composite Baseband Signal in Frequency Domain.

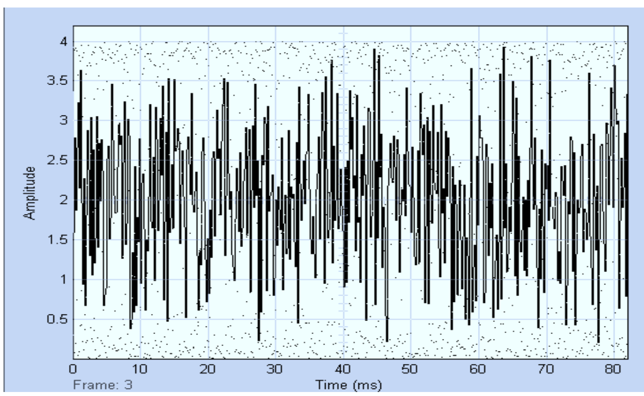


Figure 9. High frequency Random Noise.

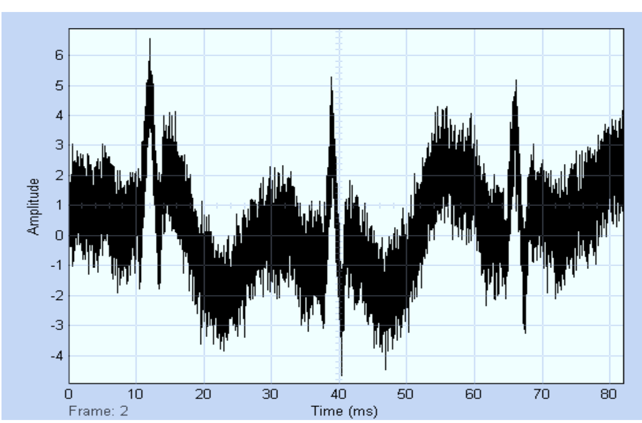


Figure 10. Recovered Noisy Baseband Signal.

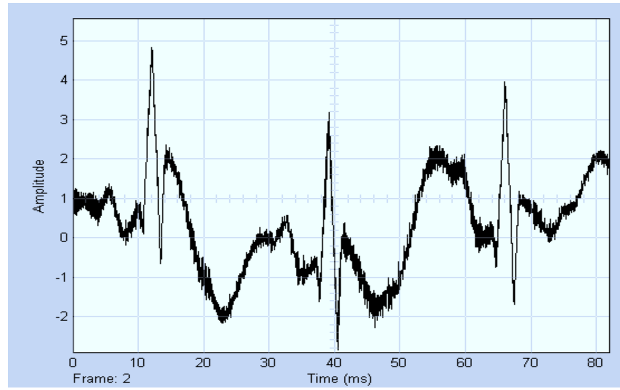


Figure 11. Original Filtered Reconstructed attenuated Signal.

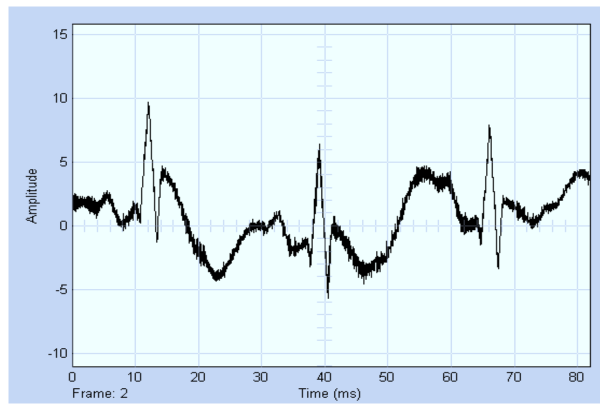


Figure 12. Original Filtered Reconstructed Signal.

VI. CONCLUSION

In this work, an attempt has been made to employ a method that can be implemented for reconstructing a high quality integrated data, especially for medical applications. Distortions of reconstructed filtered baseband signal may occur due to the affect of frequency offset in integrated data that have been eliminated. Generally, the reconstruction of data like ECG, audio and other patients' vital parameters that are acquired in the time-domain and operated in the frequency-domain. Data acquisition and the simulation are performed using Matlab. The Composite baseband signal, especially ECG, has been generated by developing a program. Filter Design Tool (FDA) has been used to eliminate the high frequency component and the original filtered baseband signal has been obtained. Distortions have been eliminated in that particular spatial domain and attenuation has been eliminated by amplifying the gain of the signal. This approach may be considered for reconstruction in the advanced technology which needs further investigation.

VII. REFERENCES

[1] R. Aston, Principles of biomedical instrumentation and measurement. Columbus: Merrill Pub. Co., 1990.

[2] Aili Li, Klaus Mueller, Thomas Ernst, “Methods for Efficient, High quality Volume Resampling in the Frequency Domain”, IEEE Visualization 2004.

[3] Dan Chazan, Ron Hoory, Gilad Cohen and Meir Zibulski, "Speech Reconstruction From Mel Frequency Cepstral Coefficients and Pitch Frequency” Acoustics, Speech, and Signal Processing, Vol 3, IEEE 2000.

[4] Xiaoying Kong , "GPS Modeling in Frequency Domain" presented at 2nd International Conference on Wireless Broadband and Ultra Wideband Communications, IEEE 2007.

[5] Monitoring Patients via Radio Signals. About W*ireless Medical Telemetry Service (WMTS).*

[6] A.K.M Fazlul Haque, M. Shamim Kaiser, S. K. Aditya, “Background Noise Cancellation Using Adaptive LMS Algorithm” Journal of Bangladesh Electronic Society, Volume 6, Issue 2.

[8] Lianming Sun, Akira Sano, "Identification and Error Analysis in Frequency Domain," Proceedings of the First International Conference on Innovative Computing, Information and Control, IEEE 2006.

[9] Mircea-Florin VAIDA, Ionel STAMATE,”Integrated Medical Image Processing Facilities Using a Dedicated Application” International Conference on Automation, Quality and Testing, Robotics , IEEE 20006.

[10] Sesar Navarro, Colin Turner, Omar Escalona, “A Method for the ECG Inverse Problem in the Frequency Domain” Proceeding of the 18th IEEE Symposium on Computer-Based Medical Systems, IEEE 2005.

AN ELECTRONIC DESIGN OF A LOW COST BRAILLE HANDGLOVE

M.Rajasenathipathi [1], M.Arthanari [2], and M.Sivakumar [3]
[1] Assistant Professor in Computer Science, N.G.M. College-642001, Tamilnadu, India.
[2] Director, Bharathidasan School of Computer Applications, Ellispettai-638116, Tamilnadu, India
[3] Doctoral Research Scholar, Anna University, Coimbatore, Tamilnadu , India
[1] Senathi_pathi@yahoo.co.in, [2] arthanarimsvc@gmail.com, [3] sivala@gmail.com

Abstract— **This paper documents a new design for a Braille Hand glove, comprising of a majority of electrical components, the design aims to produce a product to perform vibrations in six position of blind's person right hand. A low cost and robust design will provide the blind with an affordable and reliable tool also it produce the new technique and communications method for blind persons.**

Keywords- **Braille, cell, vibration, dots, motor.**

I. INTRODUCTION

Braille is an important language used by the blind to read and write. It is vital for communication and educational purposes. The Braille code has become the main system for the majority of those blind people who read and write using tactile means, and can be found in many countries around the world. Braille uses raised dots in groups of six which were arranged in three rows to two. These six arranged in three rows to two. These six positions which can be raised or flat, are used in combination to give just 64 different Braille characters. This clearly means that there can be a one to one correspondence between Braille characters and text. The blind person touch raised dots and understands the English characters. The proposed Braille Hand glove contains six vibrations motors in five fingers and center palm. These six positions are matched to six values of Braille code. So instead of touching the raised Dots in Braille sheet, this Braille Hand glove produces vibration based on English character value.

II. BACKGROUND DETAILS

A. What is Braille

All over the world, persons with visual handicaps have used Braille as the primary means to reading information. Also, the concept of Braille has been accepted as a universal approach that works across the boundaries of the world. Different countries of the world have adapted the system of Braille to suit their languages. Irrespective of these changes or modifications, Visually Handicapped persons understand standard Braille for the Roman alphabet (English Braille) making it possible to exchange information in a consistent fashion across different countries.

B. Brief Introduction to Braille

Standard Braille is an approach to creating documents which could be read through touch. This is accomplished through the concept of a Braille cell consisting of raised dots on thick sheet of paper. The protrusion of the dot is achieved through a process of embossing. A cell consists of six dots arranged in the form of a rectangular grid of two dots horizontally and three dots vertically. With six dots arranged this way, one can obtain sixty three different patterns of dots. A visually Handicapped person is taught Braille by training him or her in discerning the cells by touch, accomplished through his or her fingertips. The image below shows how this is done.

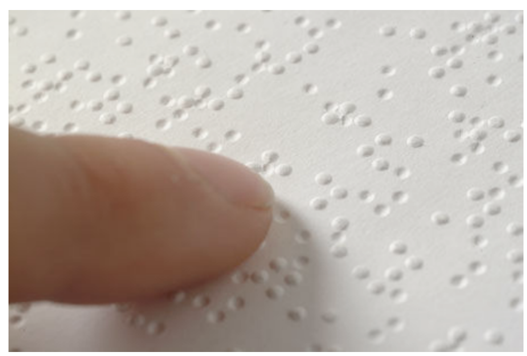


Fig. 1 Braille sheet

Each arrangement of dots is known as a cell and will consist of at least one raised dot and a maximum of six.

C. The Braille Cell

A printed sheet of Braille normally contains upwards of twenty five rows of text with forty cells in each row. The physical dimensions of a standard Braille sheet are approximately 11 inches by 11 inches. The dimensions of the Braille cell are also standardized but these may vary slightly depending on the country. The dimension of a Braille cell, as printed on an embosser is shown below.

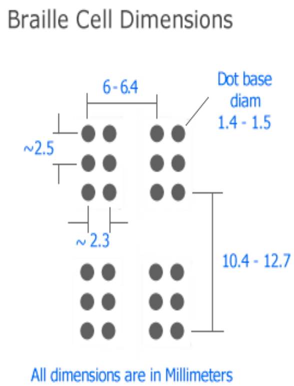


Fig. 2 A Braille cell

The six dots forming the cell permit sixty three different patterns of dot arrangements. Strictly, it is sixty four patterns but the last one is a cell without any dots and thus serves the purpose of a space. A Braille cell is thus an equivalent of a six bit character code, if we view it in the light of text representation in a computer! However, it is not related to any character code in use with computers.

D. Standard English Braille

In standard English Braille, many of the sixty three cells will correspond to a letter of the Roman alphabet, or a punctuation mark. A few cells will represent short words or syllables that are frequently encountered in English. This is done so that the number of cells required to show a sentence may be reduced, which helps minimize the space requirements while printing Braille.

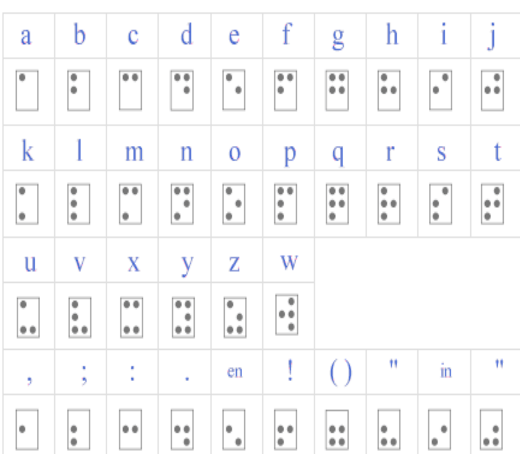


Fig. 3 Letters and Special symbols of Braille

III. COMPARISON ABOUT BRAILLE CELL AND BRAILLE HAND GLOVE

(Basic Idea behind the Braille Glove)

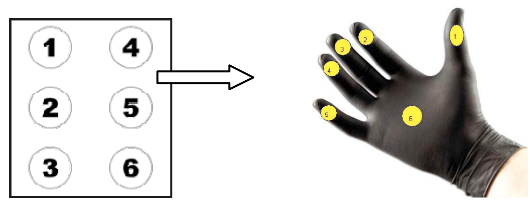


Fig. 4 Hand glove with six positions

The six dots forming the cell permit sixty three different patterns of dot arrangements. It is matched with alphabets, numbers and special symbols of the English language. The Braille glove contains six vibration motors. These are fixed in five fingers and center palm. The basic technique used in the hand glove based on retrieval value of English letter value from the user typed in the keyboard. It is converted into Braille value and activated the corresponding motors. So based on the position of vibration the blind person can understand the value of the letter. For example if the user can type the letter "r". It is converted int Braille value as 1,2,3,5 and this value activates the corresponding motors in Braille hand glove. This conversion program in written in high tech C language and it is recorded in micro controller of the hand glove. Any blind person can wear this glove in right hand, and understand the English letters through vibration instead of touch the Braille sheet. Similarly the whole word or sentence is converted into Braille vibration and send to blind person. Based on this method the visible person and deaf and blind person can communicate effectively.

IV. THE DESIGN CONCEPT

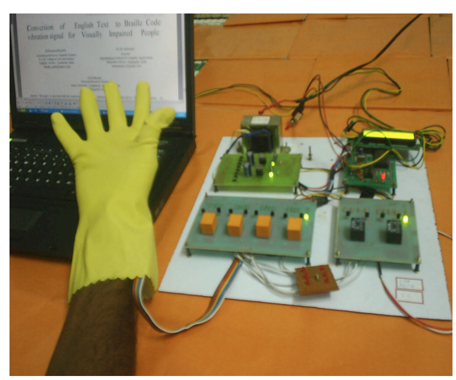


Fig. 5 Vibration Hand glove

The Braille Hand glove will be comprised of the following key components

1. 89C51 Micro controller
2. RS 232 C
3. Relay Driver and Relay
4. power supplies
5. Vibrator motor in hand glove

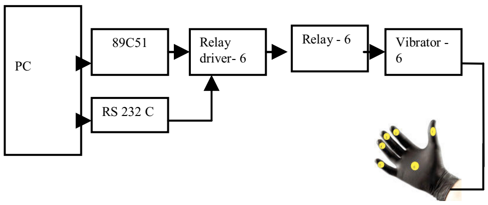


Fig. 6 Block diagram of Braille hand glove

A. Micro controller

Microcontroller is a general purpose device, which integrates a number of the components of a microprocessor system onto single chip. It has inbuilt CPU, memory and peripherals to make it as a mini computer. A microcontroller is integrated with

1. CPU Core
2. RAM and ROM
3. Some parallel digital i/o ICs

The vibration hand glove contains a microcontroller AT89C51. It is the 40 pins, 8 bit Microcontroller manufactured by Atmel group. It is the flash type reprogrammable memory. Advantage of this flash memory is we can erase the program within a few minutes. It has 4KB on chip ROM and 128 bytes internal RAM and 32 I/O pin as arranged as port 0 to port 3 each has 8 bit bin .Port 0 contains 8 data line(D0-D7) as well as low order address line(AO-A7).

The position identification and controlling the motors is programmed in hi tech c language and is loaded in microcontroller.

1) Crystal:

The heart of the micro controller is the circuitries which generate the clock pulse. Then micro controller provides the two pins. XTAL 1, XTAL 2 to correct the external crystal resonator along with capacitor. The crystal frequency is the basic clock frequency of the microcontroller. Based on the frequency rotation time of vibration motor inside the hand glove is controlled by micro controller.

2) Reset:

The memory location for 89C51 0000H to 0FFFH. Whenever switch on the supply the memory location starts from 0000H.The 89C51 micro controller provide 9th pin for Reset Function. Here the reset circuitry consists of 10Mf capacitor in series with 10K resister. When switch on the supply the capacitor is changed and discharged gives high low pulse to the 9th pin through the 7414 inverter. Here we interface LCD display to microcontroller via port 0 and port 2. LCD control lines are connected in port 2 and Data lines are connected in port 0. whenever struggle in motor speed, it is used to restart the program.

3) LCD:

Liquid Crystal Display has 16 pins in which first three and 15th pins are used for power supply. 4th pin is RS(Register Selection) if it is low data and if it is high command will be displayed. 5th pin is R/W if it is low it performs write operation. 6th pin act as enable and remaining pins are data lines

B. RS232 Communication

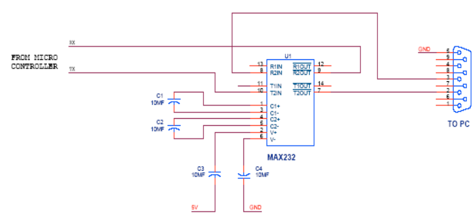


Fig. 7 circuit diagram of RS232

In viberation hand glove **RS-232** is a standard for serial binary data interconnection between a *DTE* (Data terminal equipment) and a *DCE* (Data Circuit-terminating Equipment). It is commonly used in computer serial ports.Here ascii values are converted into binary signals and send to viberation glove to activates the viberation motors.

.

Details of the character format and transmission bit rate are controlled by the serial port hardware, often a single integrated circuit called a UART that converts data from parallel to serial form. A typical serial port includes specialized driver and receiver integrated circuits to convert between internal logic levels and RS-232 compatible signal levels.

C. *Relay:*

A relay is an electrically operated switch. Current flowing through the coil of the relay creates a magnetic field which attracts a lever and changes the switch contacts. The coil current can be on or off so relays have two switch positions and they are double throw (changeover) switches. The coil of a relay passes a relatively large current, typically 30mA for a 12V relay, but it can be as much as 100mA for relays designed to operate from lower voltages. Most ICs (chips) cannot provide this current and a transistor is usually used to amplify the small IC current to the larger value required for the relay coil. The maximum output current for the popular 555 timer IC is 200mA so these devices can supply relay coils directly without amplification.

Fig. 8 Relay

Relays are usually SPDT or DPDT but they can have many more sets of switch contacts, for example relays with 4 sets of changeover contacts are readily available. Most relays are designed for PCB mounting but you can solder wires directly to the pins providing you take care to avoid melting the plastic case of the relay. The animated picture shows a working relay with its coil and switch contacts. This lever moves the switch contacts. There is one set of contacts (SPDT) in the foreground and another behind them, making the relay DPDT.

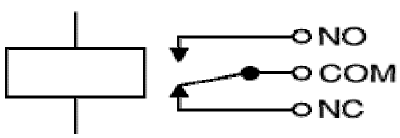


Fig. 9 Relay-Switch connection

The relay's switch connections are usually labeled COM, NC and NO:

- **COM** = Common, always connect to this, it is the moving part of the switch.
- **NC** = Normally Closed, COM is connected to this when the relay coil is **off**.
- **NO** = Normally Open, COM is connected to this when the relay coil is **on**.

D. *Power supplies*

A block diagram containing the parts of a typical power supply and the to a transformer, which steps that ac voltage down to the level for the desired dc output. A diode voltage at various points in the unit is shown in fig 19.1. The ac voltage, typically 120 V rms, is connected rectifier then provides a full-wave rectified voltage that is initially filtered by a simple capacitor filter to produce a dc voltage. This resulting dc voltage usually has some ripple or ac voltage variation. A regulator circuit can use this dc input to provide a dc voltage that not only has much less ripple voltage but also remains the same dc value even if the input dc voltage varies somewhat, or the load connected to the output dc voltage changes. This voltage regulation is usually obtained using one of a number of popular voltage regulator IC units.

E. Vibrator Motors

The main component in Braille glove is vibration motor. it is configured in two basic varieties ie coin (or flat) and cylinder (or bar). Cylinder type motors are simple brush motors with a traditional axial design. The eccentric movement of the weight attached to the rotor provides vibration during operation. In Braille glove it is best suited in finger positions. The amount of vibration is directly proportional to the voltage applied to the motor. Cylinder motors are manufactured in high volumes and are fairly inexpensive. An electrical current applied to the coil in the direction of the arrow generates upward force on the left side of the coil and downward force on the right side , causing the coil to revolve clockwise.

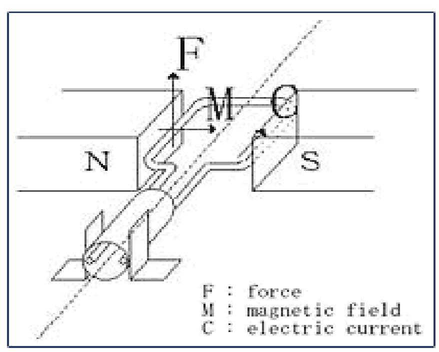


Fig. 10 working principle of viberation motor

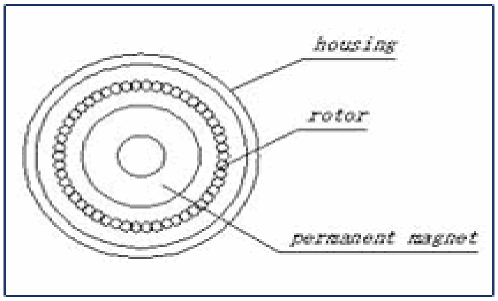


Fig. 11 cross section of coreless motors

1) *Principles of vibration:*

An offset counterweight is fitted to the end of the motor shaft. When the shaft turns, the imbalance in the counterweight causes the handset to vibrate.

Counter Weight		Speed[rpm]	Vibration[N]
4CH	R2.5×L3.5×180°	9500	1.50
	R3.0×L3.5×150°	9500	2.18
4CR	R3.0×L4.0×150°	9500	2.56
4.4CH	R3.0×L4.0×150°	7400	1.91
	R3.0×L4.0×150°	8500	2.37

Table .1 Motor speed based on weight

V. CONCLUSION AND FURTHER WORK SUGGESTED

The development of low cost Braille hand glove is necessary for visually impaired community. A number of private translation kits exist, but their private ownership restricts open development. The same technique can be used in various languages like Bengali, Hindi, Tamil, French, etc., other developments like through signal, and the blind can also convey reply to visible person produce the vibration hand glove as the best kit for two way communication.

VI. REFERENCES

[1] Basu SG ,lachikawa A (2004), Dialogue Languages and Persons with Disabilities, *IEICE transactions* , Vol. E87-D, No 6, pp 31-43.

[2] Blenkhorn (1995), A system for converting Braille into print, *IEEE Transactions on Rehabilitation Engineering* , Vol. 3 no 2, pp 215-221 .

[3] Choi H Lee W (2004), Tactile Display as Braille Display for the Visually Disabled, *in proceedings of IEEE/RSJ International Conference on Intelligent Robotics and Systems*, pp 1985-1990 .

[4] Fritz J P Barner (1999), Design of a Hepatic Visualization System for People with Visual Impairments, *IEEE Transactions on Rehabilitation Engineering*, vol. 7, No 3 pp 372-384 .

[5] Fukumoto et al (1997), Body coupled Finger Ring wireless wearable keyboard, *proceeding of the ACM conference on Human factors in computing systems ,France* , pp 147-154 .

[6] Gill J M (1992), priorities for technical research and development for visually disabled persons, *world Blind Union Res. Committee* .

[7] Golding P (1998), Design Study on a Braille Typewriter, *Western Australia, Curtin University of Technology*.

[8] Slaby W A(1990), Computerized Braille translation, *Microcomputer Applications* ,Vol.13,pp 107-113.

[9] Sleng C Lau Y(1999).,Regular feature extraction for recognition of Braille. *In Third international conference on Computational Intelligence and Multimedia Applications,ICCIM'99* Proceedings, pp 302-308.

[10] Werner H(1992),Automatic Braille production by means of computer in uses of computers in Aiding the Disabled, North Holland, pp 321-336 .

AUTHORS PROFILE

M.Rajasenathipathi (senathi_pathi@yahoo.co.in) is a third year Doctoral Research Scholar in the Research and Development Centre of Bharathiar University. He received his Master degree in computer Applications and M.Phil degree in computer science from M.K. University. He is also working as an Assistant Professor of Computer Science at N.G.M.College , Pollachi , India. He has been an active developer of systems for people with disabilities for the past 10 years.

M.Arthanari (arthanarimsvc@gmail.com) holds a Ph.D in Mathematics from Madras University as well as Masters degree in Computer Science from BITS, Pilani. He has worked as a Professor and Head , Department of CSE and IT in Tejaa Shakthi Institute of Technology for Women, Coimbatore, India and Head of the Institution in various colleges in India, He holds a patent issued by the Govt. of India for the invention in the field of Computer Science. He has directed teams of Ph.D researchers and industry experts for developing patentable products. He teaches strategy, project management, creative problem solving, innovation and integrated new product development for last 35 years.

Sivakumar.M (sivala@gmail.com) has 10+ years of experience in the software industry including Oracle Corporation. He received his Bachelor degree in Physics and Masters in Computer Applications from the Bharathiar University, India. He holds a patent for the invention in embedded technology. He is technically certified by various professional bodies like ITIL, IBM Rational Clear case Administrator, OCP - Oracle Certified Professional 10g and ISTQB.

A Test-Bed for Emergency Management Simulations

A case-study including Live Call Data in New Zealand

Anu Vaidyanathan
Department of Computer Science and Engineering
Indian Institute of Technology Ropar
Nangal Road, Rupnagar India.
anu@iitrpr.ac.in

Abstract*— *We present a test-bed for Emergency Management Simulations by contrasting two prototypes we have built, CAVIAR and Reverse 111. We outline the desirable design principles that guide our choices for simulating emergencies and implement these ideas in a modular system, which utilizes proactive crowd-sourcing to enable emergency response centers to contact civilians co-located with an emergency, to provide more information about the events. This aspect of proactive crowd-sourcing enables Emergency response centers to take into account that an emergency situation's inherent nature is dynamic and that initial assumptions while deploying resources to the emergency may not hold, as the emergency unfolds. A number of independent entities, governmental and non-governmental are known to interact while mitigating emergencies. Our test-bed utilizes a number of agents to simulate various resource sharing policies amongst different administrative domains and non-profit civilian organizations that might pool their resources at the time of an emergency. A common problem amongst first responders is the lack of interoperability amongst their devices. In our test-bed, we integrate live caller data obtained from traces generated by Telecom New Zealand, which tracks cell-phone users and their voice and data calls across the network, to identify co-located crowds. The test-bed has five important components including means to select and simulate Events, Resources and Crowds and additionally provide a visual interface as part of a massive online multi-player game to simulate Emergencies in any part of the world. We also present our initial evaluation of some resource sharing policies in our intelligent agents, which are part of our test-bed.

Keywords-test-bed, Emergency Management, Live Call Records, PCMD, Proactive Crowd-Sourcing, Agents

I. INTRODUCTION

Simulations in the field of Emergency Management, which adapt an end-to-end approach [40, 41] are important for a number of end uses including training personnel, understanding how administrative boundaries affect resource sharing and analyzing how to optimize resource-allocation and response time. Improving the efficiency of first responders is of greatest importance to response teams. Evaluations of operational alternatives are also of interest to these teams and can be easily programmed into a simulation. Normally, in the case of an Emergency or Disaster, several disparate teams work in unison to mitigate the situation at hand. This diversity in administrative units demands a dynamic system that is able to represent various entities and the resources they are able to make available to the situation. Evaluating network connectivity during an Emergency helps estimate available modes of communication with those affected and between groups of first responders. In our work, we introduce a layer to indicate how the infrastructure of Telecom New Zealand, specifically their cell-phone towers are laid out and how they handle calls during periods of high call traffic. We envision this to assist us in decision making in the case of an Emergency, during which time the call-traffic patterns are also high in volume. Dynamic Data Driven Application Systems (DDDAS) have been discussed to have the end goal of processing real-time experimental data to use as an input parameter for software simulations [1]. Since experimental data often requires simulations to make drastic changes [2], the end-goal for a *good* simulator has always been flexible design. Complex Adaptive Systems (CAS) and Agent Based Models (ABM) have also been proposed for modelling Emergency Operations Centers [3]. GEMASim models EOC operations by abstracting entities such as the SOC Chief, the Public Affairs Officer, the Operations Chief etc. into their own individual agents, to represent different functionality. A limitation of adhering too closely to prescribed service hierarchies is the lack of inclusion when it comes to local civilian groups. Integrated Gaming and Simulation for Incident Management has been proposed [4] wherein different modules exist for Plume Simulation, Crowd Simulation, Traffic Simulation (of social behaviours and vehicular traffic), health-care simulation, transportation simulation and an integrated gaming module for evaluating incident management. This does not introduce any network backbone or evaluate inter-operability of devices handled by first responders, each often working in its own frequency band [6].

A. Related Work

The Pacific Rim Visualization and Analytics Center coordinated at the University of Washington, known as the 2 PARVAC to other regional centres, has been working on developing improved shared artefacts for communicating during community-wide first response activities [8]. The use of sensors for Emergency Medical care is probably the most wide-spread. Bluetooth compatible devices have been studied in great detail to assimilate data and information and interface to a central server by means of a cellular phone [9,10,11]. Sensors are also used in the case of emergency evacuations [12] from within buildings. Emergency response simulation using sensors have been performed [13] in order to co-ordinate and control the dispatch and use of resources required. Specialized computation of emergency exit routes in case of toxic gas attack [14] have been discussed where models are used to compute the dispersion of the gases. Here the sensors are those used by the live weather web service which is able to predict and supply weather patterns for any zip code provided. The use of computers to control and co-ordinate emergency response have evolved from using static computer [15,28] to primarily using mobile sensors [16,17,18,21,27].The use of cellular phones [23, 24] has been mentioned in a few pieces of work however, there is no integrated solution for end-to-end disaster response or emergency management. The development of serious games for a variety of purposes extends across utility including role-playing in the case of Emergencies [32,37]. Several requirements exist for such games including a valid environment in which to train, practice and evaluate co-ordination amongst distributed entities, amongst various administrative domains. In general, these games can be easily extended to being *simulations* [40, 41] when played in single-player mode and therefore extend in functionality to studying performance of various approaches to live emergency response. The background of this work lies in a body of work that looked at early warning systems [38] which relied heavily on a prediction-based model to issue warnings to affected areas. In an Emergency System with online simulation models [36] that take input from sensors that measure environmental variables can be used by emergency services. [36] proposes an approach to establish information exchange and communication among many subsystems using a High Level Architecture. This work does not utilize live call information from telephony service providers as a layer in its visualization and is constrained in its ability to accept user-input to the system to simulate how first responders behave at the time of an emergency. User-Centered Design [35] is proposed in an All Hazard Exercise Development and Administration tool for emergency management agencies, once again to incorporate training elements in the simulation. The utility of knowledge sharing has been propounded in many works [33,42,43,34,44,45] The sensors proposed in these cases are very resource constrained in terms of hardware and not connected by any service-provider's network. In the case of *response* rather than warning, the work in the realm of simulations or games is rather sparse. Previous work on wireless infrastructures designed to address emergency scenarios while allowing intercommunication between teams in a scenario with existing network infrastructures and the pervasive grid network [39] do not propose citizen participation in the process of mitigation. Furthermore, the use of cell-phone data for emergency response is proposed in this work for the first time. Contacting callers on the ground by using the concept of *proactive crowd sourcing* is also proposed in order to assimilate accurate, real-time information about how the situation on the ground is changing, as an emergency evolves. By enabling modular design for rapid feedback from co-located callers, emergency response centers can better decide how to deploy resources to the emergency.

II. DESIGN PRINCIPLES FOR A TEST-BED FOR EMERGENCY MANAGEMENT

In this section, we discuss some of the desirable characteristics in a simulation environment, specifically for Emergency Management and Response, which have affected our design principles.

A. ***Evolution of an end-to-end design****:* The continuum of emergencies starts with the occurrence of the emergency or early warning systems (depending on predictors being set up), the assimilation of information and data pertaining to various aspects of the emergency, the prevention of the emergency if possible and responding to the emergency. Depending on the nature of the emergency, a variety of tools and techniques might render themselves relevant. The emergencies themselves could belong to a wide category including natural hazards, civil emergencies, medical emergencies, epidemics and large-scale terrorist attacks. The parts of the continuum in which the sensors are used are also varied. For example, certain sensors are used to detect the occurrence of an emergency. Other sensors might be used to assimilate data and information about the emergency, requiring them to have a higher hardware specification, when compared to simple sensors. Sensors can also be used to simulate emergencies, issue early warnings to people affected by the emergency and mitigate emergencies. Fig.1shows the continuum of an emergency situation.

FIGURE 1. **Emergency Continuum.** Different phases exist within emergency management including occurrence, surveillance, prevention and response.

B. ***Ability to incorporate a number of governmental and non-governmental entities and policies:*** Some observations on recent disaster statistics including the Leh Flash Flodds, the Christchurch earthquake [] and an older example in Hurricane Katrina [] indicate that governmental response agencies co-ordinate response activities with civilian groups to mitigate the situation. Therefore, it is important for any simulation tool to be able to incorporate both governmental and non-governmental entities and policies in its framework. Related work [HLA Based] proposes methods to simulate tasks within an EOC while adhering to IEEE standards on distributed simulation systems. Our work keeps with this theme of separating the run-time architecture into Software Simulations, Live components and data-viewers while extending the architecture to be able to make decisions based on interactions with non-governmental agencies and common citizens via proactive crowd-sourcing.

C. ***Efficient, inter-operable communication models****:* One of the earliest limitations identified with first-responder communications was that they were largely outdated and exposed a very big weakness in operational efficiency of these teams, especially in the case of large-scale disasters. In the US alone, there are approximately 50,000 local first responder agencies, a large number of which struggle with establishing communication channels with each other. Project 25 was one of the earlier efforts that tried to integrate groups working in the domain of Public Safety to identify issues with their devices and channels to contact each other and co-ordinate response activities. With consistent effort, P25 has evolved to assessing the availability of reliable wireless communication for these responders and introduced a variety of standards including Land Mobile Radio (LMR), services for local, state and federal public safety agencies [7]. A number of commercial service providers have also stepped in to take the lead in addressing this problem including Cyren Call, the First Response Coalition and Frontline Wireless. A good simulation test-bed must take into account varying connectivity levels in order to realistically model co-ordination between distributed teams.

D. ***Ability to change response parameters with live feedback****:* The inherent nature of an Emergency is dynamism and any good simulation test-bed, whether it is used for training personnel or evaluating response mechanisms, should take into account the fact that assumptions about the parameters change, with time. For example, a resource might be trying to traverse a path which crosses a bridge and if that bridge falls down during the emergency then that path is no longer valid, when simulating the approach of the resource dispatched for that emergency.

III. SIMULATORS FOR EMERGENCY MANAGEMENT

Our initial approach to simulating Emergency Response was captured in a somewhat static framework to demonstrate user-mobility at the time of an Emergency, in a prototype CAVIAR. This design was very limited for a variety of reasons, including the infeasibility of adding new information in a layered fashion, defining boundaries of various administrative domains and encoding the notion of *time* in a more intuitive manner. In order to understand user-mobility and plot where the cell-phone tower resources were located, we initially used OpenSceneGraph along with static terrain maps of the South Island. The system to map user-mobility in our initial simple visualization of call loads and user movement is shown in Fig. 2.

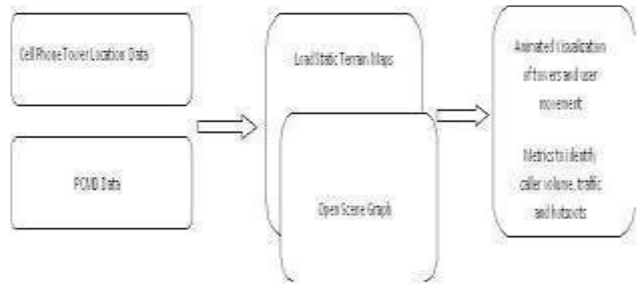


FIGURE 2. **Static Visualization - CAVIAR.** Our Initial Prototype included a lot of static components such as terrain maps, only two sources of data (tower location and PCMD) and produced animated visualizations of user-mobility and tower-load.

We evolved our initial design with CAVIAR to using a modular approach in building a simulation test-bed where all the characteristics listed in Section II are taken into consideration, including:

- Event Simulators for simulating the occurrence of geographically distributed disasters.
- Resource Simulators which place various resources such as fire-trucks, ambulances, etc. at their appropriate locations.
- Crowd Simulators which position callers on a cell-phone network at their location during the occurrence of the emergency.
- Incorporation of agents for decision making. and
- A run-time framework that is dynamic in accepting live caller feed-back on the changing nature of the emergency.

These elements are constituted in our prototype called *Reverse 111* which is a part of a larger test-bed for Emergency Management Simulations. The overview of the test-bed is shown in Fig. 3. *Reverse 111* has custom menus to simulate emergencies *in any part of the globe* wherein the NASA World Wind virtual globe is an integral part of our system. The NASA World Wind virtual globe is part of an open-source initiative where USGS satellite imagery is overlaid with several maps from ariel photography and other publicly available GIS data on 3D models of the earth and other planets. Several add-ons are available as part of this software including point layers, trial layers, line and polygon features, model features, place names, image layers and scripts.

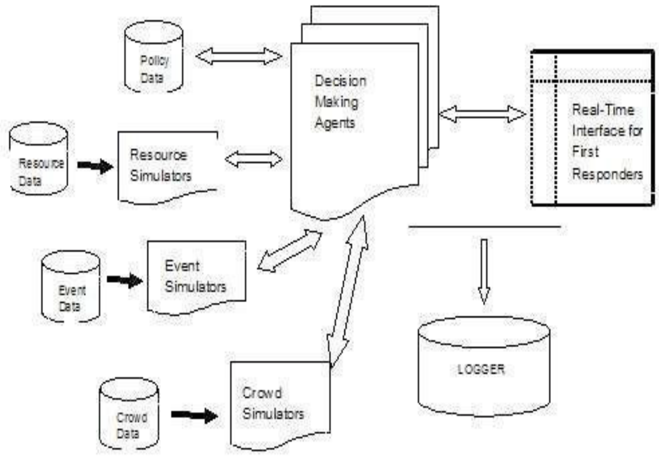


FIGURE 3. **Emergency Response Test-Bed** This figure shows the modules in our test-bed with emphasis on incorporating various sources of data and information.

In our architecture, we have special configuration files to simulate various scenarios within the emergency management framework. All emergencies are not created equal and for example, an earthquake will lead to a clustering of emergency events along a fault line, vs. a bio-weapon attack will have a different distribution of events as compared to a tsunami, where the events will be clustered along the shore. Additionally, we are able to dynamically configure *invalid paths* such as resources being unable to traverse over water bodies or gorges or roads that are blocked as a consequence of the event. Since these configurations are *dynamic* we enable constant feedback (for example from mobile phone users that are being proactively crowd-sourced for information) to modify the parameters within which our resource scheduling and planning has to work.

A. Event Selection

In our initial prototype with CAVIAR, event selection was performed by polling the busiest traffic days, as observed by the telephony service providers, Telecom New Zealand, one of the largest in the country. By characterizing call traffic and user-movement over the course of a special-event day, we successfully compared how call traffic differs between a normal day in the week versus a day on which an All-Blacks Rugby game was played in Christchurch. The data-sets used for the examples shown in Table1. The data was collected in hour-long blocks for our visualization.

TABLE 7-1. **Data-Sets used to show call-traffic.** we compare traffic over 24 hours, on two days, a normal day and the day and All-Blacks Rugby game is played in Christchurch.

Trace	**Date**	**Event**	**Sample Size**
One day dataset (Christchurch)	22nd January 2008	Norma l Day	1272184
One day dataset (Christchurch)	21st June 2008	All Blacks Rugby Game	1667914

Fig. 4 shows the call traffic within Christchurch, from the first data-set shown in Table 1. The data is broken up into hour-long samples, for one whole day, January 22nd, 2008. This is a normal weekday in the city with no special event occurring anywhere to create anomalous traffic patterns. Each tile represents an hour, starting at 8am with 12 tiles being displayed for 12 hours of the day from 8am to 7pm. The visualization is straightforward in that the density of calls is depicted by the size and the brightness of the circle depicting that statistic, which is placed over that particular cell-tower, in the city.

FIGURE 4. **Normal Day Call Traffic.** Call traffic is depicted within Christchurch on January 22nd, 2008, a normal day, over 12 hours.

As can be seen from Fig. 4, at 8am, the city wakes up and from 9am-5pm,the city is most active, which pertains to the work-hours of people residing in the city. Also seen is the concentration of calls in the cell towers closers to the city centre, which is starting to brighten (depicting greater number of calls) at around 9am, which is the second tile in the topmost-row of tiles in Fig. 4. At around lunch hour 12pm-1pm, we see that the traffic starts to move to locations outside the city, possibly commuting to get lunch at spots outside the city-centre. This is depicted in the third tile in the second row. After about 5pm 5:30pm the calls drop off as night-time approaches.

FIGURE 5. **Game Day Call Traffic.** Call traffic is depicted within Christchurch on June 21st, 2008, when an All-Blacks Rugby game was played, over 12 hours.

Fig. 5 shows the data collected for calls made on June 21st, 2008, corresponding to the date of an All Blacks game vs. England. The difference in call traffic here depicts how the city behaves during a holiday or during a special occasion. At 9am, the city is awake but not buzzing, as on a regular weekday, which was January 22nd, 2008 (a Tuesday). Furthermore, three is no migration of call traffic because most users have decided where they are going to be, at the time of the game, so the intensity stays static between tiles 3-8, depicting the hours of 10am-3pm. At the time of the game being played, 7:35pm was the telecast time, depicted in Fig. 5, in tile 1 in row 1 and for a few hours preceding that, depicted in Fig. 7-7, tiles 1-3 in row 4 (showing 5pm-7pm), the call traffic dies out as people are more focused on the game. This, in contrast to the same tiles 1-3 in row 4 (showing 5pm-7pm) in Fig. 4, is much less call traffic. In Fig. 5, between the hours of 5pm-7pm on a weekday, the city was still pretty awake and call traffic depicted that. we have co-related this observation to other All-Blacks games played in Wellington and Auckland but do not include these for length restrictions.

B. Event Simulators

In *Reverse 111, e*vents refer to emergencies that are occurring in an area, that are represented in our system by means of icons appearing at certain latitude, longitude combinations over a geographic area. The icons chosen for these represent the resource requirements for these events. There are four different types of resources an event could need, and this is indicated by different colors. Figure 6 shows the iconic depiction of an event in our system. Examples of these include fire-trucks, medical ambulances, police-vehicles or military support. In Reverse 111, an event does not have fixed resource requirements.

FIGURE 6. **Event Depiction.**We show the iconic depiction of an event in this figure wherein each color stands for a different *type* of resource that an emergency event might require.

The attributes associated with each event are outlined below:

Event Location refers to the actual physical location of the event;

Event Static Resource Requirements describes the static resource requirements that are assigned upon the creation of every event;

Event Dynamic Resource Requirements describe the changing nature of an event with the progress of time. While I borrow from the RimSim Response depiction, I extend an event's resource requirements to be more *dynamic*. In the original RimSim Response, an event is static in its resource requirements and this is not a realistic depiction of any emergency, where the needs are constantly changing over time. Additionally, the objective of Reverse 111 is not training (as was the case with RimSim Response), rather it is to make a case for obtaining better information from citizens or cell-phone users that are co-located with an emergency. Therefore, our system needs to model an emergency event more realistically in order to support end-to-end evaluation;

Event Changing Access Parameters describes how the areas surrounding an event might change over time, given that our initial assumptions about an emergency may not be valid as time goes on.

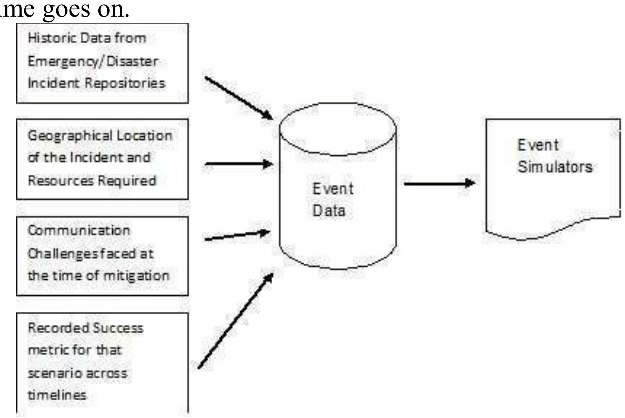


FIGURE 7. **Event Simulators.** Several data-sources are used as inputs to the Event Simulators including historic data from first responder repositories, in our test-bed.

Events occur at pre-determined locations or according to various pre-simulated scenarios. Fig. 8 shows the occurrence of three different emergencies within Seattle, including an earthquake, a bio-weapon attack and a tsunami, which have different distributions and groupings of events along the geographic spread of the region.

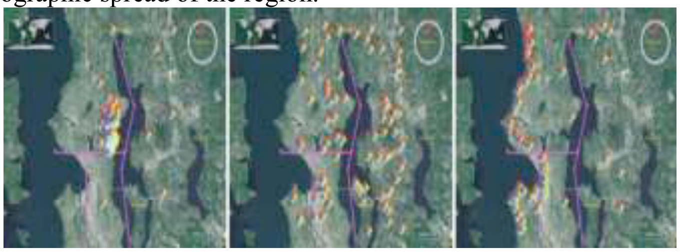


FIGURE 8. **Different Event Distributions.** Three different types of events including a bio-weapon attack, an earthquake and a tsunami show different occurrence and grouping of incidents in the greater Seattle area.

C. Resource Simulators

CAVIAR had no resource simulators. In *Reverse 111,* Resources refer to the resources that are applied in order to ameliorate the emergencies. Fig. 9 shows the types of resource used in Reverse 111. Examples include various state services and every resource has a number of attributes detailed below:

Resource Location refers to the geographical location of a resource. For example, a fire-truck might be at a fire-station when the emergency event occurs or might be returning to its home-base and therefore at a different location at the absolute time-stamp when its required.

Resource State indicates whether the resource is stationary or is in motion towards the event of interest.

Resource Owner refers to which administrative domain owns the resource. These could include state bodies such as the fire-department, police, military and can also be extended to include *commercial* emergency responders, which motivates our lottery-scheduler to evaluate a heterogeneity in such resource ownership.

FIGURE 9. **Resource Depiction.** Physical resources are static and decision-making resources are more dynamic in our system.

Resource Type refers to whether the resource is a static resource (such as the ones described before) or a more dynamic resource, such as a user providing images of the façade of the building that has fallen down. The latter is more information than an actual physical resource and can be used for decision making. This is separate from the physical resources but an integral component of Reverse 111.

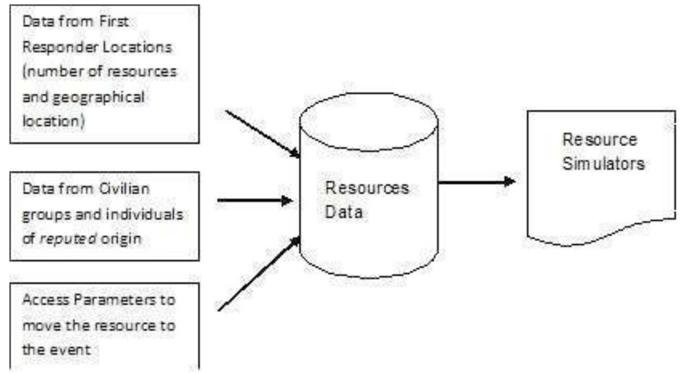


FIGURE 10. **Resource Simulators.** Physical resources are static and decision-making resources are more dynamic in our system.

D. Crowd Selection

Utilizing our first attempt at the design, we come up with a simple visualization of the cell phone towers (shown in Fig. 11 tower-statistics (shown in Fig. 12) and user-mobility (shown in Fig. 13) within CAVIAR.

FIGURE 11. **Cell Phone Towers.** We show the location of cell-phone towers in this figure, from our simple visualization tool, CAVIAR.

Cell phone tower locations are contained in their own data set outside of PCMD. We are provided with unique identifiers for each of the cell phone towers and a (latitude, longitude) for each tower. We convert these into (Easting, Northing) for each tower. Easting and Northing are terms to describe what are essentially geographic Cartesian pairs for any point. Taking a base map of New Zealand, we plot these points on that map in order to depict the location of the cell-phone towers. It is easily observed that the North Island has a lot more towers than the South Island, owing to a larger subscriber base. Added to this, the grouping of towers is also indicative of the underlying terrain, for example, in the South Island, there are not too many towers in the regions housing the Southern Alps. This is not obvious from this visualization, however as there are no terrain depictions of the region.

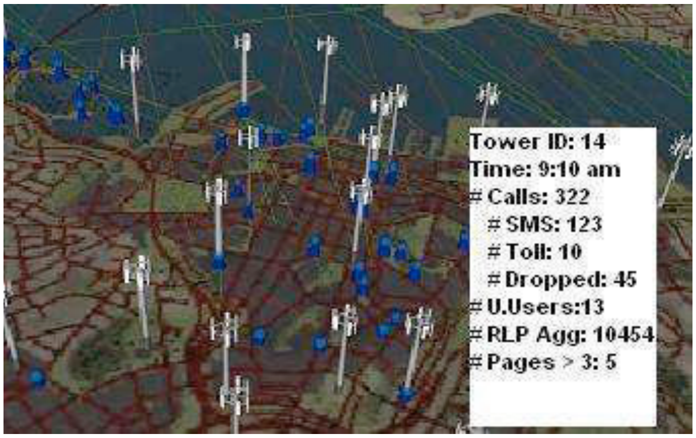


FIGURE 12. **Tower Statistics.** Cell Phone towers are the white structures, each of the users around the tower being depicted in blue. On rolling over a tower, simple statistics about the tower such as the Tower ID, the Time at which the query is being processed and other data related to the tower is displayed.

In CAVIAR, we depict phone subscribers with blue icons and enable basic zoom and roll-over functions and Fig. 12 shows the zoomed-in view where the user has rolled over a tower and the basic tower statistics are displayed to the user. In our particular example, the statistics we show the user (in this case, an emergency responder or a network capacity planner) such variables as the Unique ID of the tower, the time-stamp which the data-shown pertains to, the number of calls made, the different kinds of calls made (SMS, Toll calls, dropped calls), the number of unique users whose calls originated on the tower, the number of paging requests seen (which could provide insight into security violations on a tower) etc. The proximty of the subscriber to the tower is calculated using the call duration and the call start time and the ending cell tower. This is a static snapshot at 9:10am and does not take into account how the user is moving. Further, the orientation of the user is approximated based on movement and if the user is static, we can only estimate the distance from the center of the cell-tower.

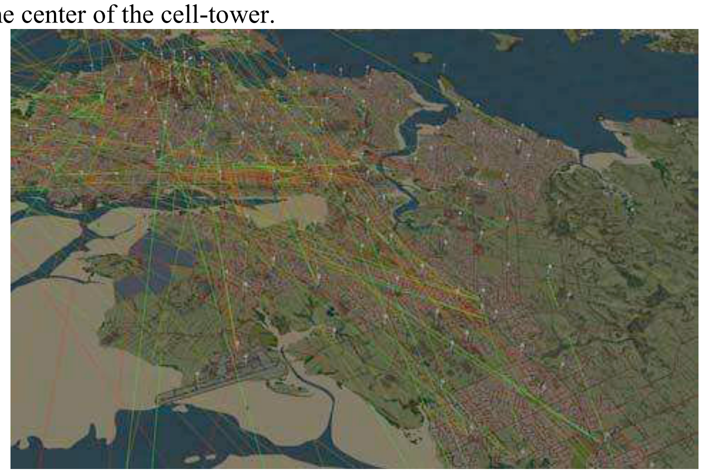


FIGURE 13. **User-Mobility.** Cell Phone users are shown moving at different rates (green being slowest, red being fastest and yellow and orange representing the range in between). This visualization is not effective in encoding time.

Fig. 13 shows user-mobility between cell-phone towers. In this example, we have filtered out the users that are moving between more than 2 cell-towers in the course of their conversation. By further filtering the data to only include a few *minutes* worth of calls, we are able to depict the users moving, along a certain trajectory (shown by green, orange and yellow lines) on the map. A user's velocity is calculated based on the starting and ending tower for a particular call, the length of the call and the location of the starting and ending towers. We depict the *shortest path* that a user can follow, during the course of the call. This is for purposes of visualization only as the user might well take a circuitous route. Added to this, our depiction of the user's movement does not address some of the infeasible paths such as traversing what are obviously water-bodies (shown in the top left-hand corner where we see users freely zooming over Lake Taupo while placing a phone call). For the initial studies on user-mobility, this depiction is sufficient for estimating call traffic and call volume as we will observe in coming examples.

E. Crowd Simulators

In Reverse 111, crowds are simulated by extending the crowd selection in CAVIAR to further incorporate the fourth design principle of changing response parameters based on live feedback. Crowds are introduced as a layer in Reverse 111 wherein the system is able to *co-locate* these crowds with the emergency event being simulated and poll them for more information with respect to the emergency. *Proactive Crowd Sourcing* involves the following steps:

1) *Identifying co-located callers* that are closest to an emergency event. In our system callers are identified by the cell tower from which their call originates at a time closest to the occurrence of an emergency. A group of callers located within a radius of 20-30km around the emergency area are similarly identified with the intention of using information that they may provide to decide on how best to mitigate the emergency.

2) *Calling co-located callers* where the Emergency Response Center contacts the crowd identified in Step 1 and polls them for information (potentially voice confirmation of the emergency, an SMS on how things are unfolding, a picture to show the changing nature of access paths to get resources to an emergency etc.).

3) *Assimilating relevant information* in various media types and processing the same for more accurate response to the situation at hand.

Reverse 111 further has the ability to avoid *invalid paths* assuming that an access bridge has fallen down during the time the response center springs into action for both resources and the crowd, unlike our initial work with CAVIAR where the *shortest path* was used to calculate how users move.

IV. A FLEXIBLE TEST BED FOR EMERGENCY RESPONSE SIMULATION

Within our test-bed for emergency response simulation, we have placed emphasis on adhering to the four design principles outlined in Section II. This prototype test-bed has been developed as a massive online multi-player game, in order to involve more than one player (or response policy) when simulating emergencies. In this section, we describe our modular implementation wherein:

1) *Evolution of end-to-end* design is achived by evolving the toold from a static framework within CAVIAR to a more dynamic framework within Reverse 111.

2) *Ability to incorporate a number of governmental and non-governmental entities and policies* is achieved by first demarcating the region in which the simulations are performed into various administrative boundaries (such as towns or muncipalities, in our case into the cities of Christchurch, Rangiora, Oxford and Ashburton in one example on the South Island). This is followed by drawing fences to ensure that resource-sharing across these administrative domains is governed by local-policies. The third step is to introduce a number of *agents* for resource scheduling including a greedy agent [], a lottery agent [] and a round-robin agent [] to simulate the interaction between governmental and non-governmental entities that might be pooling their resources and information to mitigate the emergency.

3) *Efficient, inter-operable communication models* are simulated by introducing cell-phone callers from Telecom New Zealand as an extra layer in the system. Since these work on the same frequency and provide broadband facilities where needed, we rely on a mode of communication that can be transparent across response teams. Within callers, there are high-priority callers (such as response teams) and low-priority callers (such as citizens trying to contact each other) and this stratification helps achieve channeling the important communication to the respective parties during emergencies.

4) *Ability to change response parameters with live feedback* by introducing proactive crowd-sourcing where we identify callers co-located with the emergency and the simulator introduces delays to simulate the Emergency Response Center contacting the co-located caller and receiving information from them to assess the best response mechanism.

.

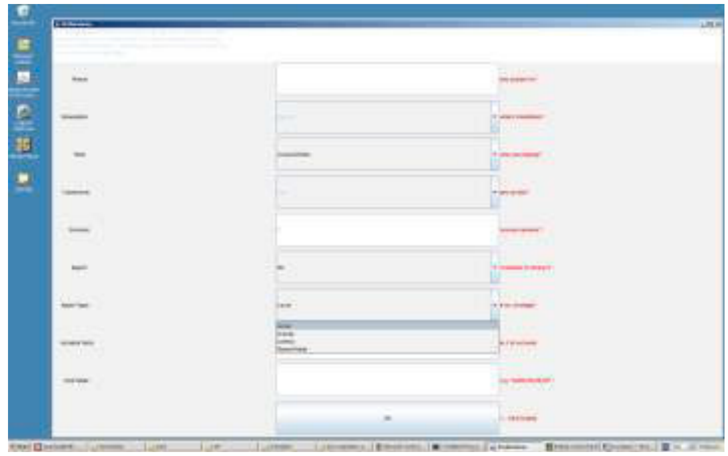


FIGURE 14. **Opening Screen.** Users of this framework are able to configure a lot of system parameters on the fly including roles assumed, playing with other online players, recording the history of their session, playing as human players or invoking scheduling agents that evaluate the effectiveness of various scheduling ideas, controlling the incident rate, etc.

Fig. 14 displays the opening screen of our prototype test-bed showing the parameters that can be dynamically configured in the test-bed starting with the region which the user wants to simulate Emergency Management in, which could include any geographic region in the world. For purposes of example (where emergency situations can assume two different scales, we limit our discussions to New Zealand and the city of Seattle in the state of Washington, in the United States). The prototype further allows user to run the simulation as an *oracle player* who controls all administrative domains (which is useful for baseline comparisons) or as a particular administrative domain (such as the mayor of Christchurch), in order to understand which policies work best, given their resource constraints. There is a further provision to allow multiple-players connected online to play simultaneously, and is a powerful instrument for training first responders. The user's actions and findings can be recorded on a per-session basis, creating a powerful tool for analyzing historical results and comparing these with moving the parameters of the simulation around. Different types of resource-schedulers (or agents) can be evaluated in this tool and in case the user does not want any agent, wants to schedule resources themselves then they can play as themselves, human users. Users are also able to choose the rate of occurrence of emergencies (once more, the emergencies will have a notion of *scale* associated with them and an emergency in India affects a lot more people and occurs quicker and more frequently if wide-spread, as opposed to an emergency in New Zealand) to keep our results realistic, when evaluating various approaches and policies to Emergency Response.

FIGURE 15. **Layered Approach.** This shows the first level view of our tool wherein the panel on the left-hand side allows the user to see the data and information that they need, selectively.

Fig. 15 shows the presentation of the top-level screen as the user enters the starting phase of the simulation. We take a layered approach by presenting the data and information, as the user sees fit, rather than all in one piece, as we did with CAVIAR. The user is notified with messages on the top-left panel and allowed to choose which of resources, incidents and *cell phone data* that they wish to see appearing on the screen. They may choose to see one or the other or some combination of this information, with the progress of time. We also add appropriate labels to indicate the actual location that the user is viewing in the OSM layer, as a reference point. This can be extended to mark administrative domains as well. A layered approach is a direct consequence of our modular framework.

The first level zoom in Reverse 111 shows a lot more information, including the scale (20km) and the Latitude and Longitude where the zoom is occurring, as opposed to CAVIAR. This is important for user-orientation. Furthermore, this visualization environment takes into account the terrain, much better than using static maps, thereby making it easier for the user to identify *valid paths* and administrative domains. In our earlier examples, these were difficult to identify, specifically elevation, but that is shown to a greater detail in this environment. Fig. 16 shows the setting of *administrative domains* demarcated by the magenta lines. In this case, we have Christchurch, Rangiora, Oxford and Ashburton as the four different administrative domains that must interact and co-ordinate resources at the time of an emergency.

In this example we show the introduction of emergency events at an interval of 4 seconds each, indicated by the rhombus-shaped objects, with the four colors standing for one kind of static resource with numbers inside them indicating the resource-requirement. For example, if a rhombus appears with 3-yellow,0-red,1green and 1-blue, that is the resource requirement for that emergency, at the time it is occurring. This is shown in Fig. 16 upon zooming in on the resource. This requirement can *dynamically change* if a co-located cell-phone user, indicates as much back to the system. The resource requirements are important to assess the scheduling policy's action.

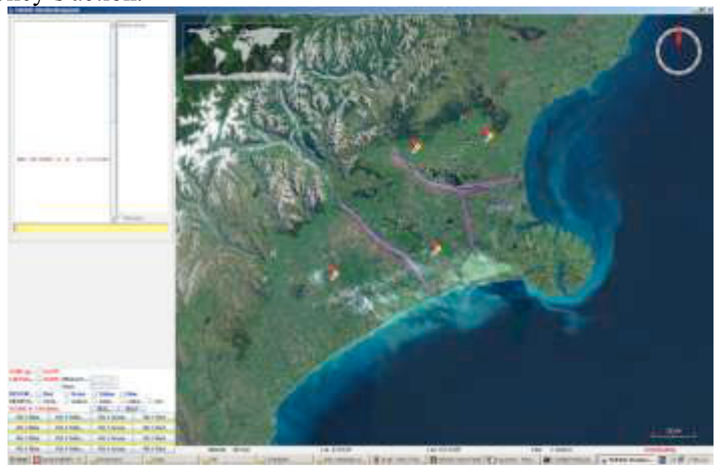


FIGURE 16. **Administrative Domains.** Four administrative domains, Christchurch, Oxford, Rangiora and Ashburton must act in tandem for this simulation situation.

Events occur at pre-determined locations or according to various pre-simulated scenarios.

FIGURE 17. Event Zoom. Upon zooming in on an event, its resource requirements are shown to the user. In this case, the event required 4-blue resources, 3 yellow, 3 green and zero red resources.

While we take into account actual numbers of fire-trucks and other resources that have been reported on public websites accounting for the State Department's response to such activities, we combine this with an intuitive approach to where and how events within an emergency scenario might occur and our visual representation confirms that our placement of events is accurate. Fig. 18 shows the location of resources of a certain kind, when the user selects that resource should be shown on screen. The location of resources also gives the user a visual clue on the performance of our schedulers, when they try to pick the resource that is suited, according to their policy, to send to a particular event. Fig. 19 shows multiple resources (of various colours on the screen, as selected by the user. In this view, the user has also chosen to play as the administrative domain of Christchurch, therefore placing a greater emphasis on that region alone.

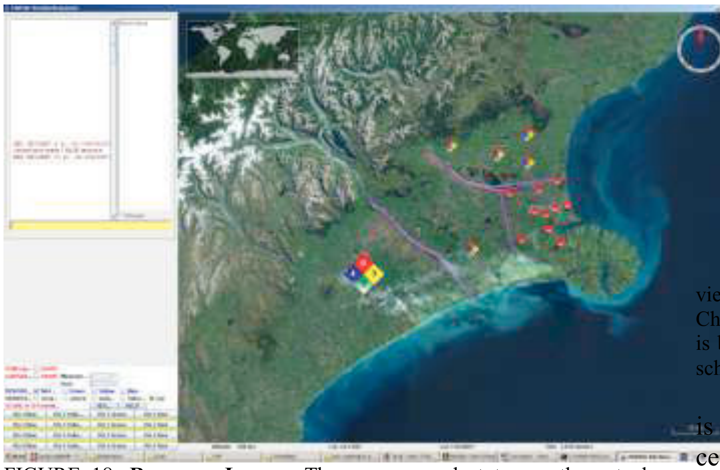


FIGURE 18. **Resource Layers.** The user can select to see the actual physical location of resources, in this case they have chosen to see the location of all the red resources. These resources are applied to Emergency management in this example but can be easily extended to logistics and visualizing Supply chains.

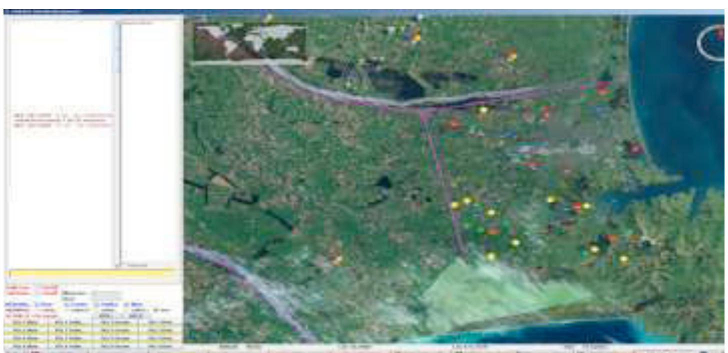


FIGURE 19. **Resource Layers for Christchurch.** The user has selected to view all resources and play as the administrative role of Christchurch, rather than all four domains indicated in earlier figures.

B. Path Selection

In this section, we discuss how paths are selected in order to get a resource to an event that requires it.

FIGURE 20. **Cell-Phone data Christchurch.** The user has selected to view all resources and events, and the cell phone users within Christchurch, indicated by the light blue circles. In this example, an event is being handled, with the white line indicating that a resource has been scheduled for the event that is highlighted at the end of the white line.

Fig. 20 shows the scheduling of a resource for an event that is highlighted at the end of the *path* for that resource and the cell-phone users in the Christchurch area, shown in light blue. Cell phone users are considered *soft resources* that provide *information*. In our prototype, the path selection process goes through the steps described in the flowchart shown in Fig. 21.

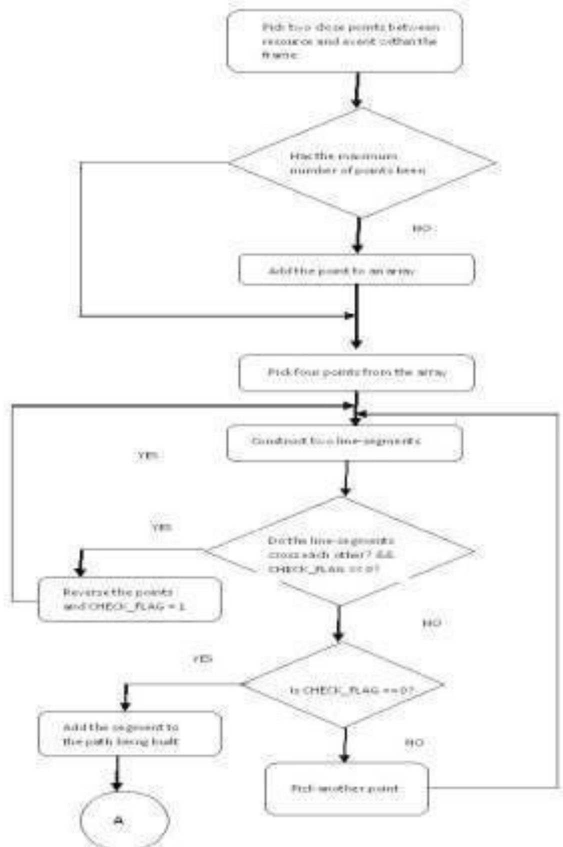


FIGURE 21. **Building a Path.** The system constructs a path in a piece-wise fashion for each valid FRAME, which is only within the region that can be traversed without intervening geographic hindrances like water-bodies, etc.

We draw the reader's attention to the fact that the user sets up the specific regions or *frames* that are valid for the resource to traverse. Our system constructs a piece-wise linear path, for each frame of valid points, located within a sub-region that can be navigated without any obstacles in the resource's way such as water-bodies or gorges (in the case of New Zealand's South Island). Fig. 21 shows the construction of the path, while ensuring that none of the segments cross each other, which would make for a circuitous route for the resource, as it reaches the event of interest.

Fig. 22 shows a co-located user, that is directly on the path being traversed, that can provide some insight into the situation towards which the resource is travelling. The user on this path has *reputation information* associated with them, besides actual physical location. For example, the user is associated with tower #27, which has seen a total call volume of about 1000 calls, 276 from distinct callers. This tower is heavily loaded at the time of the emergency (we are tracking an hour of call-time in this example and filter only users that have not moved across cell-towers, indicating static users and not perpetrators that are fleeing) and it is likely that this user is reputable as the average call time of calls on this tower are about 22seconds, indicating that most calls are being placed to check to see if the called party is safe.

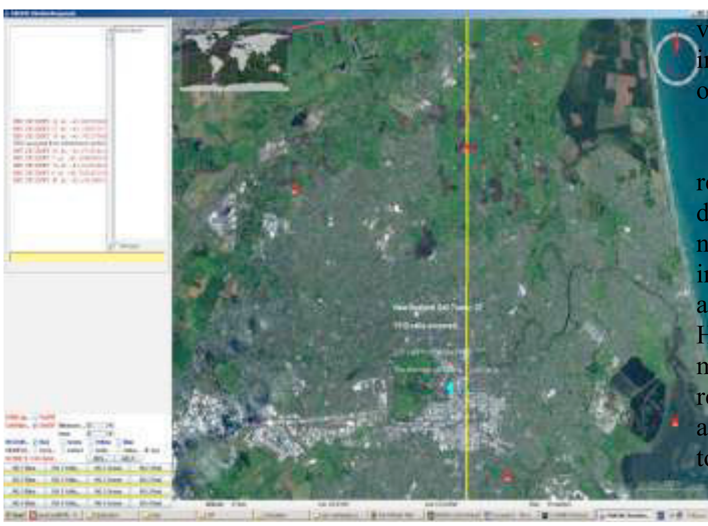


FIGURE 22 **Location + Reputation = Information.** This figure shows obtaining information from a reputed user, co-located with the emergency event and modifying the resource requirements to take that users-input into consideration.

C. Scheduling Agents

In this section, we perform a different evaluation of *four* scheduling agents, which was joint work with the University of Washington. We pick the three scenarios introduced in Fig. 17, for this evaluation, in the greater Seattle region. The scale of emergencies in this region is different, as is our evaluation method, when compared to the result presented in Chapter 5.

In order to simulate multiple first responders within a response scenario, we added an agent class that simulates four different schedulers, greedy, lottery, round-robin and first-fit. In order to encode realistic scheduler behavior, we used a combination of resource-sharing policies, and resource and event selection strategies. These heuristics are easily extensible to encode newer policies and many more scheduler types, either distributed or otherwise.

The emergency response behavior of sub-classed agent types are varied by tweaking three parameters that affect agent behavior: (1) their policy on sharing resources with neighboring administrative domains (for example, does the mayor of Christchurch share with the mayor of Rangiora, etc.), (2) their strategy for determining which active unresolved incident should be attended to next (i.e does the scheduler choose the first event which appears or perform calculations to see which event can be responded to with the closest resources), and (3) their strategy for determining which available resource to assign the targeted incident (i.e, if an event requires 3 red resources, do we get all three to the event in one go or do we phase out the resource delivery).

Each agent takes a turn in effecting the dynamic visualization until the scheduled simulation time ends or all incidents have been handled and no more are scheduled to occur. A turn is calculated as follows:

An agent figures out its resource surplus (how many resources it has free, minus the number of unmet resource demands from incidents in its jurisdiction). We call this number **R** (note: R is occasionally negative).We then interrogate the agent's R-level willingness to help other administrative domains. we call this number **HiWat**. If R > HiWat, the agent looks at the requests for help that have been made by other players (human or agent). If there is an open request and the R > HiWat, the agent chooses a resource according to its Resource Selection Strategy (RSS) and send it to the incident specified in the request. This ends a turn.

If instead R <= HiWat, we interrogate the agent's R-level willingness to ask for help from other administrative domains. we call this number **LoWat**. If R < LoWat, the agent chooses the last incident on our priority list (if there is one; priority as determined by our ISS) and broadcasts a request for help on that incident. This ends a turn.

If instead R >= LoWat, the agent chooses an incident according to its Incident Selection Strategy (ISS) subject to its Resource Sharing Policy (RSP). This RSP provides a primary ordering of the incidents according to jurisdiction, the ISS provides a secondary ordering in case of ties (and usually we find there are many ties). If there are no incidents to choose, this ends a turn.

The agent chooses a resource to send to the incident it has chosen by following its embedded heuristics, which in this case are according to an RSS. The agent sends the resource to the incident it chose in step 4. This ends a turn.

For the study we document in this thesis, the agent behavior-defining parameters we chose to test include:

- **HiWat** - an integer representing willingness to give help when asked.
- **LoWat** - an integer representing willingness to ask for help
- **Resource Sharing Policy** (RSP) - one of 5 alternatives representing willingness to volunteer help.
- **Incident Selection Strategy** (ISS) - one of 4 alternatives representing heuristic for choosing the next incident to apply resources.
- **Resource Selection Strategy** (RSS) - one of 2 alternatives representing the heuristic for choosing the next resource to assign.

D. Resource Sharing Policy

An agent's RSS policy describes under what conditions it will voluntarily respond to incidents outside its given geographic region. Five policies are defined and implemented in agent sub-classes. we refer to the policies by name in order to assist reading comprehension when describing our experiments in a later section:

- Sociopath - Never volunteer aid to another region.
- Selfish - Prefers incidents in its own region, but will volunteer aid to another region if there are no active incidents in its own region.
- Equalitarian - Does not take geographic region into account when determining which incident to handle next.
- Selfless - Prefers to volunteer for incidents in another region, but will handle incidents in its own region if there are no outside incidents to handle.
- Altruist - Never handles its own incidents, but will always volunteer for incidents outside its region.

E. Incident Selection Strategy

Within the broader resource-sharing policy, there is still the question of which incident to handle first, since there might be many active incidents within a single geographic region at any given time. Which incident is selected in the end depends first on policy and then on incident selection strategy. we implemented four representative incident selection strategies for our experiments:

- First Fit - Chooses the incident with the lowest incident number regardless of other considerations. Computationally, this is far the simplest of the strategies.
- Round Robin - Chooses the incident that has been active the longest.
- Lottery - Gives each incident a number of tickets equal to the total number of resources it requires, and chooses a ticket at random. The incident holding the winning ticket is selected.
- Greedy - Considers the resources that would have to be applied to each incident, and chooses the incident that could be handled most quickly (that is, on the basis of the furthest required resource).

F. Resource Selection Strategy

Once an incident has been identified, an agent must choose resources to assign to that incident in order to assist in incident resolution. There are likewise many possible strategies for choosing between resources to assign. For the demonstration purposes of this thesis, we encoded two resource selection strategies:

First Fit - Chooses the free resource with the lowest resource number.

Closest - Chooses the free resources closest to the incident.

G. Agent Types

The agent types that we use to demonstrate our software use are a combination of the elements above, particularly the two selection strategies, since the policy can be supplied as a parameter for every agent type:

- LocalAgent: First Fit incident selection, First Fit resource selection.
- RoundRobinAgent: Round Robin incident, Closest resource.
- LotteryAgent: Lottery incident, Closest resource.
- GreedyAgent: Greedy incident, Closest resource.

V. Simulation Runs

Each of the scenarios is divided into the same four geographic areas to represent administrative domains for the emergency response agents. we varied all possible 160 agent characteristic mixes across all four geographic regions to create 640 runs for each scenario. we watched each of the 640 simulations take place for each scenario and noticed many interesting patterns of resource allocations including obvious inefficient motions for inferior characteristic sets. The response activity across the two major bridges was especially interesting to watch as were the clustering of movements at times between administrative domains. Figure 23looks at the time to completion for each simulation across all three scenarios. Scenario completion times include the area from the top of each scenario area down to 0 on the x-axis. In all but twelve of the 640 runs, the tsunami scenario took less time to mitigate than the others (in twelve runs the completion time was identical to the earthquake scenario). The earthquake scenario took less time than the bioweapon scenario in all cases (but was much closer in the most efficient cases). The distribution of response times gives a sense of how significant agent behavior mix was to the completion time of the scenario. The more isolated the community event to one jurisdiction, the more impact agent behavior makes on completion time.

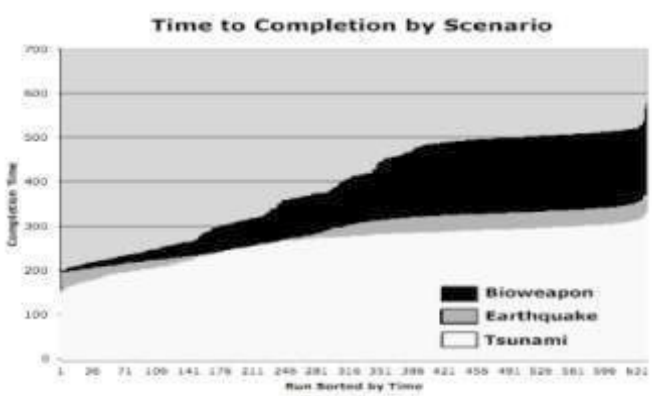


FIGURE 23 **Seattle results.** This figure shows our running various agents within the greater Seattle region. we do not take into account co-located callers in this example.

VI. Conclusions

CAVIAR was our initial prototype, which was very static in its definition of events, resources and crowds. We mapped caller traces obtained from Telecom New Zealand's network wherein the user's movement was tracked across cell-towers at which the calls originated and terminated. Events were selected across special event days such as Rugby Game days where the call-traffic patterns and user-movement patterns changed (and demonstrated crowding at certain locations downtown or close to the stadium). We evolved this prototype to accommodate four guiding design principles for efficient design of a test-bed for Emergency Management Simulation including the incorporation of an end-to-end design, incorporation of a number of governmental and non-governmental entities and policies, efficient, inter-operable communication between first responders and the ability to change response parameters with live feedback. We evaluated the performance of our prototype using 640 simulation runs and presented these results to evaluate various approaches to live emergency response.

Acknowledgment

We acknowledge the work of Konrad Schroder on help with building agents and enabling us to test these effectively.

References

1. Ito, A., Murakami, H., Watanabe, Y., Fujii, M., Yabe, T., Haraguchi, Y., Tomoy-asu, Y., Kakuda, Y., Ohta, T., Hiramatsu, Y.: ***Universal use of information delivery and display system using ad hoc network for deaf people in times of disaster***. (2008) 486–491
2. Wu, W., Cao, J., Zheng, Y., Zheng, Y.P.: ***Waiter: A wearable personal healthcare and emergency aid system.*** (2008) 680–685
3. Lee, R.G., Chen, K.C., Hsiao, C.C., Tseng, C.L***.: A mobile care system with alert mechanism. Information Technology in Biomedicine***, IEEE Transactions on 11(5) (2007) 507–517
4. Inoue, Y., Sashima, A., Ikeda, T., Kurumatani, K.: ***Indoor emergency evacuation service on autonomous navigation system using mobile phone***. (2008) 79–85
5. Filippoupolitis, A., Hey, L., Loukas, G., Gelenbe, E., Timotheou, S.: ***Emergency response simulation using wireless sensor networks.*** In: Ambi-Sys '08: Proceedings of the 1st international conference on Ambient media and systems, ICST, Brussels, Belgium, Belgium, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering) (2008)
6. Chitumalla, P., Harris, D., Thuraisingham, B., Khan, L.: ***Emergency response applications: Dynamic plume modeling and real-time routing***. Internet Computing, IEEE 12(1) (2008) 38–44
7. Basu, S.N.: Fermi 2 emergency response information system. Energy conversion, ieee transactions on EC-1(2) (1986) 16–22
8. Serpa, D.P., Walker, D.M., Jenckes, T.A.: ***Emergency monitoring, assessment and response system for diablo canyon nuclear power plant***. Nuclear Science, IEEE Transactions on 28(1) (1981) 236–241
9. Burnett, B., Bowering, G., Hagelstrom, R., Tate, D., Ramanujan, R.: ***Multi-radio nodes for mobile mesh networks for emergency response and battle command systems***. (2007) 1–7
10. Symons, L., Pavia, R., Hodges, M.: ***Emergency response in national marine sanctuaries***. (2005) 345–351 Vol. 1
11. Gertman, D.I., Bruemmer, D.J., Scott Hartley, R.: ***Improving emergency response and human- robotic performance.*** (2007) 334–340
12. Bansal, A., Patel, K., Gupta, G., Raghavachari, B., Harris, E., Staves, J.: ***Towards intelligent services: a case study in chemical emergency response.*** (2005) –758
13. Kung, H.Y., Ku, H.H., Wu, C.I., Lin, C.Y.: ***Intelligent and situation-aware pervasive system to support debris-flow disaster prediction and alerting in Taiwan***. J. Netw. Comput. Appl. 31(1) (2008) 1–18
14. Miyauchi, K., Yonezawa, Y., Ogawa, H., Maki, H., Caldwell, W.: ***A mobile phone-based safety and life support system for elderly people.*** (2005) 81–84
15. Pries, R., Hobfeld, T., Tran-Gia, P.: ***On the suitability of the short message service for emergency warning systems.*** Volume 2. (2006) 991–995
16. Drane, C., Macnaughtan, M., Scott, C.: ***Positioning gsm telephones.*** Communications Magazine, IEEE 36(4) (1998) 46–54, 59
17. Scott Shane, E.L.: ***Stumbling storm-aid effort put tons of ice on trips to nowhere***. The New York Times (2005)
18. Campbell, B., Schroder, K. (2009) Training for emergency response with ***RimSim:Response!***, In Proceedings for the Defense, Security, and Sensing 2009 Conference (Orlando, FL).

19. Filippoupolitis, A., Hey, L., Loukas, G., Gelenbe, E., Timotheou, S.: **Emergency response simulation using wireless sensor networks**. In: Ambi-Sys '08: Proceedings of the 1st international conference on Ambient media and systems, ICST, Brussels, Belgium, Belgium, ICST (2008)
20. Chitumalla, P., Harris, D., Thuraisingham, B., Khan, L.: ***Emergency response applications: Dynamic plume modeling and real-time routing.*** Internet Computing, IEEE 12(1) (2008) 38–44
21. Carl A. Waldspurger. Lottery and Stride Scheduling: Flexible Proportional-Share Resource Management, Ph.D. dissertation, Massachusetts Institute of Technology, September 1995.
22. Yishay Mansour , Boaz Patt-Shamir, ***Greedy Packet Scheduling on Shortest Paths*** (1991), Journal of Algorithms
23. Rubén Ruiz, Thomas Stützle, ***A simple and effective iterated greedy algorithm for the permutation flowshop scheduling problem*** (2006), European Journal of Operational Research
24. Linehan et. Al., ***Developing a serious game to evaluate and train group decision making skills,*** Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era, P106-113, 2009.
25. Fortes, Macintosh, ***Effective citizen relationship management: Hurricane Wilma and Miami-Dade County 311***, Proceedings of the 2006 international conference on Digital government research , Pages: 452 - 453 , 2006 , San Diego, California
26. Artigas et. Al, ***A regional approach to geographic information sharing in northern New Jersey,*** USA, Proceedings of the 2008 international conference on Digital government research, Montreal, Canada, P: 132-136, 2008.
27. L. Frishberg, ***Looking back at plan AHEAD™: exercising user-centered design in emergency management,*** Conference on Human Factors in Computing Systems, Portland, OR, 2005
28. Boukerche, et. Al., ***Web-Based Distributed Simulations Visualization and Control with HLA and Web Services,*** Proceedings of the 2008 12th IEEE/ACM International Symposium on Distributed Simulation and Real-Time Applications, P 17-23, 2008
29. R. Woltjer et. Al, ***Role-playing exercises to strengthen the resilience of command and control systems,*** Proceedings of the 13th Eurpoean conference on Cognitive ergonomics: trust and control in complex socio-technical systems, ACM International Conference Proceeding Series; Vol. 250 P71-78, 2006.
30. Artigas et. Al., ***Real-time ocean surge warning system, meadowlands district of New Jersey,*** Proceedings of the 10th Annual International Conference on Digital Government Research: Social Networks: Making Connections between Citizens, Data and Government, P216-222, 2009.
31. D. Tarchi, et. Al., ***The communication infrastructure for emergency management: the In.Sy.Eme. vision,*** Proceedings of the 2009 International Conference on Wireless Communications and Mobile Computing: Connecting the World Wirelessly, Leipzig, Germany 2009.
32. T. Clark, C. Waring***, A simulation approach to analysis of emergency services and trauma center management,*** Proceedings of the 19th conference on Winter simulation, Atlanta, Georgia, United States, 1987.
33. W. Bostian et. Al., ***Testbed for High-Speed 'End-to-End' Communications in Support of Comprehensive Emergency Management,*** Proceedings of the 2004 annual national conference on Digital government research, Seattle, WA.
34. L. Paylen et. Al., ***A vision for technology-mediated support for public participation & assistance in mass emergencies & disasters***, Proceedings of the 2010 ACM-BCS Visions of Computer Science Conference, Edinburgh, United Kingdom.
35. Mendonca et. Al., ***Collaborative adhocracies and mix-and-match technologies in emergency management,*** Communications of the ACM, SPECIAL SECTION: Emergency response information systems: emerging trends and technologies,P44-47, 2007.
36. N Adam et. Al, ***Secure information sharing and analysis for effective emergency management***, Proceedings of the 2008 international conference on Digital government research, Montreal, Canada.
37. Wu, X. Zhang, ***Supporting collaborative sensemaking in map-based emergency management and planning,*** Proceedings of the ACM 2009 international conference on Supporting group work, Sanibel Island, Florida, USA.
38. Riberio, A. Ferworn, ***Computational public safety in emergency management communications,*** Proceedings of the 6th International Wireless Communications and Mobile Computing Conference, Caen, France, 2010.
39. N. M. Patrikalakis, J. J. McCarthy, et al***., "Towards a Dynamic Data Driven System for Rapid Adaptive Interdisciplinary Ocean Forecasting,"*** In Dynamic Data Driven Application Systems, F. Darema, ed., Kluwer Academic Publishers, Netherlands, 2004.
40. Vaidyanathan, A., Johnston, D. (in press). ***The use of cell-phone technology for improved disaster response: A case-study from New Zealand.*** The Australasian Journal of Disaster and Trauma Studies.
41. Vaidyanathan, A. ***"Emergency Response with Proactive Crowd-Sourcing—Reverse-111",*** 5th International Conference on Critical Response Information Infrastructure Security (CRITIS), Athens, Greece, September 2010

AUTHORS PROFILE

Dr. Anu Vaidyanathan is Visiting Faculty at the Department of Computer Science and Engineering at IIT Ropar. Prior to joining IIT Ropar, Anu was the CEO of PatNMarks, the first Indian company to offer Value Added Services in the area of Intellectual Property. Anu graduated with a B.S in Computer Engineering from Purdue University in 2001. After getting a Masters in Computer Engineering from North Carolina in 2002 and a brief stint at IBM Austin, Anu finished her PhD in Electrical and Computer Engineering at the University of Canterbury at Christchurch, graduating in a record time of 26 months

Emerging Trends of Ubiquitous Computing

Prakriti Trivedi[1], Kamal Kishore Sagar[2], Vernon[3]

[1]HOD of CE &IT Department

[1, 2, 3] Govt. Engineering College, Ajmer
Rajasthan Technical University, Kota , India

[1]hod_ceit@rediffmail.com

[2]kamal_sa1234@yahoo.co.in

[3]vernon86@rediffmail.com

***Abstract*— Ubiquitous computing is a method of enhancing computer use by making many computers available throughout the physical environment, but making them effectively invisible to the user. Background network to support ubiquitous computing is ubiquitous network by which users can enjoy network services whenever and wherever they want (home, office, outdoors). In this paper Issues related to ubiquitous network, smart objects and wide area ubiquitous Networks have been discussed. We also discuss various elements used in ubiquitous computing with the challenges in this computing environment**

***Keywords*— Ubiquitous, sensor, network, wireless, computing**

I. Introduction

Ubiquitous Computing has potential applications in several sectors such as healthcare, Business processes, disaster management, farmland irrigation and empowering the common Man to improve the quality of life. Ubiquitous computing is a method of enhancing Computer use by making many computers available throughout the physical environment, But making them effectively invisible to the user. Disappearing computer means the Functionalities of computers will be moved to the surroundings. As technology become more embedded and invisible, it calms our lives by removing the annoyances. Ubiquitous Communication is based on the concept of ubiquitous computing, where technology recedes into background of our lives to make human computer interaction much easier. They require efficient, multimedia and power-aware technologies linking together many Heterogeneous devices distributed over small or large specifications and pear-to-pear and ad hoc paradigms. Ubiquitous communications are intended to connect and transmit/distribute the information among number of computing devices that form a ubiquitous network. In ubiquitous computing the new computing devices usually equipped with a selection of different sensors to collect data from their environment. The goal is to create context awareness, which allows intelligent things to decide and act on a decentralized basis. The new computing devices are mobile, and the tasks they are programmed to perform depend on the geographical location and neighbourhood of the devices. Ubiquitous computing means anywhere (at work, at home, in the city, in the country or on move), anytime (24 hours, day or night), anyone (adults or children, elderly or Handicapped), any thing (home appliances, individual items, cars, food products). It needs numerous base stations (BS) in cellular system and access point (AP) in wireless local area network (WLAN), which cost more. Ad- hoc networking is expected to be one of the key technologies supporting a future ubiquitous network society. Ad hoc network consists of mobile nodes equipped with wireless transceiver, which aims to establish communication anywhere anytime without the aid of infrastructure like BS and APs. Ubiquitous network allows users to exchange information through the use of broadband and mobile access. Smart objects and RFID tags are connected through wide area ubiquitous network.

II. Various Terminologies

A. Ubiquitous Network

Ubiquitous network is a federation of networks on which user-oriented services are provided anywhere and anytime to a target user in the most appropriate way with null operational cost. Ubiquitous network allows all users to access and exchange information of any kind freely at anytime, from anywhere, and from any appliance through the use of broadband and mobile access as well as intelligent home appliances and RFID tags that can access networks [4]. In ubiquitous network small computers will be embedded into wearable terminals such as clothes, watches, glasses, rings, and into every object such as desks and chairs, cupboards, refrigerators, and cars. These chips and devices will be connected to each other through wireless networks and linked to the Internet. A ubiquitous network connects lots of physical objects that exist isolated in the environment [5]. It is similar to Internet that connects computers that are far apart. In order to realize the ubiquitous networks, the combination of IP network and broadband wireless access will play an important role. Ubiquitous network combines optical communication, mobile and consumer electronics into one network [1].

B. Smart Objects

Progress in technologies for sensors (to sense environment), together with the expected increase in processing power and memory, will make every object "smart". In future, more than 50% of the devices connected to the Internet will not be PCs but they will be smart devices and appliances. Smart devices are optimized to particular tasks that they blend into the world and require little technical knowledge on the part of their uses. They should be as simple as to use calculators, telephone or toasters. Networked embedded processors which forms the heart of all smart devices, will become an important R&D field. Reliability is crucial in embedded computing systems since large interconnected information systems become unstable. The building block of an individual smart object can be derived from those of classic computer namely memory and processing, sensors and actuators as in/out devices, wireless communication to connect smart objects, architecture (hardware and software components and the interface between them) and middleware (defines how different participants communicate with one another in a network). The interaction of a person with these networks of smart things requires novel human-computer interface (HCI) i.e. special in/out devices, which enables a natural, almost unconscious interaction

C. RF Identification

One of the major problems in ubiquitous computing is the identification of the objects. RFID tags represent a newer and more interesting concept for identifying objects. A smart device is a small, low power microchip combined with antenna, implanted in a physical object. Each device has a unique serial number and can contain other programmable information. The information contained in a device can be transmitted to a nearby reader device by a RF signal. By holding RFID card near computing device, the reconfigured as your own. Regardless of type of computing device anything, anywhere can be configured as your own computing device.

D. Wireless Technology

In wireless technology Wi-Fi (Wireless Fidelity) corresponds to IEEE802.11a/b/g/n WLAN, WiMAX (Worldwide interoperability microwave access) to IEEE802.16 wireless metropolitan area network (WMAN), and Bluetooth, ultra wide band (UWB), ZigBee to IEEE802.15 wireless personal area network (WPAN). Wireless LAN operating at 2.4GHz can provide high speed data rates for short distances while cellular system operating at 900MHz covers a wide area with relatively low speed data rates. WLAN 802.11n is scheduled to be completed in June 2009 can provide data rate above 100Mb/s in 40MHz BW operating at 2.4GHz. The new regulations released 255Mb/s data rate in 5GHz band for indoor/outdoor use. WiMAX originally developed for fixed wireless access but later on applied to mobile wireless access (IEEE802.16e).

E. Wide Area Ubiquitous Network (WAUN)

WAUN communicate small amount of information (several Kbytes per month per device) over a wide area (5kms radius) using VHF/UHF bands [3]. Application of WAUN could be remote management, gas/water/power meter reading, health care, child tracking, stolen property (car) tracking, vending machine monitoring, environment monitoring, and community security [2]. There can be billions of objects such as mobile nodes, various services, computers, home appliances etc. connected to ubiquitous network. Taking into account the recent advances in wireless technologies such as adaptive modulation and coding (AMC), diversity antennas, and large scale monolithic microwave IC (MMIC) using CMOS devices, WAUN can be made more efficient to cover large range with high bit rate wireless link.

III. Elements Of Ubiquitous Computing

Elements in ubiquitous computing mainly categorized into three types .First one is devices Second one regarding the communication link last one according to the Interfacing technology

Devices

1. Computing Nodes / Sensor-Compute Nodes (miniature to large, single to multi-core)
2. Display devices (hard and soft surface display devices)
3. Input devices (voice, video, touchpad, keypad etc.)
4. Storage devices (short-term to long-term, slow to very fast)

Communication devices (wireless and wire line

Communication Links

1. Physical Links:
- Fixed wire line links
- Fixed wireless links
- Mobile wireless links
- Hybrid multi-links

2. Logical / Virtual links

Interfacing technologies

1. Navigation technologies
2. On-screen / Touch-panel technologies
3. Voice interfacing technologies
4. Video-interfacing technologies
5. Handwriting-based interfacing technologies
6. Hybrid interfacing technologies

IV. Challenges Of Ubiquitous Computing

As with most new technological advances, obstacles are encountered and must be overcome for them to succeed. The idea driving ubiquitous computing is to make computers that are unseen by their users because the devices are so embedded, so natural, so ergonomic, so friendly, and so fitting, they use them without even noticing. The name alone implies computers everywhere, but in order for success to be achieved, they must dissolve into the background. To do this, ubiquitous computing devices must overcome six challenges [4] [5].

A. The "Accidentally" Smart Environment -

If you walk into an environment anywhere in the world, you would probably not find an infrastructure suitable for ubiquitous computing devices. Instead, you would find an infrastructure suited towards established and well-grounded technologies, such as electricity running through the walls, phone lines running into and out of buildings, and conveniences such as indoor plumbing. You are not likely to see newly constructed buildings equipped with devices to support pervasive computing.

B. Improper Interoperability –

It is common knowledge many technology-producing companies desire to produce their own proprietary products speaking their own proprietary language. This leads to "no interoperability" between devices from different companies, stunting the growth of ubiquitous computing. Even if companies agreed on a communication standard to further the new computing age, we still may see stunted growth. Even if two devices are able to communicate, there is still the question of "How well they communicate?"

C. No Systems Administrator –

Challenge Two above ties in nicely with the next test for ubiquitous computing to take root – lack of a systems administrator. Most individuals who operate a personal computer have no intimate knowledge of how to administer a single workstation. It would be unrealistic for the manufacturers of ubiquitous computing devices to expect their users to administer a complex network consisting of multiple devices. How does ubiquitous computing respond to this challenge?

D. Social Implications of Aware Technologies –

Ubiquitous computing will have a social impact on our society just as the previous two eras of computing did – there is no question about that. However, as these devices are constantly bombarded with input from their surroundings, the question then becomes what do the devices do with this input? How will it affect privacy? Will society turn to a social solution, legal solution, ethical solution, or technological solution to protect privacy?

E. Reliability –

If you take a look at the current age of computing and compare the reliability issue of today to what the reliability issue of tomorrow must be, the difference is staggering. Today's personal computers are, in a sense, becoming more and more reliable. However, they still have a long road ahead for them to catch up with the reliability exhibited by other well-founded technologies, such as televisions, telephones and even washing machines. These well-founded technologies have been successful, in part, due to their reliability.

V. Conclusions

Combining every object with identification, localization, and communication technologies will have a significant impact on private life of millions of people with decreasing cost of ubiquitous information and communication technology. The penetration of the technologies will become more widespread and no doubt a large number of business opportunities will emerge within this sector. Multiple devices from multiple vendors with different technological capabilities, equipped with varying communication technologies must cooperate in an ever-changing environment. Much progress in information science, computer science, and material science is necessary to render the vision economically feasible and to overcome current technological hurdles such as energy consumption. Progress in material science, chemistry and physics will change the appearance of information appliances, e.g. light emitting polymer (LEP) display, offers flexible large area or curved displays capable of delivering high resolution video images at low power consumption, visible in daylight and with wide viewing angle.

.

References

[1] P. Tandler, et al, "Room ware-Moving towards ubiquitous computers", IEEE Micro, Nov.Dec, 2002, pp 36-47

[2] R. Want et al, "Disappearing hardware", IEEE Pervasive Computing, Jan-March, 2002, pp 36-47

[3] T. Starner, "The challenges of wearable computing: part 1 &2", IEEE Micro, July- Aug. 2001, pp 44-52 & 54-67

[4].Tandler, P.: Software Infrastructure for a Ubiquitous-Computing Environment Supporting Collaboration with Multiple Single- and Multi-User Devices. Proceedings of UbiComp'01. Lecture Notes in Computer Science, Springer, Heidelberg, 96-115, 2001.

[5].Langheinrich, M. Privacy by Design - Principles of Privacy-Aware Ubiquitous Systems. in International Symposium on Ubiquitous Computing. 2001. Atlanta, USA. p. 273-291.

AUTHOR PROFILE

Prakriti Trivedi received her B.E. degree from MBM Engineering College,Jodhpur Rajasthan affiliated to Jai Narain University, India in 1994 and ME from NITTTR ,Punjab University ,Chandigarh , India in Computer Science & Engineering, She is currently head of the department of CE & IT in the Govt. Engineering College, Ajmer Rajasthan, India.

Kamal Kishore Sagar received his B.E. degree from Rajasthan University, Jaipur and currently he is pursuing M.Tech specializing in Software Engineering from Govt. Engineering College, Ajmer affilated to Rajasthan Technical University, Kota. His field of reasearch are Adhoc Networks, Security and Software engineering.

Vernon received his B.E. degree from Ajmer Institute of Technology affiliated to Rajathan University and currently he is pursuing M.Tech specializing in IT from Govt. Engineering College, Ajmer affilated to Rajasthan Technical University, Kota. His field of reasearch are Artifical Intelligence and Wireless communication technologies.

Modelling and Analysing of Software Defect Prevention Using ODC

Prakriti Trivedi[1] Som Pachori[2]
[1]HOD, Computer Department, [2]M Tech IV Semester (SE),
Govt Engineering College Ajmer
[1]hod.ceit@rediffmail.com [2]sompachori@in.com

***Abstract*— As the time passes the software complexity is increasing and due to this software reliability and quality will be affected. And for measuring the software reliability and quality various defect measurement and defect tracing mechanism will be used .Software defect prevention work typically focuses on individual inspection and testing technique. ODC is a mechanism by which we exploit software defect that occur during the software development life cycle. Orthogonal defect classification is a concept which enables developers, quality managers and project managers to evaluate the effectiveness and correctness of the software**

***Keywords*— Defect Prevention, ODC, Defect Trigger**

I. Introduction

Software defect prevention is an important part of the software development. The quality; reliability and the cost of the software product heavily depend on the software defect detection and prevention process. In the development of software product 40% or the more of the project time is spent on defect detection activities. Software defect prevention research has proposed new inspection and testing methods and has studied and compared different inspection and testing methods.

In the paper the basic idea is to provide implementation of ODC in real world application. It begins with and overview of various defect classification schemes followed by ODC concepts. The latter part will describe how will we adopt ODC in software development.

The end of this paper describes the Improvement in software project after implementing ODC.Page Layout

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

II. DEFECT CLASSIFICATION SCHEME

Since 1975, a number of classification schemes have been developed by different organizations, such as HP and IBM, to classify software defects and to identify common causes for defects in order to determine corrective action

A. Hewlett-Packard - "Company-Wide Software Metrics[9][10]

It classifies defects from three perspectives in three area

(1) Identifying where the defect occurred (e.g., in the design or the code).

(2) Finding out what was wrong (e.g., the data definition or the logic description may be incorrect).

(3) Specifying why it was wrong, missing or incorrect.

B. The IBM Orthogonal Defect Classification Scheme[1]

The IBM Orthogonal Defect Classification (ODC) was originally described in the paper by Chillarege et al. in 1992 [1][4]. As described by Chillarege, the goal of ODC is to provide a scheme to capture the key attributes of defects so that mathematical analysis is possible. The software development process is evaluated based on the data analysis. According to ODC, the defect attributes that need to be captured include: defect trigger, defect type, and defect qualifier. The "defect type" attribute describes the actual correction that was made. For example, if the fix to a defect involves interactions between two classes or methods, it is an interface defect. The "defect trigger" attribute represents the condition that leads the defect to surface. For example, if the tester found the defect by executing two units of code in sequence, the defect trigger is "Test sequencing". "The defect qualifier" indicates whether the defect is caused by a missing or wrong element

III. ODC Concepts

ODC is a defect classification scheme by which we characterize and capture defect information. ODC is a measurement system for software processes based on the semantic information contained in the defect stream. And it can help us evaluate the effectiveness and efficiency of testing, enable error tracking, and evaluate customer satisfaction via an analysis mechanism behind the scheme

A. Defect Trigger

It provides surface to the fault and results in a failure. It just provides a measurement for the development process.

It is very hard for the developer to find the fault during testing Process. For this purpose various verification and testing activities are conducted to find that fault

DEFECT TRIGGER	CLASSIFICATION
REVIEW/INSPECTION TRIGGER	DESIGN CONFORMANCE LOGIC/FLOW BACKWARD COMPATIBILITY LATERAL COMPATIBILITY CONCURRENCY INTERNAL DOCUMENT LANGUAGE DEPENDENCY SIDE EFFECTS RARE SITUATION
FUNCTION TEST TRIGGER	SIMPLE PATH COMPLEX PATH COVERAGE VARIATION SEQUENCING INTERACTION
SYSTEM TEST TRIGGER	WORKLOAD STRESS RECOVERY STARTUP/RESTART HARDWARE CONFIGURATION SOFTWARE CONFIGURATION BLOCKED TEST/NORMAL MODE

B. Defect Type

ODC becomes more understandable when we discuss the defect type. The defect types are chosen so as to be general enough to apply to any software development independent of a specific product. [2] ODC provides a framework for identifying defect types and the sources of errors in a software development effort. Using the feedback provided by analysis of the defects it inserts in its systems.

- FUNCTION: Affects significant capability, end-user interfaces, product interfaces, and interface with hardware architecture or global data structure.
- LOGIC: Affects the functionality of a code module and can be fixed by re-implementing an algorithm or local data structure without a need for requesting a high level design change.
- INTERFACE: Affects the interaction of components via macros, call statements and/or parameters.
- CHECKING: Affects program logic that would properly validate data and values before they are stored or used in computation.
- ASSIGNMENT: Requires change in a few lines of code, such as initialization of control blocks or data structures.
- TIMING/SERIALIZATION: Affects the proper management of shared and real-time resources.
- BUILD/PACKAGE/MERGE: Affects productversion or configuration; requires formal changes to reconfigure or rebuild the product.

IV. ADOPTING ODC IN SOFTWARE DEVELOPMENT

ODC has its own life cycle, which we can integrate into the Iterative software development lifecycle .With this integration, we can monitor software quality status at each development phase[7]. And if we find some abnormalities in our result we

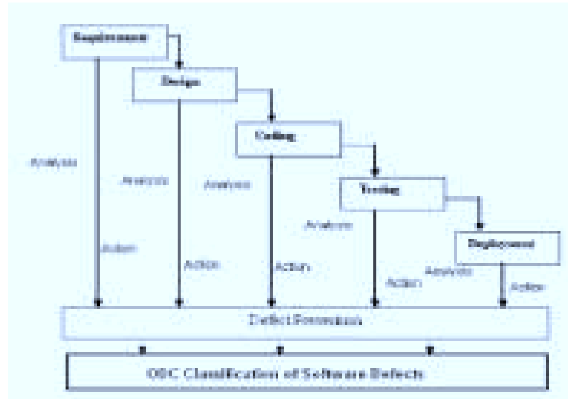


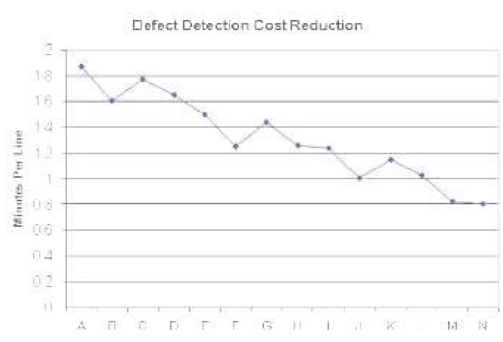
Figure 1:Relation between SDLC and ODC

Therefore, for any given phase, defect detection measures should be taken. The measures must be appropriate to the typical type of defects injected and the information or work product produced. The goal is to minimize the amount of defects that propagates to the subsequent phases.

V. IMPROVEMENT AFTER IMPLEMENTING THE ODC IN SOFTWARE PROJECTS

After implementing ODC in various software project at college level we see that the defect detection rate will be reduced. These projects are very different in terms of the languages (procedure language VB 6 and Object-Oriented language VB .Net ,Web Application develop through ASP.Net, Php), the architectures (client-server and multi-tier), the databases (as simple as SQL and as complicated as Orcale 8i), the resources the complexity (from one week for a single person to several months for eight people), and the characteristics (adding new functionalities to an old system, fixing the bugs in an old system, and developing a new system). With only these projects and the large variations, it is too early to draw a statistical conclusion on what was improved as experience was gained.

The cost of defect detection has dropped dramatically,



although there are some fluctuations

Figure .2: Defect Detection Cost Reduction v/s per project

VI. CONCLUSION AND FUTURE WORK

In this paper we have presented an approach for defining, introducing orthogonal defect classification.ODC can help improve efficiency and effectiveness of development and testing. These are all critical for quality improvement. This paper builds some fundamental work that demonstrated the existence of a relationship between the type of defects and their effect on the software development. By predicting software defect introduction and removal rates, ODC is useful for identifying appropriate defect reduction strategies. The extension for ODC defect types provides more granular insight into defect profiles and their impacts to specific risks. The use of value-neutral software engineering methods often causes software projects to expend significant amounts of scarce resources on activities with negative returns on investment.. The ODC defect detection efficiency functions are being evaluated for different domains and operational scenarios. This research investigated the software defect detection process to address: how to conduct the process better, how to evaluate and control the process better, and how to continuously improve the process.

REFERENCES

[1] Ram Chillarege (www.chillarege.com)

[2] R. Chillarege, W.-L. Kao, andR. G. Condit, "Defect Type and its Impact on the Growt h Curve," in Proceedings of The 13th International Conference on Software Engineering, 1991

[3] Chillarege, R; Bhandari, I; Chaar, J; Halliday, M; Moebus, D; Ray, B; Wong, M; Orthogonal defect classification - A concept for in- process measurements, IEEE Transactions on Software Engineering, vol. 18, pp. 943-956, Nov. 1992

[4]. IBM Research Center for Software Engineering (http://www.research.ibm.com/softeng/ODC/ODC.HTM)

[5] "Software Triggers as a function of time -ODC on field faults", Ram Chillarege and Kathryn A. Bassin, DCCA-5: Fifth IFIP Working Conference on Dependable Computing for Critical Applications, Sept 1995.

[6] "Improving software testing via ODC: Three case studies", M. Butcher, H. Munro, T. Kratschmer, IBM Systems Journal, Vol 41, No. 1, 2002.

[7]. "Identifying Risk using ODC Based Growt h Models", R. Chillarege, S. Biyani, Proceedings, 5th International Symposium on Software Reliability Engineering, IEEE, Monterey, California, pp 282- 288, November 1994

[8] Butcher, M., Munro, H., Kratschmer, T., Improving Software Testing via ODC: Three Case Studies - Orthogonal Defect Classification, IBM Systems Journal, March 2002.

[9] Kan, S. H., Parrish, J., Manlove, D., In-process Metrics for Software Testing, IBM Systems Journal, 40(1), pages 220- 241, 2001.

[10] Grady, Robert B., Caswell, D., L. Software Metrics: Establishing a Company-Wide Program, Prentice Hall, Englewood Cliffs, NJ, 1987.

Enhanced Segmentation Procedure for Intima-Adventitial Layers of Common Carotid Artery

V.Savithri[1] and S.Purushothaman[2]
[1]Department of Computer Science, Mother Teresa Women's University, Kodaikanal,India
[2] Principal, Sun College of Engineering and Technology, Kanyakumari, India
Email: savi3_8@yahoo.co.in, dr.purushothaman.s@gmail.com

***Abstract*— This paper presents an enhanced Segmentation technique for use on noisy B-mode ultrasound images of the carotid artery. This method is based on Image Enhancement, Edge detection and Morphological operations in boundary detection. This procedure may simplify the job of the practitioner for analyzing accuracy and variability of segmentation results. Possible plaque regions are also highlighted. A thorough evaluation of the method in the clinical environment shows that inter observer variability is evidently decreased and so is the overall analysis time. The results demonstrate that it has the potential to perform qualitatively better than applying existing methods in intima and adventitial layer detection on B-mode images.**

***Keywords*— Artery, boundary detection, imaging, Ultrasonic, parallel programming**

I. Introduction

According to Madras Medical College, stroke disease costs India more than 1100 billion rupees per year, and this value continues to increase with each passing year [1]. This amount covers the cost of care and research to help alleviate the dreadful effects of this disease. In order to help such costs, techniques to help simplify and reduce a medical practitioner's daily workload need to be developed.

Accurate measurement and understanding of the geometry of carotid arteries are important in the assessment and management of risk for stroke. An indication of early symptoms, which can be used as a pre-clinical disease marker, is a thickening of carotid plaque levels may block the passage of blood traveling to the upper body and brain. However, plaque and boundary identification in ultrasound images is still a tedious and difficult task due to lack of its automation. Currently in India, practitioners identify plaque regions and other myocardial tissues in ultrasound images by manually tracing their borders to best fit the data.

Several studies are being conducted to improve the quality and detection within ultrasound images [2] designed an efficient dynamic programming algorithm to detect boundaries of carotid artery. Their algorithm creates an accurate boundary detection.[3] designed a dual boundaries with the optimal solution based on a given cost function. However, their algorithms may not give depth study of vessel detection in far-wall adventitia. The basic research on the feasibility study of using B-mode sonographic images to measure the arterial IMT is reported in [20]. Afterwards, some related problems such as measuring the arterial elasticity have been studied [4],[15],[16]. Also, the automatic measurement of IMT variation and lumen diameter variation over cardiac cycles have been investigated [12]. Early studies of automatic detection on static B-mode images can be found in [6] and [17]-[19]. However, they have been performed without any anatomic knowledge on intima and adventitia. Dynamic programming (DP) approaches[13] [21],[22] have been employed to solve many image analysis and computer vision problems such as stereo matching[23], contour detection [24],[25], contour tracking [25], and medical applications [18],[26],[27]. The DP method used for IMT measurement is known at least since 1994 [7]. The advantage of DP techniques is that optimal solution can be guaranteed. Generally, dynamic programming can handle primitive knowledge such as image features and curve smoothness. Seldom researchers [27],[29] take the structural knowledge into consideration.

The authors in [27],[14] proposed a sectional dynamic programming that can detect quasi-parallel lines such as the spline boundaries. [5],The proposed method is piecewise linear and is operated with some segments with a fixed distance defining the width of the spinal cord.[8] The authors treat the structural knowledge as geometric and consider hard constraints that the dynamic programming has to satisfy. Their method is specifically applicable to the spine boundaries. However, some applications require detection of curves with exact positions in pixel such as intima and adventitia detection of CCA wall of B-mode sonographic images [9],[10],[19]. Under [11] this requirement, the constraints might fail on this pixel scale. We illustrate this point by an example: the normal IMT of a normal adult is about 0.5 mm. If the pixel size such as that of a Toshiba SSA-'380-A' scanner(Japan) is 0.096 mm, then one pixel error results in about 20% error in a normal patient, which is unacceptable in clinical measurements.

The study in [29] proposed a method that takes the pixel scale into consideration in the multi dynamic programming. In order to reduce the computation time, some disjoint ness, ordering, spacing and heuristic constraints were suggested. The disjoint ness and spacing constraints can not only reduce the computation time but also help to satisfy the geometric constraints that are suitable for our application.

In this paper, a new technique is implemented to help identify carotid regions semi-automatically by segmenting ultrasound images. This technique would dramatically decrease the time needed to analyze the ultrasonic images. In order to accomplish this segmentation solution, image processing technique such as image enhancement, image threshold, image denoising, edge detection, boundary detection and segmentation, have been used to manipulate the image data.

II. METHODOLOGY

Problem Definition

The noise created during ultrasound scanning leads to difficulty in defining the boundary of the vessel. The image is further deteriorated by the occurrence of lipid rich plaque a poorly angled transducer during image acquisition. Difficult in highlighting plaque region.

A. Enhanced Segmentation Procedure

I. Image Enhancement

The goal of image enhancement is to improve the perception of the image through modification of intensity functions, image spectral content, or a combination of these functions. Removing blurring and noise, increasing contrast, and revealing details are examples of image-enhancement operations. The histogram of an image gives the frequency of occurrence of the gray levels.

Procedure to perform Histogram equalization

Step 1: Find the running sum of the histogram values.

Step 2: Normalise the values from step (1) by dividing by the total number of pixels.

Step 3: Multiply the values from step (2) by the maximum gray-level value and round.

Step 4 : Map the gray level values to the results from step (3) using a one-to-one correspondence.

II. Image Thresholding

Thresholding is a part of a more general segmentation problem. Thresholding is required to extract a part of an image which contains all the information. Thresholding is useful technique for establishing boundaries in images that 22contain solid objects resting on a contrasting background. In hard thresholding, pixels having intensity lower than the threshold T are set to zero and the pixels having intensity greater than the threshold are set to 255 or left at their original intensity depending on the effect that is required. This type of hard thresholding allows us to obtain a binary image from a grayscale image. The mathematical expressions are:

The derivation of the area functions for an object whose boundary is defined by thresholding:

$$H(D) = -d / dD \ A(D) \tag{1}$$

Where D- gray level, A(D) – area of the object at gray level D, and H(D) refers to Histogram.

III. Image Denoising

An image is often corrupted by noise in its acquisition or transmission. Noise is any undesired information that contaminates an image. Noise appears in images from a variety of sources. The goal of denoising is to remove the noise while retaining as much as possible the important signal features. Denoising can be done through filtering. Filters reduce noise. Gaussian highpass filter helps to reduces a noise.

$$H_{hfe}(u,v) = a + b_{Hhp}(u,v) \tag{2}$$

Where a is the offset, b is the multiplier, and $H_{hp}(u,v)$ is the transfer function of the highpass filter.

IV. Edge Detection

This procedure defines edges as Zero-crossings of second derivatives in the direction of the greatest first derivate. This works in multistage process (i) image is smoothed by Gaussian convolution (ii) 2D first derivate operator is applied to the smoothed image to highlight region of the image with high spatial derivatives. The effectiveness of this algorithm is determined by three parameters (i) width of the Gaussian kernel (ii) upper threshold (iii) lower threshold used by tracker.

[g, t] = edge (f, 'canny', T, sigma) (3)

Where T is a vector, T = [t1 , t2], containing the two thresholds and sigma is the standard deviation of the smoothing filter.

V. Morphological operations for Boundary detection

Morphological operations are very effective in the detection of boundaries in a binary image X. The following boundary detectors are widely used:

$$Y = X - (X \theta B) \tag{4}$$

$$Y = (X \oplus B) = X \text{ or}$$

$$Y = (X \oplus B) - (X \theta B) \tag{5}$$

where Y is the boundary image, operator θ denotes erosion operator ⊕ denotes dilation ' – ' denotes set theoretical subtraction.

```
%Boundary detector
Close all;
Clear all;
Clc;
a=imread('carotid.jpg');
b=[010;111;010];
a1=imdilate(a,b);
a2=imerode(a,b);
a3=a-a2;
a4=a1-a;
a5=a1-a2;
imshow(a)
figure,imshow(a1),title('Dilated Image')
figure, imshow(a2),title('Eroded Image')
```

VI. Segmentation of the intima-media region

We introduce a split and merge method for the segmentation of the imtima-media region in ultrasound images, which combines splines(for the adventitia detection), dynamic programming(dp), smooth intensity thresholding surfaces and a successful geometric active contour model and known for its accuracy, flexibility and robustness. Several image features are used in the segmentation. Human interaction is minimal.

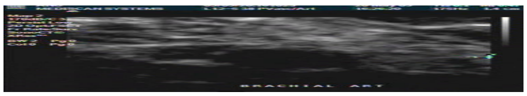
It is able to segment both near-end and far-end carotid walls; it supports to detect plaques of different sizes, shapes and classes.
The basic steps of splitting and merging images are

Split into four disjoint quadrants any region Ri for which
$P(R_i)$ = FALSE.

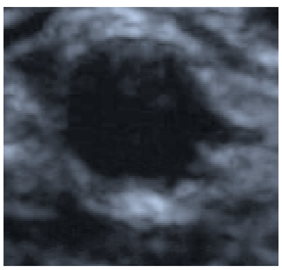
When no further splitting is possible, merge any adjacent regions R_j and R_k for which
$P(R_j \cup R_k)$ = TRUE.

Stop when no further merging is possible.

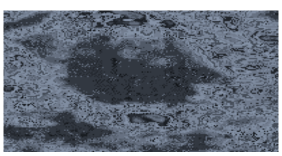
III. RESULTS



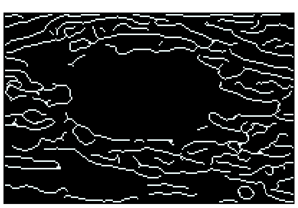
Original Image



gray



indexed



morphed

Segmentation

IV. DISCUSSION

We have proposed a enhanced segmentation technique for detecting intima and adventitia in B-mode ultrasonic images. Our method inherits the property of guaranteeing the global minimal result from TDP. Simultaneously it increases the robustness against the speckle noise effect in the ultrasonic images. Our experimental results show that this method can effectively alleviate the problems caused by the TDP when applied in CCA(far-wall) IMT measurement. In general, Our segmentation technique demonstrates its superior performance over the TDP in this application

Table-1
Reading Variability (%) when measurements were Performed by three readers before applying Procedure.

Reader1 (Accuracy)	Reader2 (Accuracy)	Reader3 (Accuracy)	Average(%) (Accuracy)
82%	80%	84%	82%

Table-2
Reading Variability (%) when measurements were Performed by three readers after applying Procedure.

Reader1 (Accuracy)	Reader2 (Accuracy)	Reader3 (Accuracy)	Average(%) (Accuracy)
92%	92.5%	94%	93%

Our future aim is to test this system on images made by some other sonographic instrumentation. The system should be able to select features fully automatically according to different instrumentations, The final goal is to make this system portable within the current B-mode ultrasonic instrumentation. As a final remark, we want to point out that in addition to the intima and adventitia detection, several other medical applications are based on detecting near-parallel contours. The technique proposed in this paper is potentially applicable in these situations as well.

V. CONCLUSION

A method is proposed based on Enhanced segmentation procedure to automatically measure ultrasonic artery images. The human knowledge of the artery image is incorporated in the system, which makes the system capable in processing images of different quality. Human factors in the determination of the boundaries are reduced. Evaluation of the system shows reduced inter observer variability as well as overall analysis time. The automated artery boundary detection and segmentation, system can replace the old manual system in a clinical application environment.

ACKNOWLEDGMENT

We would like to thank members of Madras Medical College and New hope Scan centre and Hospital , in the Department of Radiologist and Neurologist , offered valuable comments advices and for providing patients personal reports

REFERENCES

[1] Heart and Stroke Foundation, Annual report Findings, 2008.
[2] Q.Liang, I.Wendelhag,J.wikstrand, and T.Gustavasson "A multiscale dynamic programming procedure for boundary detection in ultrasonic artery images,"IEEE Trans.Med.imag.vol.19,No.2,pp 127-142,feb.2000.
[3] Da-Chuan Cheng, Member,IEEE, and Xiaoyi Jiang, member, IEEE, "Detections of Arterial wall in Sonographic Artery images using Dual Dynamic Programming," IEEE TransMed Imag. Vol.12, No.6, November 2008.
[4] L. Angrisani, P. Daponte, C. Liguori, and A. Pietrosanto, "An image based measurement system for the characterization of automotive gaskets," *Measurement*, vol. 25, pp. 169–181, 1999.
[5] M. G. Bond and S. K. Wilmoth *et al.*, "Detection and Monitoring of Asymptomatic Atherosclerosisi in Clinical Trials," *Amer. J. Med.*, vol.86, (suppl 4A), pp. 33–36, 1989.
[6] N. M. El-Barghouty, T. Levine, S. Ladva, A. Flanagan, and A. Nicoladeis, "Histological verification of computerized carotid plaque characterization," *Eur. J. Vascular Endovascular Surg.*, vol. 11, pp. 414–416, 1996.
[7] F. De Man, I. De Scheerder, M.C. Herregods, J. Piessens and H. De Geest Role of Intravascular Ultrasound in Coronary Artery Disease: A new gold standart? Beyond Angiography. Intravascular Ultrasound State-Of-The-Art XX Congres of the ESC, Vol 1 (August 1998)
[8] D. Hausmann, Andre J.S. Lundkvist, Guy Friedrich, Krishnankutty Sudhir, Peter J. Fitzgerald and Paul G. Yock Lumen and Plaque Shape in Atherosclerotic Coronary Arteries Assesed by In EVO Intracoronary Ultrasound Beyond Angiography. Intravascular Ultrasound: State-Of-The-Art XX Congres of the ESC, Vol 1 (August 1998)
[9] F. Escolano, M. Cazorla, D. Gallardo and R. Rizo Deformable Templates for Plaque Thickness Estimation of Intravascular Ultrasound Sequences Pattern Recognition and Image Analysis. Preprints of the VI1 National Symp. On Patt. Recog. and Im. An.Vol 1 (April 1997)
[10] M.A.Bottalico, A.Starita, "EcoStudio:A computer tool to support carotid ultrasound images analysis,Engineering in Medicine and Biology Soc.,IEEE,pp.2428-2430,2000.
[11] Song Chun Zhu, Alan Yuille, Region Competition:Unzfiin Snakes, Region Growing, and BayesIMDL for Multiband Image Segmentation. IEEE Trans. Pattern An. Mach. Intelligence, Vol. 18, No 9 , (September 1996).
[12] Nobuyuki Otsu A Threshold Selection Method from Gray-Level Histograms. IEEE Trans. on Sys. Man and Cybernetics, Vol. SMC-9,Na 1, pp 62-65, (January 1979)
[13] I. Wendelhag, T. Gustavsson, M. Suurkula, G. Berglund, and J. Wikstrand, "Ultrasound measurement of wall thickness in the carotid artery. Fundamental principles and description of a computerized analyzing system," *Clin. Physiol.*, vol. 6, no. 11, pp. 565–577, 1991.
[14] P. Pignoli, E. Tremoli, A. Poli, and R. Paoletti, "Intimal plus medial thickness of the arterial wall: A direct measurement with ultrasound imaging," *Circulation*, vol. 74, pp. 1399–1406, 1986.
[15] R. E. Bellman and S. Dreyfus, Appled Dynamic Programming. Princeton, NJ: Princeton University Press, 1962.
[16] A. A. Amini, T. E.Weymouth, and R. C. Jain, "Using dynamic programming for solving variational problems in vision," IEEE Trans. Pattern Anal. Machine Intell., vol. 12, pp. 855–867, Sept. 1990.
[17] R. J. Kozick, "Detectinh interfaces on ultrasound images of the carotid artery by dynamic programming," *SPIE*, vol. 2666, pp. 233–241, 1996.
[18] W. Liang, R. Briwning, R. Lauer, and M. Sonka, "Automated analysis of brachial ultrasound time series," in *Proc. SPIE Conf. Physiol. Function Multidimensional Images*, vol. SPIE 3337, San Diego, CA, Feb. 1998, pp. 108–118.
[19] P.Pignoli,E.Tremoli,A.Poli,P.Oreste,and R.Paoletti, "Intimal plus medical thickness of the arterial wall: A direct measurement with ultrasound imaging", circulation, vol.74 pp. 1399-1406, 1986.
[20] L.Scharf and H.Elliott,"Aspects of dynamic programming in signal and image processing",IEEE, Trans Autom.control,Vol.AC-26,no.5,pp.1018-1029,Oct.1981.
[21] H.Burkhardt and H.Schorb,"Maximum-a-posteriori restoration of images-an application of the viterbi algorithm to two-dimenstional filtering" in Proc.Int.Conf.Pattern Recog.(ICPR).oct.,1982,pp.98-101.
[22] I.Cox S.Hingorani,S.Rao,B.Maggs,"A maximum likehood stereo algorithm", Computer Vs image understanding,vol.63,pp.24-45,1996.
[23] A.Amini,T.Weymouth, and R.Jain,"Using dynamic programming for solbving variational problems in vision", IEEE Trans.Pattern Analy.Mach. Intell.vol.12,no.9,pp 855-867,sep.1990.
[24] D.Geiger,A.Gupta,L.A.CostaandJ.Vlontzos,"Dynamic programming for detecting,tracking and matchingdeformablecontours,"IEEETrans.Pattern.Anal .Mach. Intell, vol. 17, no.3, pp.294302, mar.1995.
[25] Y.chen,T.Hung and Y.Rui,"optimal radial contour Tracking by dynamic programming, "in proc. Int. Conf .Image Process. (ICIP), 2001 pp.626-629.
[26] G.Wei,J.Qian, and H.Schramm, "Generalized Dynamic programming approaches for object detection:Detecting spine boundaries and vertebra endplates" in Proc.IEEE Conf. Comput.Vis.Pattern Recognis.(CVPR),2001 vol.1,pp.954-959.
[27] G.Aboutanos, J.Nikanna N.Watkins B.Dawant, Model creatinn and deformation the automatic segmentation of the brain in MR images , IEEE, Trans.Biomed.Eng.vol.46,no.11,pp.1346-1356,nov.1999.
[28] C.Su and B.Appleton, " Multiple paths extraction in images using aconstrained expanded Trellis", IEEE Trans. Pattern Anal.Mach. Intell.,vol.27,no.12,pp.1923-
[29] L.MacLeod, "Comments on techniques for edge detection,"Proc.IEEE,vol.60,no.3,p.344,Mar.1972.

Application of Locality Preserving Projections in Face Recognition

Shermina.J[#,]
[#] Department of Computing, Muscat College, University of Stirling, UK,
Shermina@muscatcollege.edu.om

***Abstract*– Face recognition technology has evolved as an enchanting solution to address the contemporary needs in order to perform identification and verification of identity claims. By advancing the feature extraction methods and dimensionality reduction techniques in the application of pattern recognition, a number of face recognition systems has been developed with distinct degrees of success. Locality preserving projection (LPP) is a recently proposed method for unsupervised linear dimensionality reduction. LPP preserve the local structure of face image space which is usually more significant than the global structure preserved by principal component analysis (PCA) and linear discriminant analysis (LDA). This paper focuses on a systematic analysis of locality-preserving projections and the application of LPP in combination with an existing technique This combined approach of LPP through MPCA can preserve the global and the local structure of the face image which is proved very effective. Proposed approach is tested using the AT & T face database. Experimental results show the significant improvements in the face recognition performance in comparison with some previous methods.**

***Keywords*- Image Processing; Face Recognition; Image Compression; Multilinear Systems; Locality preserving projection**

I. INTRODUCTION

Biometric technologies have been evolved as an enchanting solution to perform secure identification and personal verification. The need for highly secure identification and personal verification technologies is becoming apparent as the level of security breaches and transaction fraud increases. The increasing use of biometric technologies in high security applications and beyond has created the requirement for highly dependable face recognition systems. Face recognition system is used to verify an identity of a person by matching a given face against a database of known faces. It has become an alternative to the traditional identification and authentication methods such as the use of keys, ID cards and passwords.

Face recognition involves computer recognition of personal identity based on geometric or statistical features that are derived from the face images [2],[7],[8]. Even though human can detect and identify faces in a scene easily, building an automated system is challenging. Face recognition technology can be applied to a wide variety of application areas including access control for PCs, airport surveillance, private surveillance, criminal identification and for security in ATM transaction. In addition, face recognition system is moving towards the next-generation smart environment where computers are designed to interact more like humans [12].

In recent years, considerable progress has been made in the area of face recognition with the development of many techniques. In the recent past, face recognition research has witnessed a growing interest in subspace analysis techniques [2],[7]. Two of the classical algorithms, principal component analysis (PCA) and linear discriminant analysis (LDA)[2], which are well-known for feature extraction and dimension reduction, has been widely used in face recognition. However, both PCA and LDA effectively see only the linear manifold that based on the Euclidean structure. They fail to capture the underlying structure which lies on a nonlinear submanifold hidden in the image space. Recently, some nonlinear methods have been developed to discover the nonlinear structure of the manifold, algorithms e.g. Isomap, locally linear embedding (LLE), and Locality Preserving Projections (LPP)[3],[7] in which the first two algorithms are nonlinear but the LPP algorithm is a linear dimensionality reduction algorithm.

PCA aims to preserve the global structure of the face image space and the LDA method aims to preserve the discriminating information, but the goal of LPP method is to preserve the local structure of image samples. Locality Preserving Projections describe face images by mapping the face data onto a low-dimensional face feature subspace called "Laplacianfaces". The advantages of LPP algorithm are that LPP is a linear method and it preserve the local information of the face image space. The limitation of LPP is that it represents an image by a vector in high-dimensional space which is often confronted with the difficulty that sometimes the image matrix is singular. It is unsupervised and hence some of the important information with the classification are neglected. To overcome the complication of the singular problem, Laplacianface method projects the image set onto a PCA subspace to confirm to be non-singular.

In this paper, we are projecting the face data onto a Multilinear Principal Component Analysis (MPCA) subspace, and LPP algorithm is further used to preserve the local structure information. This combined approach considering the global and local structure of the face image space can obtain a more effective optimal subspace for face representation and recognition. First, it compresses and preserves the principal information in a matrix form, so it removes more inherent redundancy, and a much lower dimensional face representation is acquired by which the recognition speed is enhanced greatly. Second, the low dimensional represent exempts the consequent LPP step from the singularity problem and it also achieves a more competitive accurate recognition rate than the Laplacianface. The organization of the paper is as follows: Literature review is presented in Section 2. Description about the face recognition techniques used in our research is presented in section 3. Methodology of the approach is presented in Section 4. Experimental results and the comparative analysis is given in section 5 and finally the conclusions are summed up in Section 6.

II. LITERATURE REVIEW

Many face recognition methods have been developed in the past few decades. Most common feature extraction methods are principal component analysis (PCA)[1] and linear discriminant analysis (LDA) [2]. Another linear technique which is used for face recognition is Locality Preserving Projections (LPP) [3],[4], which finds an embedding that preserves local information, and gains a face subspace that best detects the essential face manifold structure[9],[11]. Xiaofei He introduced the Locality Preserving Projections (LPP) as when the high dimensional data lies on a low dimensional manifold embedded in the ambient space, the Locality Preserving Projections are obtained by finding the optimal linear approximations to the eigen functions of the Laplace Beltrami operator on the manifold. Because of this, LPP, being linear, shares many of the data representation properties of nonlinear techniques such as Laplacian Eigenmaps or Locally Linear Embedding.

Lin Kezheng *et al* [13] proposed the Enhanced Locality Preserving Projections, to identify the underlying manifold structure of a data set. ELPP considers both the between-class scatter and the within-class scatter in the processing of manifold learning. Equivalently, the goal of ELPP is to preserve the within-class geometric structure, while maximizing the between-class distance. Zhonglong Zheng *et al.*[18] proposed supervised locality preserving projection (SLPP), using class labels of data points to enhance its discriminant power in their mapping into a low dimensional space. The GSLPP method, which is robust to variations of illumination and facial expression, applies the SLPP to an augmented Gabor feature is vector derived from the Gabor wavelet representation of face images.

Yu Weiwei *et al.* [17] proposed two-dimensional discriminate locality preserving projections (2D-DLPP), which benefits from three techniques, i.e., locality preserving projections (LPP), image based projection and discriminant analysis. Deng Cai *et al.* [14] have proposed an appearance based face recognition method, called orthogonal Laplacian face in which face data may be generated by sampling a probability distribution that has support on or near a sub-manifold of ambient space. Earlier works based on PCA or LDA [2] suffer from not preserving the local manifold of the face structure whereas the research works on LPP lacks to preserve global features of face images[3]. Some papers [3], [7] uses the combination of both PCA and LPP, captures only the most expressive features.

Yi Jin *et al.* [5] presented a new manifold learning algorithm in which a bilateral-projection-based 2DPCA (B2DPCA) for image matrix compression is performed before supervised locality preserving projections. The bilateral-projection-based DPCA algorithm is used to obtain the meaningful low dimensional structure of the data space. Also those works that uses PCA captures the variation in the samples without considering the variance among the subjects. The combination of global feature extraction technique LDA and local feature extraction technique LPP to achieve a high quality feature set called Combined Global and Local Preserving Features (CGLPF) that captures the discriminate features among the samples considering the different classes in the subjects which produces the considerable improved results in facial image representation and recognition[12].

Motivated by this research, our approach combines the global feature preservation technique MPCA and the local feature preservation technique LPP to form the high quality feature set. The concept is to project the face data to an MPCA space for preserving the global information and then projecting to Locality Preserving Projection (LPP) space by using the distance preserving spectral methods, to add the local neighbourhood manifold information. Analysing these techniques, a unique feature extraction technique is not felicitous when the dimensionality of face images attempts to reach its peek and combination of two feature extraction methods on distinctly separate subspaces seems to be more effective in the performance of face recognition.

III. METHODOLOGY

A. Locality Preserving Projections

Locality Preserving Projections (LPP) are linear projective maps that arise by solving a variational problem that optimally preserves the neighbourhood structure of the data set. LPP represents a linear approximation of the nonlinear Laplacian eigenmaps introduced in [3]. When high-dimensional data lies on a low dimension manifold embedded in the data space, then LPP approximate the eigenfunctions of the Laplace-Beltrami operator of the manifold. LPP aims at preserving the local structure of the data. This is unlike PCA and LDA, which aims at preserving the global structure of the data. LPP is unsupervised and performs a linear transformation. It models the manifold structure by constructing an adjacency graph, which is a graph expressing local nearness of the data. This is highly desirable for face recognition compared to non-linear local structure preserving, since it is significantly less computationally expensive and more importantly it is defined in all points and not just in the training points as Isomaps and Laplacian Eigenmaps. Let xi, $i = 1, 2, \cdots, n$, denote the training patterns of m classes. We use $X = [x1, x2, \cdots, xn]$ to denote the data matrix and use $l(xi)$ to denote the label of xi, say, $l(xi) = k$ implies that xi belongs to class k. LPP aims at preserving the intrinsic geometry of the data by forcing neighboring points in the original data space to be mapped into closely projected data. The algorithm starts by defining a similarity matrix **W**, based on a (weighted) k nearest neighbors graph, whose entry **W**$i\ j$ represents the edge between training images (graph nodes) **x**i and **x**j . Gaussiantype weights of the form $W_{ij} = e^{-\frac{\|x_i - x_j\|^2}{t}}$ have been proposed in [23], although other choices (e.g., cosine type) are also possible. Based on matrix **W**, a special objective function is constructed, enforcing the locality of the projected data points by penalizing those points that are mapped far apart. Basically, the approach reduces to finding a minimum eigenvalue solution to the generalized eigenvalue problem.

B. Algorithm

Locality Preserving Projection (LPP) is one of the linear approximation obtained from the nonlinear Laplacian Eigenmap [3]. The algorithmic procedure of LPP is stated below:

1) Construction of adjacency graph: Let G denote a graph with m nodes and an edge between nodes i and j , if x_i and x_j are close. There are two variations:

(a) ε-neighbourhoods: Nodes i and j are connected by an edge if $\| x_i - x_j \|^2 < \varepsilon$, where the norm is the usual Euclidean norm in R^n .

(b) k nearest neighbors: Nodes i and j are connected by an edge if i is among k nearest neighbors of j or j is among k nearest neighbors of i .

2) Choosing the weights: We have two variations for weighting the edges. W is a sparse symmetric $m \times m$ matrix with W_{ij} having the weight of the edge joining vertices i and j , and 0 if there is no such edge.

(a) Heat kernel: If nodes i and j are connected,

$$W_{ij} = e^{-\frac{\|x_i - x_j\|^2}{t}}$$

(b) Simple-minded: $W_{ij} = 1$, if and only if vertices i and j are connected by an edge.

3) Eigenmaps: Compute the eigenvectors and eigenvalues for the generalized eigenvector problem:

$$XLX^T a = \lambda XDX^T a$$

Where D is a diagonal matrix whose entries are column (or row, since W is symmetric) sums of W , $D_{ii} = \sum_j W_{ji} . L = D - W$ is the Laplacian matrix. The i^{th} column of matrix X is x_i .

Let the column vectors $a_0, \cdots, a_{l-1}$ be the solutions of equation (1), ordered according to their eigenvalues, $\lambda_0 < \cdots < \lambda_{l-1}$. Thus, the embedding is as follows:

$$x_i \to y_i = A^T x_i , A = (a_0, a_1, \cdots, a_{l-1})$$

Where y_i is a l-dimensional vector and A is a $n \times l$ matrix.

C. Implementation

Initially, image processing techniques such as normalization and resizing of the face images are employed in preprocessing in order to improve the face image since variations in lighting conditions dramatically decrease recognition performance. When processing a face, the features like variations in light, image quality, persons' pose, facial expressions are taken into account. Normalization is used to outweigh any illumination variations or relative sizes between two sets of faces. The face images in the database are usually of different dimensions. Therefore the input samples necessitate to be resized to standard dimensions. Hence, image pre-processing and normalization is the important part of face recognition systems. Next, the feature extraction is achieved by merging the MPCA along with LPP to calculate the feature projection matrices. MPCA receives the set of face image samples of the same dimensions as input for feature extraction. The resultant output of the MPCA is the dimensionally reduced feature projection matrix of face images. The dimension reduced feature projection matrices of face image samples obtained using MPCA is then fed as an input to the LPP algorithm. Locality Preserving Projection (LPP) is one of the linear approximation obtained from the nonlinear Laplacian Eigenmap [9]. The dimensional reduced feature matrices of the training sample images obtained using the MPCA and LPP techniques are stored in a database. While we are testing the face images, the aforesaid techniques are applied to generate the feature matrix and thereby a similarity measure is carried out on the sample face images. The face recognition can be done by comparing the test feature matrix with the enrolled face features in the database using L2 distance that measures the query images with the nearest database images. The similarity distance measure for a pair of face images is computed in which a threshold determines whether the pair of face is identified as similar or not.

IV. COMPARATIVE ANALYSIS

Existing experiments on AT&T facial images formerly called as ORL database were analyzed and the performance is observed as per the procedures applied to AT&T facial images [15]. Accuracy in face recognition is to computed with the false acceptance rate (FAR) which is percentage of incorrect acceptances and false rejection rate (FRR) which is the percentage of incorrect rejections. The genuine acceptance rate (GAR) is computed using these factors and the overall accuracy measurement of the proposed approach is calculated using the formula *100-(FAR/FRR)/2* which is shown in Fig. 1. The comparative results of existing appearance based methods in face recognition were analyzed and found that the performance of LPP supersedes the PCA and LDA approaches. Moreover, the latest experimental results show that the combination of LPP with MPCA improves the accurate recognition rate in face recognition. The table clearly shows the efficiency of LPP in comparable with the other approaches. The comparative recognition rate of various approaches is shown in the following table1.

TABLE I
Comparison of recognition rates on AT&T databases

Data base	**Performance Comparisons**	
	Method	**Accurate Recognition Rate**
AT&T	PCA(Eigenface)	88.50 %
	LDA(Fisherface)	91.00%
	LPP(Laplacianface)	94.00%
	MPCA + LPP	96.5%

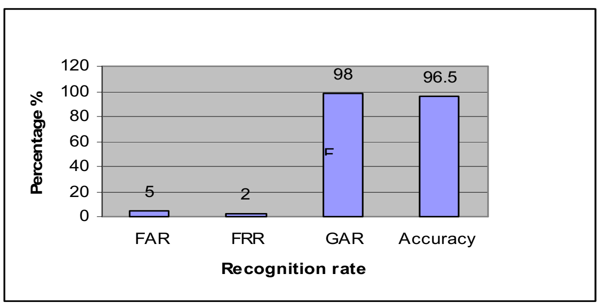


Fig 1 Accuracy rate of LPP + MPCA approach

V. CONCLUSION

Face Recognition is the active area of research over the past two decades which have resulted in successful techniques for recognizing the facial images. In this article, the importance of Locality Preserving Projection is studied and the comparison of existing face recognition technique with the impact of LPP is also analysed. The combined appearance based technique such as LPP with MPCA outperforms the previously developed locality preserving projection approach to yield a good recognition rate. With these results, we conclude that the combination of LPP with any other approaches could improve the recognition accuracy. Therefore, in future work, the combination of LPP with other appearance based approaches could be experimented for accurate recognition rate in order to determine the most efficient approach in appearance based face recognition.

ACKNOWLEDGMENT

I would like to thank Dr.V.Vasudevan for providing the guidance in the preliminary discussions in order to proceed further in my area of research.

REFERENCES

[1] H. Moon and P.J. Phillips, “Computational and Performance aspects of PCA based Face Recognition Algorithms”, *Perception*, vol. 30, pp. 303- 321, 2001.

[2] P.N. Belhumeur, J.P. Hepanha, and D.J. Kriegman, “Eigen faces vs.fisherfaces: recognition using class specific linear projection” *IEEE. Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997.

[3] X. He and P. Niyogi, “Locality preserving projections, ”in Advances in *Neural Information Processing Systems*. Cambridge, MA: MIT Press, 2003.

[4] Li, Ruidong; Zhu, Lei; Zou, Shaofang; “Face Recognition Based on an Alternative Formulation of Orthogonal LPP*”, Control and Automation, IEEE International Conference* on May 30 2007-June 1, pp.2374 – 2376, 2007.

[5] Yi Jin and Qiu-Qi Ruan, “An image matrix compression based supervised LPP for face recognition” *International Symposium on Intelligent Signal Processing and Communication Systems*, 2007

[6] M. Turk and Pentland, “Face Recognition Using Eigenfaces”, in *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, Maui, Hawaii, 1991.

[7] Xiaofei He, Shuicheng Yan, Yuxiao Hu, Partha Niyogi, and Hong-Jiang Zhang, “Face Recognition Using Laplacianfaces” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, No. 3, March 2005

[8] M. Belkin and P. Niyogi, “Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering,” *Advances in Neural Information Processing Systems* 14, Vancouver, British Columbia, Canada, 2002.

[9] Borut Batagelj and Franc Solina, “ Face Recognition in Different Subspaces: A Comparative Study”

[10] Haiping Lu, K.N. Plataniotis and A.N. Venetsanopoulos, "MPCA: Multilinear Principal Component Analysis of Tensor Objects", *IEEE Transactions on Neural Networks*, Vol. 19, No. 1, January 2008.

[11] Zhao, W. Chellappa, R., Phillips, P. J. and Rosenfeld, A., “Face Recognition: A Literature Survey”, *ACM Computing Survey*, December, pp: 399-458, 2003.

[12] Ruba.S.Kathavarayan and Murugesan Karuppasamy, “Preserving Global and Local Features for Robust Face Recognition under Various Noisy Environments”, *International Journal of Image Processing*, Volume(3), Issue(6)

[13] Lin Kezheng, Lin Sheng and Chen Dongmei: “Enhanced Locality Preserving Projections”. CSSE (1) 2008: 985-988

[14] Deng Cai, Xiaofei He, Jiawei Han and Hong-Jiang Zhang, "Orthogonal Laplacianfaces for Face Recognition", *IEEE Transactions On Image Processing*, 2006

[15] "AT&T database of faces" {Online} Available:http://www.cl.cam.ac.uk/ research/dtg/attarchive/ facedatabase.html

[16] Zhonglong Zheng, Fan Yang, Wenan Tan, Jiong Jia and Jie Yang, “Gabor feature-based face recognition using supervised locality preserving projection”, *Signal Processing* Volume 87, Issue 10, October 2007, Pages 2473-2483

[17] Yu Weiwe “Two-dimensional discriminant locality preserving projections for face recognition” *Pattern Recognition* Letters Volume 30, Issue 15, 1 November 2009, Pages 1378-1383