Online ISSN: 0975-4172
Print ISSN: 0975-4350

Global Journal of Computer Science and Technology

DISCOVERING THOUGHTS AND INVENTING FUTURE

Technology 10 Reforming Ideas

September 2011

Pinnacles
- Analysis of Data Mining
- Maintenance Mechanism
- Wireless Sensor Network
- Energy Efficient Network

The Volume 11
Issue 16
Version 1.0

ENG

Global Journal of Computer Science and Technology

Global Journal of Computer Science and Technology

Volume 11 Issue 16 (Ver. 1.0)

Open Association of Research Society

Incorporation No.: 0423089
License No.: 42125/022010/1186
Registration No.: 430374
Import-Export Code: 1109007027
Employer Identification Number (EIN):
USA Tax ID: 98-0673427

Global Journals Inc.

(A Delaware USA Incorporation with "Good Standing"; Reg. Number: 0423089)

Sponsors: Open Association of Research Society

Open Scientific Standards

Publisher's Headquarters office

Global Journals Inc., Headquarters Corporate Office, Cambridge Office Center, II Canal Park, Floor No. 5th, ***Cambridge (Massachusetts)***, Pin: MA 02141 United States

USA Toll Free: +001-888-839-7392
USA Toll Free Fax: +001-888-839-7392

Offset Typesetting

Open Association of Research Society, Marsh Road, Rainham, Essex, London RM13 8EU United Kingdom.

Packaging & Continental Dispatching

Global Journals, India

Find a correspondence nodal officer near you

To find nodal officer of your country, please email us at *local@globaljournals.org*

eContacts

Press Inquiries: *press@globaljournals.org*
Investor Inquiries: *investers@globaljournals.org*
Technical Support: *technology@globaljournals.org*
Media & Releases: *media@globaljournals.org*

Pricing (Including by Air Parcel Charges):

For Authors:
22 USD (B/W) & 50 USD (Color)
Yearly Subscription (Personal & Institutional):
200 USD (B/W) & 250 USD (Color)

Editorial Board Members (HON.)

Dr. Bart Lambrecht
Director of Research in Accounting and FinanceProfessor of Finance
Lancaster University Management School
BA (Antwerp); MPhil, MA, PhD (Cambridge)

Dr. Carlos García Pont
Associate Professor of Marketing
IESE Business School, University of Navarra
Doctor of Philosophy (Management), Massachusetts Institute of Technology (MIT)
Master in Business Administration, IESE, University of Navarra
Degree in Industrial Engineering, Universitat Politècnica de Catalunya

Dr. Fotini Labropulu
Mathematics - Luther College
University of ReginaPh.D., M.Sc. in Mathematics
B.A. (Honors) in Mathematics
University of Windso

Dr. Lynn Lim
Reader in Business and Marketing
Roehampton University, London
BCom, PGDip, MBA (Distinction), PhD, FHEA

Dr. Mihaly Mezei
ASSOCIATE PROFESSOR
Department of Structural and Chemical Biology, Mount Sinai School of Medical Center
Ph.D., Etvs Lornd University
Postdoctoral Training,
New York University

Dr. Söhnke M. Bartram
Department of Accounting and FinanceLancaster University Management SchoolPh.D. (WHU Koblenz)
MBA/BBA (University of Saarbrücken)

Dr. Miguel Angel Ariño
Professor of Decision Sciences
IESE Business School
Barcelona, Spain (Universidad de Navarra)
CEIBS (China Europe International Business School).
Beijing, Shanghai and Shenzhen
Ph.D. in Mathematics
University of Barcelona
BA in Mathematics (Licenciatura)
University of Barcelona

Philip G. Moscoso
Technology and Operations Management
IESE Business School, University of Navarra
Ph.D in Industrial Engineering and Management, ETH Zurich
M.Sc. in Chemical Engineering, ETH Zurich

Dr. Sanjay Dixit, M.D.
Director, EP Laboratories, Philadelphia VA Medical Center
Cardiovascular Medicine - Cardiac Arrhythmia
Univ of Penn School of Medicine

Dr. Han-Xiang Deng
MD., Ph.D
Associate Professor and Research
Department Division of Neuromuscular Medicine
Davee Department of Neurology and Clinical NeuroscienceNorthwestern University
Feinberg School of Medicine

Dr. Pina C. Sanelli
Associate Professor of Public Health
Weill Cornell Medical College
Associate Attending Radiologist
NewYork-Presbyterian Hospital
MRI, MRA, CT, and CTA
Neuroradiology and Diagnostic Radiology
M.D., State University of New York at Buffalo,School of Medicine and Biomedical Sciences

Dr. Roberto Sanchez
Associate Professor
Department of Structural and Chemical Biology
Mount Sinai School of Medicine
Ph.D., The Rockefeller University

Dr. Wen-Yih Sun
Professor of Earth and Atmospheric SciencesPurdue University Director
National Center for Typhoon and Flooding Research, Taiwan
University Chair Professor
Department of Atmospheric Sciences, National Central University, Chung-Li, TaiwanUniversity Chair Professor
Institute of Environmental Engineering, National Chiao Tung University, Hsin-chu, Taiwan.Ph.D., MS The University of Chicago, Geophysical Sciences
BS National Taiwan University, Atmospheric Sciences
Associate Professor of Radiology

Dr. Michael R. Rudnick
M.D., FACP
Associate Professor of Medicine
Chief, Renal Electrolyte and Hypertension Division (PMC)
Penn Medicine, University of Pennsylvania
Presbyterian Medical Center, Philadelphia
Nephrology and Internal Medicine
Certified by the American Board of Internal Medicine

Dr. Bassey Benjamin Esu
B.Sc. Marketing; MBA Marketing; Ph.D Marketing
Lecturer, Department of Marketing, University of Calabar
Tourism Consultant, Cross River State Tourism Development Department
Co-ordinator , Sustainable Tourism Initiative, Calabar, Nigeria

Dr. Aziz M. Barbar, Ph.D.
IEEE Senior Member
Chairperson, Department of Computer Science
AUST - American University of Science & Technology
Alfred Naccash Avenue – Ashrafieh

Contents of the Volume

Global Journal of Computer Science and Technology
Volume 11 Issue 16 Version 1.0 September 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Analysis of Data Mining Based Software Defect Prediction Techniques

By Naheed Azeem, Shazia Usmani

Federal Urdu University

Abstract - Software bug repository is the main resource for fault prone modules. Different data mining algorithms are used to extract fault prone modules from these repositories. Software development team tries to increase the software quality by decreasing the number of defects as much as possible. In this paper different data mining techniques are discussed for identifying fault prone modules as well as compare the data mining algorithms to find out the best algorithm for defect prediction.

Keywords : *Defect prediction, Data Mining.*

GJCST Classification : *H.2.8, D.2.9*

Strictly as per the compliance and regulations of:

Analysis of Data Mining Based Software Defect Prediction Techniques

Naheed Azeem[α], Shazia Usmani[Ω]

Abstract - Software bug repository is the main resource for fault prone modules. Different data mining algorithms are used to extract fault prone modules from these repositories. Software development team tries to increase the software quality by decreasing the number of defects as much as possible. In this paper different data mining techniques are discussed for identifying fault prone modules as well as compare the data mining algorithms to find out the best algorithm for defect prediction.

General Terms : Reliability
Keywords : Defect prediction, Data Mining

I. Introduction

Software life cycle is a human activity, so it is impossible to produce the software without defects. To deliver a defect free software it is imperative to predict and fix the defects as many as possible before the product delivers to the customer.

Software repositories have lots of information that is useful in assessing software quality. Data mining techniques and machine learning algorithms can be applied on these repositories to extract the useful information.

The aim of this research is to explore the different issues and problems in the area of defect prediction as well as provide the solutions to improve the product quality via defect prediction mechanism.

In this survey four type of research issues, formulated as questions, need to be addressed to understand the problems of defect prediction mechanism based on data mining techniques.

Research questions:

- How can we resolve the problem of ceiling effects as well as imbalanced and highly skewed datasets?
- What software repositories and datasets should be mined for defect prediction?
- How can we get better results in identifying defects from large features and high level software modules?
- How machine learning algorithms and data mining techniques can be proved more effective in defect extraction from repository?
- Is there any good data mining technique that performs the best in all situations?

The remainder of this paper begins with a background and description. (Section 2), followed by Issues and problems regarding data mining in defect prediction and its solution(section 3) , future work and open issues are discussed in section 5and finally summarizes the paper (section 4).

Author α : Department of Computer Science Federal Urdu University
E-mail : naheedazeem@fuuast.edu.pk
Author Ω : Department of Computer Science Federal Urdu University
E-mail : shaziausmani@fuuast.edu.pk

II. Background and Descriptions

A ***software defect*** is an error, flaw, mistake, failure, or fault in a computer program or system that produces an incorrect or unexpected result, or causes it to behave in unintended ways [24].

Software defects are expensive in terms of quality and cost. Moreover, the cost of capturing and correcting defects is one of the most expensive software development activities. It will not be possible to eliminate all defects but it is possible to minimize the number of defects and their severe impact on the projects. To do this a defect management process needs to be implemented that focuses on improving software quality via decreasing the defect density. A little investment in defect management process can yield significant returns.

a) Software Defect Prediction

Software defect prediction is the process of locating defective modules in software. To produce high quality software, the final product should have as few defects as possible. Early detection of software defects could lead to reduced development costs and rework effort and more reliable software. So, the study of the defect prediction is important to achieve software quality.

The most discussed problem is software defect prediction in the field of software quality and software reliability. As Boehm observed finding and fixing a problem after delivery is 100 times more expensive than fixing it during requirement and design phase. Additionally software projects spend 40 to 50 percent of their efforts in avoidable rework [25].

b) Data Mining and Machine Learning Techniques

Data mining techniques and machine learning algorithms are useful in prediction of software bug estimation. Machine learning models and Data mining techniques can be applied on the software repositories to extract the defects of a software product. Common techniques include decision tree learning, Naive Bayesian classification and neural networks, j48 and ONER.

III. Issues and Problems

Software prediction model only works well when enough amount of data is available in software repository within the organization to initially feed the model. Extraction of defects from software bug repository accurately is not done without a good data mining model. There is a need of good data mining model to predict the software defects from a bug repository.

a) Highly skewed and imbalanced datasets

- Existing prediction models based on un sampling as well as training dataset does not contain any information about number of fault per module and distribution of fault among modules [3].
- Data mining algorithms lack of business knowledge and hit a performance ceiling effect when cannot extract the additional information that related to software metrics with fault occurrence [16].
- Fit datasets are usually imbalanced that cause the degradation of defect prediction models [22].
- Highly skewed dataset is considered as the main cause of unsatisfactory prediction result. However the results of more balanced dataset are also unsatisfactory [23].

b) Early life cycle and multiple dataset

- Early life cycle data cannot be useful in identifying fault prone modules [9, 20].
- No change in defect predictions results when different software repositories are mined [11].
- Single classifier is technically unfit to make use of all the features. However the problems of combing different classifier still remain unresolved [14].

c) Large number of features and high level software modules

- Most of the machine leading algorithms are not capable of extracting defects from the database that store continuous features [7].
- Supervised learning are useful for defect prediction at same logical levels but it is not suitable for high level software modules [8].
- Existing classifier based defect prediction model are insufficient accurate for practical use and use of a large number of features [13].

d) Accurate defect prediction model

- There is a need of accurate defect prediction model for large-scale software system which is more robust to noise [2].
- Traditional decision tree are used in classification of defective and non-defective modules. However traditional decision trees induction method contain several disadvantages [4].
- There is a need of good data mining model to predict the software defects from a bug repository [5].
- Data transformation can improve the performance of software quality models [21].

e) Consistent data mining technique

- A good data mining technique to build a better prediction model is an open issue [1].
- Quality professional cannot find appropriate defect prediction techniques because there is no comparative study that asses the performance of these techniques [6].
- A good data mining technique to build a better prediction model is an open issue [10].
- Evaluation of different prediction model is still an open issue as well as effort reduction gain by using such model is ignored during evaluation [15].
- Various fault prediction techniques have been proposed but no one has proven to be consistently accurate [17].
- A good data mining technique to build a better prediction model is an open issue [18].
- Different prediction techniques are used to assess the software quality but there is a lack of comparative study to evaluate the effectiveness of various models [19].

IV. Approaches and Methodologies

a) Sampling effect on imbalanced datasets

An oversampling method is proposed that using the number of fault per modules and distribution of fault among modules. Two prediction models Naïve Bayes and Logistic regression are applied to two dataset from NASA MDP project .Sampling and over sampling method are used. The result of T test and the non-parametric method of Wilcoxon test showed that oversampling method significant influence on the prediction of both LR and NB model [3].

Author in [16] proposed a human-in-the-loop CBR tool that ad business knowledge to the data mining algorithm. CBR build better prediction model that detect the lower bound on the number of instances. Using three sub sampling techniques (over, under and micro sampling) to find the lower bound the number of training instances. Naive Bayes and j48 methods are used in case of over and under sampling and Naive Bayes is used in case of micro sampling.

Another technique used Sampling method to improve the performance of defect prediction models when data sets are imbalanced [22].Four sampling methods (random over sampling, synthetic minority over sampling, random under sampling and one-sided selection) applied to four fault-proneness models(linear discriminant analysis, logistic regression analysis, neural network and classification tree) by using two module sets of industry legacy software.

A method SimBoost is used to handle the software defect prediction problem when high skewed datasets are used, with a fuzzy based classification. A novel method SimBoost is applied on the NASA project dataset to reduced the effects of skewed dataset but the

prediction on more balanced dataset are still not accurate. So, fuzzy classification was used to accurate the prediction result [23].

b) Effect of early life cycle and multiple dataset

Most of the researchers raise the issue that relying on single data source can limit the accuracy of defect prediction models. However, a combination of different data sources is better to utilize in order to built more accurate fault prediction models.

Both papers [9, 20] analyzed that early lifecycle data can be highly useful in defect prediction. In [9] a hybrid Defect prediction models consisting of K-means clustering and C 4.5 are built. Requirement metrics and code metrics and the combination of both requirement and code metrics are applied on these models. Compare the result of models on three NASA projects i.e. CM1, JM1 and PC1. Result shows requirement metric plays an important role in identifying defects. While in the paper [20] author built a Defect prediction models using requirement metrics and code metrics and the combination of both requirement and code metrics. Compare the result of models on three NASA projects. Result shows requirement metric plays an important role in identifying defects.

Author [11] claimed that Defect prediction results improve significantly with different data sources. Three repositories static analysis, version control and release management are used for defect prediction. Learning algorithm ID3, J48 and SVM are used to assess the accuracy of different data sources.

A method is proposed to build a software quality model using multiple learners induced on multiple training datasets to take advantage of their respective biases. Seventeen classifier models were used on seven NASA datasets. Multiple classifiers were combined by majority voting of experts. Four classification scenarios were used to evaluate the result [14].

c) Large number of features and high level software modules

The paper [7] proposed a new data mining model to predict the software bug estimation more accurately. This technique used an entropy based splitting criteria and minimum description stopping criteria (decide when to stop discretization).The binary discretization was always select the best cut point and was applied recursively.

Author investigated that a novel Multi-instance learning technique is much better in identifying defects for high level software modules. Four multi-instance learning algorithm i.e. Statistical learner, Set Kernel, Citation KNN (k Nearest Neighbor) and MI EM-DD (Expectation-Maximization version of Diverse Density) are investigated against three supervised learners Naïve Bayes, Multi-layer Perceptron and logistic Regression [8].

A feature selection algorithm is proposed in [13] that decrease the number of features used by a machine learning classifier for fault prediction. Perform a feature selection process using gain ratio to reduce the set of features in an iterative form. These reduced features are then used to train the two classification model i.e. Naïve Bayes and SVM. Finally the performance of two classifier are assessed in terms of overall prediction accuracy, buggy precision, recall, F-measure, and ROC area under curve (AUC).

d) Need of accurate defect prediction model

The paper [2] present a software defect prediction model based on random forest which is more robust to outliers and noise than other classifiers and beneficial for large-scale software system. They applied Random forest on five different data set of NASA project using two machine learning tools WEKA and See5. Finally they compare the accuracy of random forest with other statistical methods such as logistic regression and discriminant analysis.

Earlier studies have addressed the use of evolutionary decision tree in classification of defective and non-defective modules. But in [4] author used Evolutionary decision tree in a multi population genetic algorithm. SAEDT is applied on promise dataset using software metrics. The result shows better generalization and higher accuracy.

In [5] a two step data mining model is proposed to predict software bug estimation. In first step, a weighted similarity model is used to match the summary and description of new bug from the previous bug in the bug repository. In the second step calculate the duration of all the bugs and the average is calculated.

The authors [21] criticized that data transformation can improve defect prediction model. They proved it with four data transformation methods applied on ten software quality models on nine dataset from MDP. The performances of models are compared through different test i.e. the Friedman test, the Nemenyi test and the Wilcoxon test.

e) Need of a Consistent data mining technique

This paper focused on using and comparing the performance of different machine learning algorithms to build a prediction models based on source code measures and history data. Confusion matrix may be inappropriate for evaluation criteria. Nine different machine learning algorithms are used to build prediction models for a java legacy system to identify the fault prone modules. Compare the performance of each model using confusion matrix and cost sensitive criteria [1].

In [6], author Evaluate the performance of five data mining algorithm named J48, CART, Random Forest, BFTree and Naïve Bayesian classifier (NBC). The performances of algorithms are evaluated using WEKA tool on software metric dataset KC2 from NASA

database. Cross validation test are applied to verify the results. Result shows that performance of algorithm is depends on various factors like problem domain , nature of dataset etc.

Another comparison is done in [10]. This paper compares the three most used data mining techniques. The performance of J48 is better than ONER and ONER is better than Naïve Bayes. Two datasets having 1212 modules and 101 modules was used to evaluate the performance of three machine learning algorithms i.e. J48 , ONER and Naïve Bayes with the help of WEKA tool.10 fold cross validation was applied to confirm the result.

Performances of five classifier prediction model based only on the size of modules measured in LOC are evaluated. Data sets from NASA MDP are used to evaluate the performance of trial defect prediction model based only on the size of modules measured in LOC. Compare the performance of five classifier including Naive Bayes, Logistic Regression, CART decision tree learner, bagging and random forest.When model is evaluated using AUC it shows surprising well results while evaluated using proposed performance measure, the result becomes worst [15].

Researchers [17] evaluate the performance of different fault prediction techniques on different real time software data sets. But no particular technique that prove consistently accurate. Seven different learning methods are applied on Real time data sets from NASA MDP repository to predict the fault prone modules. Also different methods are trained to combine with statistical method PCA. Assess the performance of machine learning algorithm.

MCLP method to build a better prediction model and assess the performance by comparing with other classification algorithm was reported in [8]. Different method are used for generating prediction model include C4.5, Decision Tree, Support Vector Machine (SVM), Neural Network(NN) and Multiple Criteria Linear Programming (MCLP) and applied to data set taken from NASA MDP. Assess the performance of the prediction models based on accuracy, probability of detection (PD), and probability of false alarm (PF).

While in [19], author proposed an ideal a software defect management system based on data mining techniques and data mining models. Proposed methodology of this paper based on three data mining techniques classification, clustering and association rule with two specific data mining models Bayesian Network and Probabilistic Relational Model.

V. Conclusions

Defects can assess in directing the software quality assurance measures as well as improve software management process if developers find and fix them early in the software life cycle.

Effective Defect prediction is based on good data mining model. In this we surveyed different data mining algorithms used for defect prediction. We also discuss the performance and effectiveness of data mining algorithms. This survey also has showed that all the issues for selecting a data mining technique for defect prediction and their provided solutions have been discussed.

Our most important finding is that there is no single data mining technique that is more powerful or suitable for all type of projects. In order to select a better data mining algorithm, domain expert must consider the various factors like problem domain, type of data sets, nature of project, uncertainty in data set etc. Multiple classifiers were combined by majority voting of experts to get more accurate result.

Our findings indicate that early life cycle data can be highly effective in defect prediction. However, a combination of different data sources can utilize to get better prediction results.

Another finding of this paper is Sampling method are useful to improve performance when dataset are highly skewed. Data transformation cannot improve the performance of defect prediction. Integration of discretization method with classification algorithm improves the defect prediction accuracy by transforming the continuous features into discrete features. However different techniques are applied in identifying defects for large features and high level software modules.

VI. Future Work and Open Issues

Future work in this area should:

- Study other unsupervised or semi supervised learning framework and compare the performance of other different data mining algorithms to find out the best algorithm for defect prediction.
- Establish an improved method for predicting software quality via combining different classifier based on different software measures and different voting schemes.
- Analyze the affect of classifier with feature selection and find out whether the cost sensitive learning algorithms can be used to build better defect prediction models.

References References Referencias

1. E.Arisholm,L.C.Briand and M.Fuglerud,"Data Mining Techniques for Building Fault-proneness Models in Telecom Java Software", Proceedings of The 18th IEEE International Symposium on Software Reliability,pp.215-224.
2. L.Guo,Y.Ma,B.Cukic and H.Singh,"Robust prediction of fault-proneness by random forests",Proceedings of the 15th International Symposium on Software Reliability Engineering,2004, pp.417-428.

3. L.Li and H.Leung,"Using the Number of Faults to Improve Fault-Proneness Prediction of the Probability Models", Proceedings of the 2009 WRI World Congress on Computer Science and Information Engineering - Volume 07, pp.722-726.
4. M.Chiş,"Evolutionary Decision Trees and Software Metrics for Module Defects Identification", International Journal of Computer and Information Science,2008, pp. 273-277.
5. N. K. Nagwani and S. Verma ,"Predictive Data Mining Model for Software Bug Estimation Using Average Weighted Similarity" , 2010 IEEE 2nd International Advance Computing Conference, pp.373-378.
6. N.Gayatri,S.Nickolas,A.V.Reddy and R.Chitra,"Performance Analysis of Data Mining Algorithms for Software Quality Prediction", International Conference on Advances in Recent Technologies in Communication and Computing, 2009. ARTCom '09,pp.393 - 395.
7. P. Singh and S. Verma, "An Investigation of the Effect of Discretization on Defect Prediction Using Static Measures", IEEE International Conference on Advances in Computing, Control, and Telecommunication Technologies (2009), pp.837-839.
8. P.Huang and J.Zhu,"Predicting Defect-Prone Software Modules at Different Logical Levels", International Conference on Research Challenges in Computer Science, 2009. ICRCCS '09, pp.37 - 40.
9. P.S.Sandhu, R.Goel, A.S.Brar, J.Kaur and S.Anand ,"A Model for Early Prediction of Faults in Software Systems",The 2nd International Conference on Computer and Automation Engineering (ICCAE), 2010 , pp.281-285
10. P.Singh," Comparing the effectiveness of machine learning algorithms for defect prediction", International Journal of Information Technology and Knowledge Management, 2009 ,pp. 481-483.
11. R.Ramler,S.Larndorfer and T.Natschläger,"What Software Repositories Should Be Mined for Defect Predictors?", Proceedings of the 2009 35th Euromicro Conference on Software Engineering and Advanced Applications , pp.181-187.
12. S.Shafi,S.M.Hassan,A.Arshaq,M.J.Khan and S.Shamail,"Software quality prediction techniques: A comparative analysis", 4th International Conference on Emerging Technologies, 2008. ICET 2008, pp.242-246.
13. S.Shivaji,E.J.Whitehead,R.Akella and S.Kim, "Reducing Features to Improve Bug Prediction", 24th IEEE/ACM International Conference on Automated Software Engineering, ASE'09, pp.600-604.
14. T.M.Khoshgoftaar, P.Rebours and N.Seliya ,"Software quality analysis by combining multiple projects and learners" , Software quality journal 2009, Springer, vol. 17, pp.25-49.
15. T.Mende and R.Koschke,"Revisiting the Evaluation of Defect Prediction Models", Proceedings of the 5th International Conference on Predictor Models in Software Engineering 2009.
16. T.Menzies,B.Turhan,A.Bener,G.Gay,B.Cukic and Y.Jiang, "Implications of ceiling effects in defect predictors" , Proceedings of the 4th international workshop on Predictor models in software engineering, 2008,Leipzig,Germany ,pp.47-54.
17. V.U.B.Challagulla,F.B.Bastani,I.Yen and R.A.Paul, "Empirical Assessment of Machine Learning based Software Defect Prediction Techniques", Proceedings of the 10th IEEE International Workshop on Object-Oriented Real-Time Dependable Systems,2005,pp.263-270.
18. X.Zhao,Y.Liu and S.Yong, "Predicting Software Defects using Multiple Criteria Linear Programming", Proceedings of the International Symposium on Intelligent Information Systems and Applications (IISA'09)2009, pp.583-585.
19. Y.Chen,X.Shen,P.Du and B.Ge,"Research on software defect prediction based on data mining",The 2nd International Conference on Computer and Automation Engineering (ICCAE), 2010, pp.563-567.
20. Y.Jiang ,B.Cukic and T.Menzies ,"Fault Prediction using Early Lifecycle Data", Proceedings of The 18th IEEE International Symposium on Software Reliability 2007, pp.237-246.
21. Y.Jiang,B.Cukic andT.Menzies,"Can data transformation help in the detection of fault-prone modules?" , Proceedings of the 2008 workshop on Defects in large software systems, July 20-20, 2008, Seattle, Washington ,pp.16-20.
22. Y.Kamei,A.Monden,S.Matsumoto,T.Kakimoto and K.Matsumoto,"The effect of over and under sampling on fault–prone module detection", First International Symposium on Empirical Software Engineering and Measurement,2007.ESEM 2007,pp.196-204.
23. Z.Li and M.Reformat,"A practical method for the software fault-prediction", IEEE International Conference on Information Reuse and Integration, 2007. IRI 2007, pp.659-666.
24. en.wikipedia.org/wiki/Software_bug
25. www.cs.umd.edu/projects/SoftEng/ESEG/papers/82.78.pdf

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 11 Issue 16 Version 1.0 September 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Proposing a New Approach to Applying Pervasive Computing In Agriculture Environments

By Mohammadreza Mohammadrezaei, Nima Attarzadeh
Islamic Azad University, Ramhormoz, Iran

Abstract - The resource management in agriculture environments is very important. Using smart controls will be one of the most eminent ways of managing. These resources such as water and plant nutrition. In this paper researcher are going to present a special program in which provide necessary resources for growing plant by using data sensors based on environment conditions. Firstly, it is gained a few data from soil, climate and plant conditions by using sensors and made context by processing all the data. In next stage the presented approach will do its own calculations on the basis of conditions. It can be said that researchers are used fuzzy logic for calculations because of complex data. Then researchers by using actuators can make decision for environment. In this paper, because of injecting nutrition on the basis of its conditions in to soil and plant necessary, plants can frequently use suitable quanta of nutrition and …, won't be on stress danger

Keywords : *pervasive computing, agriculture environment ,sensor network, fuzzy logic.*

GJCST Classification : *I.2.3*

Strictly as per the compliance and regulations of:

Proposing a New Approach to Applying Pervasive Computing In Agriculture Environments

Mohammadreza Mohammadrezaei[α], Nima Attarzadeh[Ω]

Abstract - The resource management in agriculture environments is very important. Using smart controls will be one of the most eminent ways of managing. These resources such as water and plant nutrition. In this paper researcher are going to present a special program in which provide necessary resources for growing plant by using data sensors based on environment conditions. Firstly, it is gained a few data from soil, climate and plant conditions by using sensors and made context by processing all the data. In next stage the presented approach will do its own calculations on the basis of conditions. It can be said that researchers are used fuzzy logic for calculations because of complex data. Then researchers by using actuators can make decision for environment. In this paper, because of injecting nutrition on the basis of its conditions in to soil and plant necessary, plants can frequently use suitable quanta of nutrition and …, won't be on stress danger.

Keywords: pervasive computing, agriculture environment, sensor network, fuzzy logic.

I. Introduction

The technology progress, improves its way whole the world and life. These progresses are affected in the forms of porches, relation with others, mobility from one place to another and other aspect of human life. In fact, human life is improved toward pervasive computing. This kind of calculations make an incredible situation for computers context in which while they are available and usable for people, are concealed and invisible.

For gaining this point, computers must be little in accounting tools shape and put them in walls, buildings and furniture's pervasive computing may be defined as the utilization of all the computers power in physical users area, in which are invisible from users viewer. It will be main point, using computers in human life without necessary to their presentation [1]. Pervasive computing context aware, are kind of this calculations in which are depend on context and automatically able to react and update itself with due to context.

Having information from context will automatically make active system and cause reducing The rate of user's disorder with program and intelligently assist them.

The context can be consisting of any information's in which are usable in order to create characters to a Situation/presentation. A presentation can be a person, place or an object in which are depended to relate between user and program in which consist of their owns users and programs. In fact, the pervasive computing on the basis of basic conditions making decision without human's actions. There is needed to sensors and actuator in order to relate system to area. Sensors will be a group of tools in which are able to collect every comprehension from environments conditions and situations. They usually received qualities of environment and convert into digital amounts.

Actuator will be a group of tools in which are able to grant all the users wanted on the environment [2]. In this discourse reviewing the annals of research's struggles in pervasive computing on the context and also historical records of using sensors network in agriculture environments in section 2, we present fuzzy logic concepts in section3, then section4 describes methodology. In section5 we present our results and conclusions.

II. The Annals of Researcher's Activity

a) *Researchers activity in pervasive computing based on context*

Sensors, in fact are a part of pervasive computing system based on context, in which will use them in this system in order to collect data and we are able to use this kind of calculations in different ways. In Aware-home project, researchers create an extraordinary home in which understand residents movements and assist them [3], Coal Town project, connected to… company, image a city in the future in which all the people, places, object, furniture, will introduce as the number one citizens of wireless and wire of global communications. In this visionary city, all the services and tools calculations will be context aware and available in communication network global [4]. In pervasive health care projects from Denmark Arouse university, general services will have created in or out sides hospital in order to assist patients and make

Author α : Department of Computer Engineering, Ramhormoz Branch, Islamic Azad University, Ramhormoz, Iran.
E-mail : Mohammadrezaei@iauahvaz.ac.ir
Author Ω : Department of Computer Engineering, Mahshahr Branch, Islamic Azad University, Mahshahr, Iran.
E-mail : n.attarzadeh@mahshahriau.ac.ir

available all the patients and physicians wanted, automatically and wisely [5], it can be said that, suggestions have been propounded [6] graphic tools have been designed by human and computers cooperation group of Cornel university by using these tools, attaching text notes from one place to another will be feasible and possible [7].

b) Hystorical records of using sensor networks in agriculture environments

Firstly, sensors used in military applications but by time passing, their utilization improved. Sensor network are a group of small sensors in which insist and cooperate together in order to collect information's [8].

III. Fuzzy Logic

These kinds of network are powerful and organized and also able to guard soil and etc. sensor networks are used in agriculture environments in order to resources management, pest controls and etc. Zhouho.Zhang could develop wireless sensor for golden house monitoring. In this network sensors are used in order to collect soil, wet, and environment temperature [9]. Aline Baggio has presented a design in order to utilize sensor network for controlling pests and used it for potato yield [10].

IV. The Proposed Method

Methods having different information from soil will be one of the most important factors in order to make discussion about changing soil characteristics. But enabling to obtain information about cheap and fast characteristics of soil in one of the greatest limitations in agriculture part. In order to solve this problem, researchers will present a new approach for applying pervasive computing context aware in agriculture environments (PCAE). The approach for applying pervasive computing context aware in agriculture environments will be an understanding from soil and water conditions in order to grow plant. In this part, utilizing pervasive computing context aware would be explained. The utilizing pervasive computing context aware in agriculture environment and its architectural are showed in figure2 and fig.3 respectively.

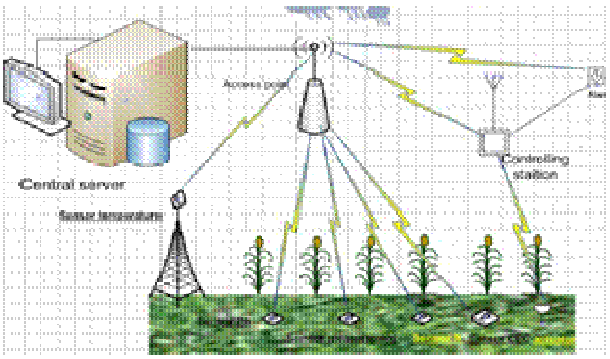


Fig1 : views of pervasive computing context aware in agriculture environments

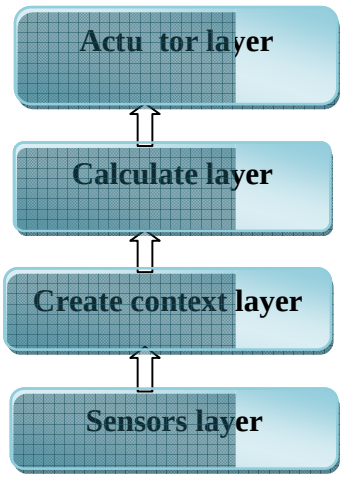


Fig2 : A model from pervasive computing context aware in agriculture environments

Generally, using pervasive computing context aware in agriculture environments has four layers constructions.

a) The layer of sensors

In this layer, hardware sensors are placed in to layer in order to create context. It can be said that, researchers need lot of information's about soil, water and plant conditions. For that reason, are classified in three groups.

b) A layer for creating context

A system pervasive computing context aware needs some information's about context in order to change behavior according to information's, in order to make available context in every time. It can be assume that a compilations as the context manager will be presence by sensors. Context manager will control available situations, time by time and update different context and finally, the present material in environment will provide on intelligent atmosphere and give service more than olden time. This form of communications and calculations will be available in environment in which can be named host. In order to create context, researchers will study and research the kind of plant and climate conditions.

Table1 : the static and dynamic conditions of created materials of context

Agent / Condition	Soil Tissue	Particulate Air Pollution	Weather Condition	Nutrit-ional substa-nces in the soil	Soil's PH
Static	X	…	…	…	…
Dynamic	…	X	X	X	X

c) The layer of calculations

In this layer, in which resolution subject will be available, all the calculations most be done on context aware. In this section, because of data intricate, researchers utilize fuzzy logic in order to calculate during process. After creating context, presented plan or context aware and laws, in which have been explained, the rate of necessary materials of plants are calculated and sent to actuator layer. In fact, in this layer a fuzzy logic are placed in which will control the rate of necessary materials in extern and context aware (fig.4) fuzzy controller , all the conditions will check and select, according to lows and selected lows will calculate the rate of injecting materials in to fuzzy block and send to next layer.

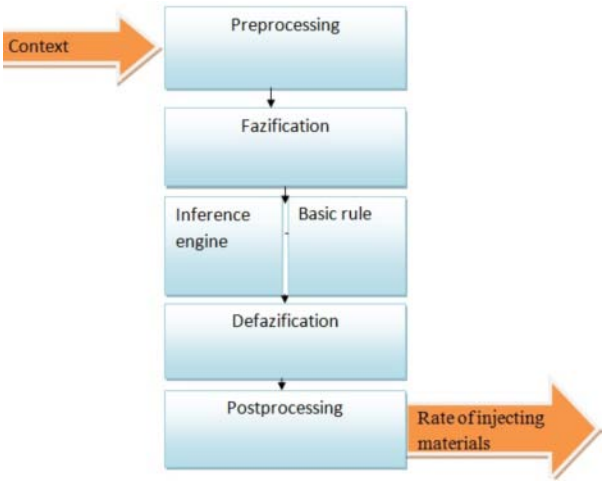


Fig 3 : fuzzy controller

d) Actuators layer and alarm system

Actuators layer in which are connected to external environment. In this layer a hardware actuators will be available in which can be use it in order to exert calculations environment. This pare include irrigation systems, spraying systems, injecting chemical fertilizer digital systems. In which will be able to inject nutritive materials in to soil and also on alarm system will find in this layer. Alarm system is used in emergency situations and conditions when plant will be at danger and orders can network harmoniously. When critical conditions threat a plan, alarm system will work and operate.

V. Simulation Results

The proposal requires the data to simulate real conditions for growing a particular plant. This data is used to grow corn in a laboratory. Table2 [13] shows the corn-fed conditions under which the parameters are defined by fuzzy logic.

Corn	Nitrogen	PH	Conductivity of Soil	
Bad	< 39000	< 9	Low	> 6
Good	7000	5.5 – 7	---------	----------
Critical	35000	> 4	High	< 10

Table 2 : Elements needed to grow corn

For corn, the third parameter has been studied in three favorable, unfavorable and critical shown are Figures 5, 6 and 7.

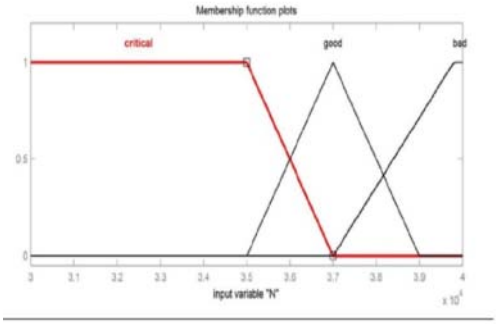


Fig 4 : Phase diagram of nitrogen needed for corn

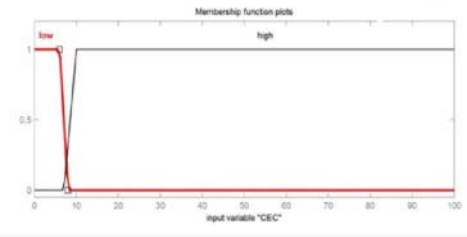


Fig 5 : Phase diagram's PH for grow corn

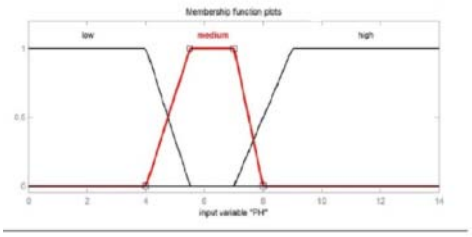


Fig 6 : Phase diagram for the conductivity of soil

Fuzzy rules for adjusting the soil nitrogen is presented in Figure 8. In these rules, depending on soil's PH used of three types of fertilizer for regulate the soil nitrogen. If PH is low, the oure is used, but if PH is medium, the nitrate ammonium is used, and if PH is high, the sulfate-ammonium is used.

1. If (N is critical) and (PH is low) and (CEC is low) then (oure is very high)
2. If (N is critical) and (PH is medium) and (CEC is low) then (nitrate ammonium is very-high)
3. If (N is critical) and (PH is high) and (CEC is low) Then (sulfate-ammonium is very high)
4. If (N is critical) and (PH is low) and (CEC is high) then (oure is high)
5. If (N is critical) and (PH is medium) and (CEC is high) then (nitrate-ammonium is high)
6. If (N is critical) and (PH is high) and (CEC is high) then (sulfate - ammonium is high)
7. If (N is good) and (PH is low) and (CEC is low) then (oure is high)
8. If (N is good) and (PH is medium) and (CEC is low) then (nitrate-ammonium is high)
9. If (N is good) and (PH is high) and (CEC is low) then (sulfate-ammonium is high)
10. If (N is good) and (PH is low) and (CEC is high) then (oure is medium)
11. If (N is good) and (PH is medium) and (CEC is high) then (nitrate-ammonium is medium)
12. If (N is good) and (PH is high) and (CEC is high) then (sulfate-ammonium is medium)
13. If (N is bad) and (PH is low) and (CEC is low) then (oure is medium)
14. If (N is bad) and (PH is low) and (EC is high) then (oure is low)
15. If (N is bad) and (PH is medium) and (CEC is low) then (nitrate -ammonium is medium)
16. If (N is bad) and (PH is medium) and (CEC is high) then (nitrate-ammonium is low)
17. If (N is bad) and (PH is high) and (CEC is low) then (sulfate-ammonium is medium)
18. If (N is bad) and (PH is high) and (CEC is high) then (sulfate-ammonium is low)

Fig 7 : Fuzzy rules for adjusting the soil nitrogen

Simulation results are expressed in Figures 9, 10 and 11. Figures extracted from the simulations are very similar to Figure 5, 6 and 7 showed that the ideal condition, In this case, shows performance the proposed scheme.

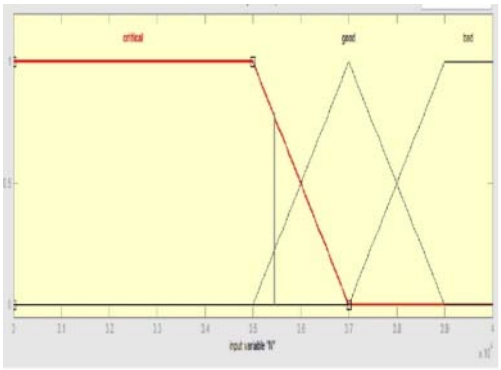


Fig 8 : Simulation result for nitrogen needed for corn

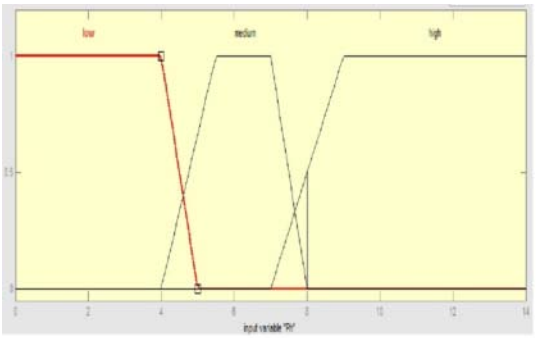


Fig 9 : Simulation result for the conductivity of soil

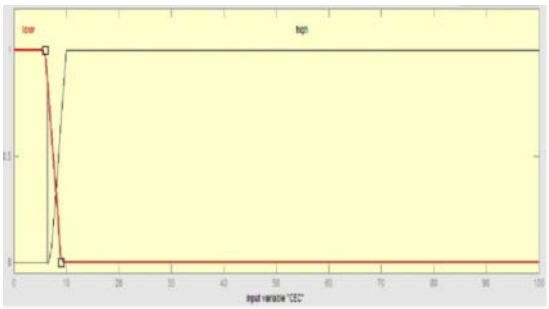


Fig 10 : Simulation result for PH for grow corn

VI. Result

In this research a new approach has been presented in order to apply pervasive computing context aware in agriculture environments. In this approach sensors are used in order collect data about soil, water, plant and climate conditions and send to layer of creating context and after that send to layer of calculation, then, the rate of injecting materials in to soil are calculated by fuzzy logic and context aware. After that are exerted on environment by actuators in the last projects, researchers have utilized sensors in agriculture environment but in primary form such as designing Drip irrigation in primary sensor networks form. But in this paper approach, sensors are generally used in order to collect data of environment and all the next decisions will be on the basis of collected data.

References References Referencias

1. Weiser, M., "The Computer for the 21st Century", Scientific American, 94 -104, September 1991.
2. Walker, K., kabashi, A., Abdelnour, J., Ngugi, K., Underwood, J., Elmirghani, J., and Prodanovic, M., "Interaction design for rural agricultural sensor networks", Internal Environmental Modeling and Software society(iEMSs) , 2008 .
3. Lun, W.Y and Lau, F.C.M., "A Context -Aware Decision Engine for Content Adaptation", IEEE Pervasive Computing Vol.1., no.3, pp.41-49., jul-sep -2002.

4. Saha, D., Mukherjee, A., “Pervasive Computing:A paradigm for 21st Century”.
5. Judd, G., Steenkiste, “providing contextual Information to Pervasive Computing Applications”, Proc, IEEE International Conference on Pervasive Computing, March 2003.
6. Chi, Ed.H., Borriello, G., Hunt,G., Davies,N., ”Pervasive Computing in Sports Technologies”, IEEE PERVASIVE computing, JULY-SEPTEMBER 2005.
7. Burrell, J., Gay, G., ”E -Graffiti:Evaluation Real-World use of a Context -Aware systwm”, Integrating with Computers : Special Issue on Universal Usability, 4. 301 -312, 2003.
8. Baggio, A., "Wireless sensor networks in precision agriculture”, In Proc.ACM Workshop Real-Word Wireless Sensor Network, 2005.
9. Loke, S.,”Context–Aware Pervasive Systems”, Auerbach Publications, Pages 2-13, 2006.
10. Tanaka, T., " an Introduction to Fuzzy logic for practical applications", springer vaerlay, 1996.
11. Fuller, R., Carlsson, C., “Fuzzy multiple criteria decision making”, Recent developments, Fuzzy Sets and Systems 78(2) 139-153., 1996.
12. Grabisch, M., Fuzzy integral in multicriteria decision making, Fuzzy Sets and System 69(3) 279-289,1995
13. “Corn starters”,available at http://www.agtest.com/articles , 2000.
14. Zhang, Z., “Investigation of wireless sensor networks for precision agriculture”, American Society of Agricultural and Biological Engineers, 2004.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 11 Issue 16 Version 1.0 September 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Effective File Replication and Consistency Maintenance Mechanism in P2P Systems

By K. Shalini, Y. Surekha

PVP Siddhartha Institute of Technology Vijayawada, Andhra Pradesh, India

Abstract - In peer–to-peer file sharing systems, file replication and consistency maintenance are widely used techniques for high system performance. Despite significant interdependencies between them, these two issues are typically addressed separately. Most file replication methods rigidly specify replica nodes, leading to low replica utilization, unnecessary replicas and hence extra consistency maintenance overhead. Most consistency maintenance methods propagate update messages based on message spreading or a structure without considering file replication dynamism, leading to inefficient file update and hence high possibility of outdated file response. This paper presents an Integrated file Replication and consistency Maintenance mechanism that integrates the two techniques in a systematic and harmonized manner. It achieves high efficiency in file replication and consistency maintenance at a significantly low cost. Instead of passively accepting replicas and updates, each node determines file replication and update polling by dynamically adapting to time-varying file query and update rates, which avoids unnecessary file replications and updates. It dramatically reduces overhead and yields significant improvements on the efficiency of both file replication and consistency maintenance approaches

Keywords : *File replication, consistency maintenance,peer-to-peer, distributed hash table.*

GJCST Classification : *C.2.3*

Strictly as per the compliance and regulations of:

Effective File Replication and Consistency Maintenance Mechanism in P2P Systems

K. Shalini[α], Y. Surekha[Ω]

Abstract - In peer–to-peer file sharing systems, file replication and consistency maintenance are widely used techniques for high system performance. Despite significant interdependencies between them, these two issues are typically addressed separately. Most file replication methods rigidly specify replica nodes, leading to low replica utilization, unnecessary replicas and hence extra consistency maintenance overhead. Most consistency maintenance methods propagate update messages based on message spreading or a structure without considering file replication dynamism, leading to inefficient file update and hence high possibility of outdated file response. This paper presents an Integrated file Replication and consistency Maintenance mechanism that integrates the two techniques in a systematic and harmonized manner. It achieves high efficiency in file replication and consistency maintenance at a significantly low cost. Instead of passively accepting replicas and updates, each node determines file replication and update polling by dynamically adapting to time-varying file query and update rates, which avoids unnecessary file replications and updates. It dramatically reduces overhead and yields significant improvements on the efficiency of both file replication and consistency maintenance approaches.

Keywords : File replication, consistency maintenance, peer-to-peer, distributed hash table.

I. Introduction

Over the past years, the immerse popularity of Internet has produced a significant stimulus to peer-to-peer (P2P) file sharing systems. A recent large-scale characterization Of HTTP traffic has shown that more than 75 percent of Internet traffic is generated by P2P applications. The percentage of P2P traffic has increased significantly as files such as videos and audios have become almost pervasive. File replication is an effective method to deal with the problem of overload condition due to flash crowds or hot files. It distributes load over replica nodes and improves file query efficiency. File consistency maintenance to maintain the consistency between a file and its replicas is indispensable to file replication. Thus, file replication should proactively reduce unnecessary replicas to minimize the overhead of consistency maintenance, which in turn provides guarantee for the fidelity of consistency among file replicas considering file replication dynamism.

Author α Ω : Department of Computer Science & Engineering, PVP Siddhartha Institute of Technology Vijayawada, Andhra Pradesh, India
E-mails : shalini.hi2007@gmail.com, rekha_18y@yahoo.com

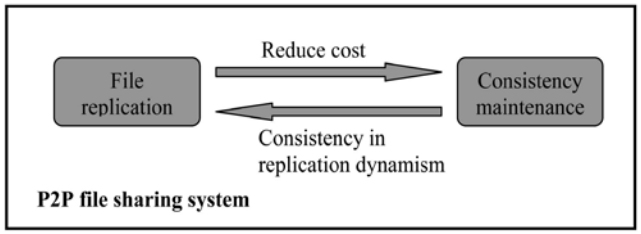


Fig. 1 : Interrelationship between file replication and consistency maintenance

This paper presents an Integrated File Replication and Consistency Maintenance mechanism that achieves high efficiency in file replication and consistency maintenance at a significantly lower cost.

II. Related Work

File replication in P2P systems is targeted to release the load in hot spots and meanwhile decrease file query latency. Generally, the methods replicate files near file owners [2], [3], [4], file requesters [5], [6], or along a query path from a requester to a owner [1], [7]. PAST [2], CFS [3], and Backslash [4] replicate each file on close nodes near the file's owner. In LAR [5] and Gnutella [6], overloaded nodes replicate a file at requesters. Freenet [1] replicates files on the path from a requester to a file owner. CFS, PAST, LAR [5] cache routing hints along the search path of a query. Cox et al. [7] studied providing DNS service over a P2P network as an alternative to traditional DNS. Other studies of file replication investigated the relationship between the number of replicas, file query latency, and load balance [8], [9], [10], [11], [12], [13] in unstructured P2P systems. In most of these methods, file owners rigidly determine replica nodes and nodes passively accept replicas. They are unable to keep track replica utilization to reduce underutilized replicas and ensure high utilization of existing replicas. In our previous work, we proposed an efficient and adaptive decentralized file replication algorithm in P2P file sharing systems called EAD [14]. In the method, traffic hubs that carry more query load and frequently requesters are chosen as replica nodes. The nodes periodically compute their query load to create replicas and remove underutilized replicas. Replication in a structured P2P system is to decrease file query time, while replication in an unstructured P2P system is to decrease the search time. Unstructured P2P systems allow for more proactive

replications of objects, where an object may be replicated at a node even though the node has not requested the object.

In most of these file replication and consistency maintenance methods, nodes passively accept replicas and update messages. They are unable to keep track the utilization of replicas to determine the need of file replicas and replica updates. Minimization of the number of replicas helps to reduce unnecessary updates in consistency maintenance, but it should still keep the efficiency of file replication to release the load in hot spots and to improve query efficiency.

III. Integrated File Replication and Consistency Maintenance(IRM)

Instead of passively accepting replicas and update messages, it harmonically integrates file replication and consistency maintenance by letting each node autonomously determine the need for file replication and update based on actual file query rate and update rates. File replication places replicas in frequently visited nodes to guarantee high utilization of replicas, and meanwhile reduce underutilized replicas and overhead of consistency maintenance.

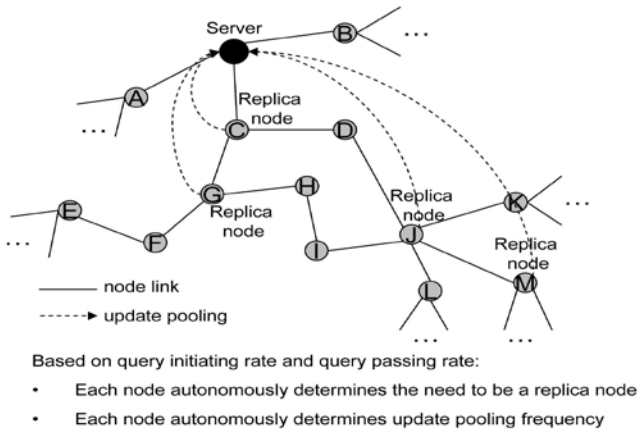


Fig.2 : IRM file replication and consistency maintenance

Consistency maintenance in turn aims to guarantee file fidelity of consistency at a low cost with file replication dynamism consideration. Using adaptive polling, this ensures timely update operation and avoids unnecessary updates. The basic idea of IRM is to use file query and update rate to direct file replication and consistency maintenance.

a) Adaptive File Replication

Integrated File Replication and Consistency maintenance mechanism is developed by leveraging EAD [14] file replication algorithm. The replication algorithm achieves an optimized trade-off between query efficiency and overhead in file replication. We introduce file replication component by addressing two main problems in file replication: 1) Where to replicate files so that the file query can be significantly expedited and the replicas can be fully utilized? 2) How to remove underutilized file replicas so that the overhead for consistency maintenance is minimized?

b) File Consistency maintenance

Maintaining consistency between frequently updated or even infrequently updated files and their replicas is a fundamental reliability requirement for a P2P system. P2P systems are characterized by dynamism, in which node join and leave continuously and rapidly. IRM employs adaptive polling for file consistency maintenance to cater to file replication dynamism.

In IRM poll-based consistency maintenance, each replica node polls its file owner or another node to validate whether its replica is the up-to-date file, and updates its replica accordingly. IRM addresses two main issues in consistency maintenance: 1) How to determine the frequency that a replica node probe a file owner in order to guarantee timely file update? 2) How to reduce the number of polling operations to save cost and meanwhile provide the fidelity of consistency guarantees?

IRM associates a time-to-refresh (TTR) value with each replica. It denotes the next time instant a node should poll the owner to keep its replica updated. The TTR value is varied dynamically based on the results of each polling. IRM combines file query rate into consideration for poll time determination. TTRquery and TTRpoll denotes the next time instant of corresponding operation of a file.

Algorithm 1. Pseudo-code for the IRM adaptive file consistency maintenance algorithm

```
//operation at time instant T_poll
if there is a query for the file then
   include a polling request into the query for file f
else
   send out a polling request
if get a validation reply from file owner then{
   if file is valid then
TTR = TTR_old + α
   if file is stale then{
   TTR = TTR_old / β
   update file replica}
if TTR > TTR_max or TTR < TTR_min then
  TTR = max(TTR_min, min(TTR_max, TTR))
if TTR ≤ T_query then
  TTR_poll = T_query
else
 TTR_poll = TTR}
```

When $TTR > T_{query}$, that is, the file is queried at a higher rate than change rate, then the file should be updated timely based on TTR. As a result, TTR_{poll} should be calculated based on the following formula:

$$TTR_{poll} = \begin{cases} T_{query} & TTR \leq T_{query}, \\ TTR & TTR > T_{query}. \end{cases}$$

IV. Performance Evaluation

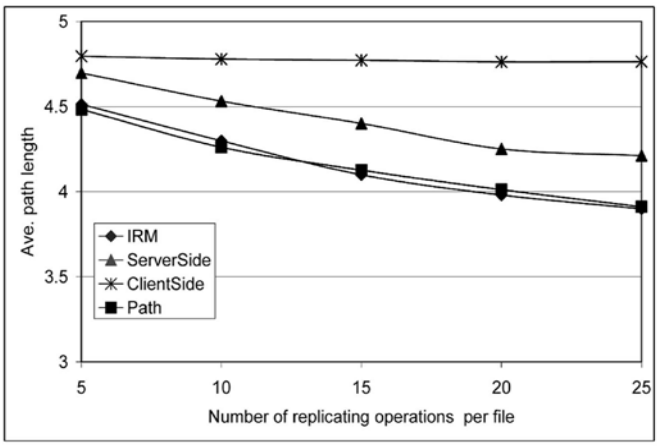
We designed and implemented a simulator for evaluating the IRM mechanism based on Chord P2P system [8]. We compared IRM with representative approaches of file replication and consistency maintenance. Experiment results show that IRM file replication algorithm is highly effective in reducing file query latency, the number of replicas, and file consistency maintenance overhead. IRM file consistency maintenance in turn provides a guarantee of file fidelity of consistency even in churn and dramatically reduces consistency maintenance overhead. Table 1 lists the parameters of the simulation and their default values. In practice, a node has various capacities in terms of bandwidth, memory storage, processing speed, etc. We assume that different capacities can be represented by one metric.

a) File Replication

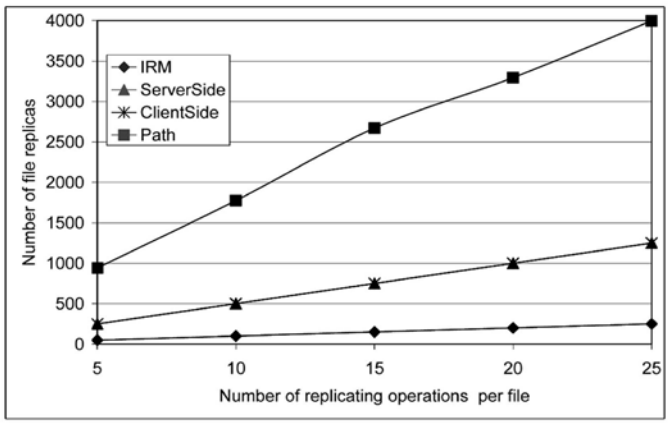
We choose the works in [2], [5], and [7] as representative works of the three categories of file replication proaches, Server Side, Client Side, and Path, respectively. We compared the performance of IRM with Server Side [2], Client Side [5], and Path [7] in terms of average lookup path length, hot spot reduction, and the total number of file replicas versus the number of replicating operations per file. In each replicating operation, IRM, Server Side and Client Side replicate a file to a single node, while Path replicates a file to a number of nodes along a query path.

Table 1 : Simulated Environment and Algorithm Parameters

Parameter	Default value
Object arrival location	Uniform over ID space
Number of nodes	4096
Node capacity c	Bounded Pareto: shape 2 lower bound: 500 upper bound: 50000
Number of queried files	50
Number of queries per file	1000
Number of replicating operations per file	5-25
T_l, T_q	2
Observation period	1 second
α	0.5
β	1.5



(a)



(b)

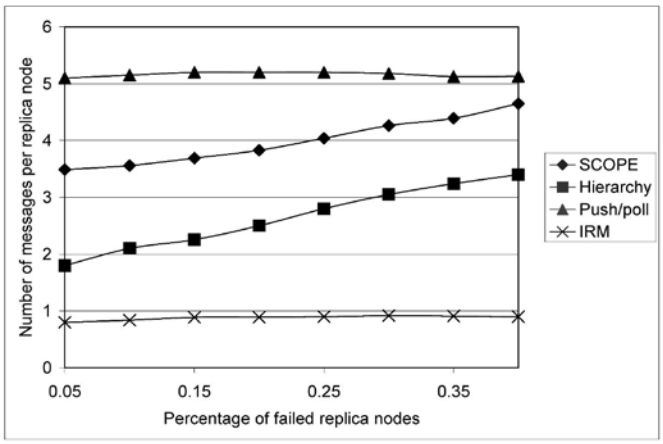
Fig. 3 : Performance of File Consistence maintenance algorithms
(a) *number of messages with churn (b) stale file responses with churn*

Fig. 3a plots the average path length of different approaches. We can see that Path generates shorter path length than Server Side and Client Side, and IRM leads to approximately the same path length as Path. Fig. 3b illustrates the number of replicas versus the number of replicating operations per file. The figure shows that the number of replicas increases as the number of replicating operations per file increases. This is due to the reason that more replication operations for a file lead to more replicas. The figure also shows that the number of replicas of Path is excessively higher than others. It is because in each file replication operation, a file is replicated in multiple nodes along a routing path in Path but in a single node in Server Side, Client Side, and IRM.

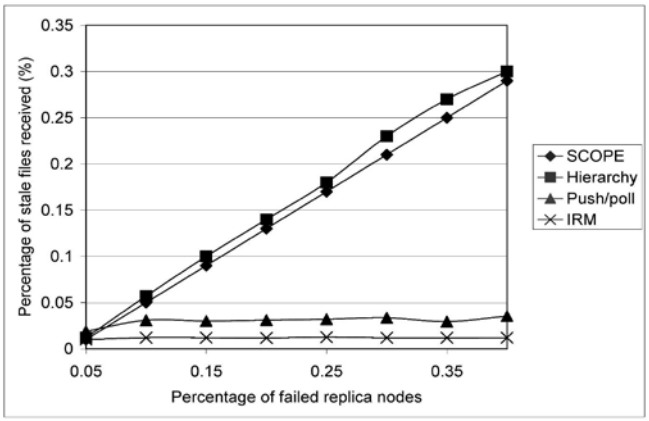
b) File Consistency Maintenance

We use Hierarchy to denote the work in [15] that builds a hierarchical structure for file consistency maintenance. We compared the performance of IRM with SCOPE [16], hierarchy [15], and Push/poll [17]

methods in terms of file consistency maintenance cost and the capability to keep the fidelity of file consistency. In Hierarchy, we set the number of nodes in a cluster to 16. We assumed four types of file: highly mutable, very mutable, mutable, and immutable. The percentage of the files in each category and their update rates were (0.5 percent, 0.15 sec), (2.5 percent, 7.5 sec), (7 percent, 30 sec), and (90 percent, 100 sec). File queries were successively generated. The query interval time was randomly chosen between 1 and 500 seconds.



(a)



(b)

Fig. 4 : Effectiveness of IRM in overhead reduction. (a) Update rate and (b) number of update messages.

This experiment evaluated the performance of different Consistency maintenance methods with churn in P2P systems. In the experiment, the number of replica nodes was set to 4,000 and the failed nodes were randomly chosen. Fig. 4a shows the average number of update messages per replica node versus the percentage of failed replica nodes. We can see that the number of update messages increases as the percentage of failed replica nodes increases in SCOPE and Hybrid, but remains constant in IRM and Push/poll. SCOPE constitutes nodes into a tree structure for file updating.

Fig. 4b depicts the percentage of stale files received by requesters versus the percentage of failed nodes. We can see that the percentages of stale files received in SCOPE and Hierarchy increase rapidly as the failed replica nodes grow, while the percentages of stale files received in IRM and Push/poll keep almost constant regardless of the percentage of failed replica nodes. The figure also demonstrates that SCOPE and Hierarchy incur much higher percentage rates than IRM and Push/poll. SCOPE relies on tree structure for update propagation, and if a node fails, all the node's children cannot get the update message in time until the tree is fixed.

V. Conclusion

This paper proposes an IRM that achieves high efficiency at a significantly lower cost. Instead of passively accepting replicas and updates, nodes autonomously determine the need for file replication and validation based on file query rate and update rate. It guarantees the high utilization of replicas, high query efficiency and fidelity of consistency. Meanwhile, IRM reduces redundant file replicas, consistency maintenance overhead, and unnecessary file updates. Simulation results demonstrate the effectiveness of IRM in comparison with other file replication and consistency maintenance approaches. Its low overhead and high effectiveness are particularly attractive to the deployment of large-scale P2P systems.

We find that IRM relying on polling file owners still cannot guarantee that all file requesters receive up-to-date files, although its performance is better than other consistency maintenance algorithms. We plan to further study and explore adaptive polling methods to fully exploit file popularity and update rate for efficient and effective replica consistency maintenance.

References Références Referencias

1. I. Clarke, O. Sandberg, B. Wiley, and T.W. Hong, "Freenet: A Distributed Anonymous Information Storage and Retrieval System," Proc. Int'l Workshop Design Issues in Anonymity and Unobservability, pp. 46-66, 2001.
2. A. Rowstron and P. Druschel, "Storage Management and Caching in PAST, a Large-Scale, Persistent Peer-to-Peer Storage Utility," Proc. ACM Symp. Operating Systems Principles (SOSP), 2001.
3. F. Dabek, M.F. Kaashoek, D. Karger, R. Morris, and I. Stocia,"Wide Area Cooperative Storage with CFS," Proc. ACM Symp. Operating Systems Principles (SOSP), 2001.
4. T. Stading, P. Maniatis, and M. Baker, "Peer-to-Peer Caching Schemes to Address Flash Crowds," Proc. First Int'l Workshop Peerto- Peer Systems (IPTPS), 2002.
5. V. Gopalakrishnan, B. Silaghi, B. Bhattacharjee, and P. Keleher, "Adaptive Replication in Peer-to-Peer

Systems," Proc. 24th Int'l Conf. Distributed Computing Systems (ICDCS), 2004.
6. Gnutella Home Page, http://www.gnutella.com, 2008. [7] R. Cox, A. Muthitacharoen, and R.T. Morris, "Serving DNS Using a Peer-to-Peer Lookup Service," Proc. First Int'l Workshop Peer-to-Peer Systems (IPTPS), 2002.
7. Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker, "Search and Replication in Unstructured Peer-to-Peer Networks," Proc. 16th Int'l Conf. Supercomputing (ICS), 2001.
8. E. Cohen and S. Shenker, "Replication Strategies in Unstructured Peer-to-Peer Networks," Proc. ACM SIGCOMM, 2002.
9. S. Tewari and L. Kleinrock, "Analysis of Search and Replication in Unstructured Peer-to-Peer Networks," Proc. ACM SIGMETRICS, 2005.
10. S. Tewari and L. Kleinrock, "On Fairness, Optimal Download Performance and Proportional Replication in Peer-to-Peer Networks," Proc. IFIP Networking, 2005.
11. S. Tewari and L. Kleinrock, "Proportional Replication in Peer-to-Peer Network," Proc. IEEE INFOCOM, 2006.
12. D. Rubenstein and S. Sahu, "Can Unstructured P2P Protocols Survive Flash Crowds?" IEEE/ACM Trans. Networking, vol. 13, no. 3, pp. 501-512, June 2005.
13. H. Shen, "EAD: An Efficient and Adaptive Decentralized File Replication Algorithm in P2P File Sharing Systems," Proc. Eighth Int'l Conf. Peer-to-Peer Computing (P2P '08), 2008.
14. G. Xie, Z. Li, and Z. Li, "Efficient and Scalable Consistency Maintenance for Heterogeneous Peer-to-Peer Systems," IEEE Trans. Parallel and Distributed Systems, vol. 19, no. 12, pp. 1695-1708, Dec. 2008.
15. X. Chen, S. Ren, H. Wang, and X. Zhang, "SCOPE: Scalable Consistency Maintenance in Structured P2P Systems," Proc. IEEE INFOCOM, 2005.
16. Datta, M. Hauswirth, and K. Aberer, "Updates in Highly Unreliable, Replicated Peer-to-Peer Systems," Proc. 23rd Int'l Conf. Distributed Computing Systems (ICDCS), 2003.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 11 Issue 16 Version 1.0 September 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Study & Analysis of Security Issues in Wireless Sensor Networks

By Payal Jain, Sameer verma
Maharishi Markandeshwar University, Mullana, Ambala Haryana

Abstract - Wireless sensor networks are successfully used in the conditions of war as well as natural calamities like earthquake, flood, volcanoes etc. Rapid technological advances in the area of micro electro-mechanical systems have spurred the development of small inexpensive sensors capable of intelligent sensing. A significant amount of research has been done in the area of connecting large numbers of these sensors to create robust and scalable Wireless Sensor Networks (WSNs). Proposed applications for WSNs include habitat monitoring, battlefield surveillance, and security systems. WSNs aim to be energy efficient, self-organizing, scalable, and robust. Relatively little work has been done on security issues related to sensor networks. The resource scarcity, ad-hoc deployment, and immense scale of WSNs make secure communication a particularly challenging problem. The primary consideration for sensor networks is energy efficiency, security schemes must balance their security features against the communication and computational overhead required to implement them. This paper will describe the fundamental challenges in the emergent field of sensor network security and the initial approaches to solving them.

Keywords : *Sensor network, Seismic, Message authentication code, Hopping, Spread spectrum etc.*

GJCST Classification : *C.2.1*

Strictly as per the compliance and regulations of:

Study & Analysis of Security Issues in Wireless Sensor Networks

Payal Jain[α], Sameer verma[Ω]

Abstract - Wireless sensor networks are successfully used in the conditions of war as well as natural calamities like earthquake, flood, volcanoes etc. Rapid technological advances in the area of micro electro-mechanical systems have spurred the development of small inexpensive sensors capable of intelligent sensing. A significant amount of research has been done in the area of connecting large numbers of these sensors to create robust and scalable Wireless Sensor Networks (WSNs). Proposed applications for WSNs include habitat monitoring, battlefield surveillance, and security systems. WSNs aim to be energy efficient, self-organizing, scalable, and robust. Relatively little work has been done on security issues related to sensor networks. The resource scarcity, ad-hoc deployment, and immense scale of WSNs make secure communication a particularly challenging problem. The primary consideration for sensor networks is energy efficiency, security schemes must balance their security features against the communication and computational overhead required to implement them. This paper will describe the fundamental challenges in the emergent field of sensor network security and the initial approaches to solving them.

Keywords : *Sensor network, Seismic, Message authentication code, Hopping, Spread spectrum etc.*

I. Introduction

Wireless Sensor Networks is composed of hundreds or thousands of inexpensive, low-powered sensing devices with limited computational and communication resources, provide a useful interface to the real world with their data acquisition and processing capabilities. Applications include burglar alarms, inventory control, medical monitoring and emergency response monitoring remote or inhospitable habitats, target tracking in battle fields, disaster relief networks early fire detections in forest and environmental monitoring. Sensor devices, also called motes or nodes, typically consist of a sensing unit, a transceiver unit, a processing unit, and a power source unit. Depending on the application, the sensing unit may monitor various types of data including acoustic, seismic, visual, and temperature data. The transceiver unit is a low-power radio capable of short range communication (tens of meters). The processing unit contains memory and a processor with severely limited size and speed. Wireless sensor motes are powered by a battery energy source which is not intended to be recharged. Communication usually consists of source nodes which sense the data and return it to sink nodes over multiple hops. Sink nodes may be ordinary sensor nodes or specialized base stations with greater resources.

In the future thousands to millions of sensor devices will be embedded in almost every aspect of life. The main aim is create an intelligent environment which is capable of collecting massive amounts of information, recognizing significant events automatically and responding appropriately. Sensor networks facilitate "large-scale, real-time data processing in complex environments" [Wood and Stankovic 2002].

Although two of the most security-orientated applications of WANs are military and medical solutions. Sensor networks can be applied to a large number of areas and its applications are continuously growing. Sensor networks are extremely vulnerable against any type of internal or external attacks, due to resource constraints, lack of tamper-resistant packaging, and the nature of its communication channels.

If sensor networks are to attain their potential, however, secure communication techniques must be developed in order to protect the system and its users. WSNs are ideal for detecting chemical, biological, or environmental threats over large areas, but maliciously induced false alarms could completely negate the value of the system. As [1] point out, if security is weak, sensor networks "will only be suitable for limited, controlled environments – falling far short of their promise." The widespread deployment and overall success of sensor networks will be directly related to their security strength.

II. Security Issues and Goals

a) Data confidentiality

Confidentiality means keeping information secret from unauthorized parties. A sensor network should not leak sensor readings to neighboring networks. The standard approach for keeping sensitive data secret is to encrypt the data with a secret key that only intended receivers possess, hence achieving confidentiality. Most of the proposed protocols use symmetric key encryption methods.

Author α : Lecturer, Department of Information and Technology, Maharishi Markandeshwar College of Engineering Maharishi Markandeshwar University, Mullana, Ambala Haryana.
E-mail : payaljain2006@gmail.comb, Mobile No: 91-9466742552.
Author Ω : Lecturer, Department of Information and Technology, Maharishi Markandeshwar College of Engineering Maharishi Markandeshwar University, Mullana, Ambala, Haryana.
E-mail : sameer. verma1986@gmail.com, Mobile no: +919416320512.

b) Data authenticity

In a sensor network, an adversary can easily inject messages, so the receiver needs to make sure that the data used in any decision-making process originates from the correct source. Data authentication prevents unauthorized parties from participating in the network and legitimate nodes should be able to detect messages from unauthorized nodes and reject them. In Data authentication can be achieved through a purely symmetric mechanism: The sender and the receiver share a secret key to compute a message authentication code (MAC) of all communicated data. When a message with a correct MAC arrives, the receiver knows that it must have been sent by the sender. However, authentication for broadcast messages requires stronger trust assumptions on the network nodes.

c) Data integrity

Data integrity ensures the receiver that the received data is not altered in transit by an adversary.

d) Data freshness

Data freshness implies that the data is recent, and it ensures that an adversary has not replayed old messages. A common defense is to include a monotonically increasing counter with every message and reject messages with old counter values. With this policy, every recipient must maintain a table of the last value from every sender it receives. For RAM constrained sensor nodes, this defense becomes problematic for even modestly sized networks. Assuming nodes devote only a small fraction of their RAM for this neighbor table, an adversary replaying broadcast messages from many different senders can fill up the table. At this point, the recipient has one of two options: ignore any messages from senders not in its neighbor table, or purge entries from the table. Neither is acceptable; the first creates a DOS attack and the second permits replay attacks.

The protection against the replay of data packets should be provided at the application layer and not by a secure routing protocol as only the application can fully and accurately detect the replay of data packets (as opposed to retransmissions, for example). The reason that by using information about the network's topology and communication patterns, the application and routing layers can properly and efficiently manage a limited amount of memory devoted to replay detection.

There are two types of freshness identified: weak freshness, which provides partial message ordering, but carries no delay information, and strong freshness, which provides a total order on a request-response pair, and allows for delay estimation. Weak freshness is required by sensor measurements, while strong freshness is useful for time synchronization within the network.

e) Robustness and Survivability

The sensor network should be robust against various security attacks, and if an attack succeeds, its impact should be minimized. The compromise of a single node should not break the security of the entire network.

III. Threats to WSNS

There are a large and increasing number of threats and attacks to which WSNs are susceptible. They can be broadly classified as attacks against the privacy of the network data, denial of service (DOS) attacks, impersonation or replication attacks and physical attacks. A denial of service (DoS) attack aims to deny access to legitimate users to shared services or resources. Attacks can be launched at any point in the network. This implies that certain attacks may be more effective at different layers of the communication protocol,

Table I : sensor network layer and attack

Layer	Attack
Physical Layer	DOS – Jamming, Tampering
Data-link Layer	DOS – Collision, Exhaustion, Unfairness
Network Layer	DOS – Neglect & Greed, Homing, Misdirection (Spoofing), Black Holes, Flooding Sybil Wormhole Attack

a) Physical layer

Attacks at the physical level include radio signal jamming and tampering with physical devices.

i. *Jamming*

Jamming is interference with the radio frequencies used by a device's transceiver. It represents an attack on the availability of a network. Jamming is only different from normal radio propagation in that it is unwanted and disruptive, thus creating a denial-of-service condition **[4]**. The degree of the jamming is determined by physical properties such as the available power, antenna design, obstacles, and height above ground **[4]**. This attack is extremely effective against single frequency networks.

Defense against jamming involves the use of spread-spectrum or frequency hopping techniques. Spread-spectrum communication uses a wider band for radio transmission. Frequency hopping is a type of spread-spectrum in which a pseudorandom sequence is used to change the frequency of transmission. The receiver, who also knows the hopping sequence, can "dehop" the signal to reconstruct the original message **[2]**. Frequency hopping also protects against unintentional jamming, i.e., interference. Spread-

spectrum techniques provide protection in high noise environments in which sensor networks will certainly be deployed. The inherent complexity involved in spread-spectrum systems is particularly costly for sensor motes. Frequency hopping requires greater power and financial cost, two scarce resources in sensor networks.

Prevention of denial of service attacks is a difficult task. Since most sensor networks currently use single frequency communication, [4] have proposed a Jammed Area Mapping (JAM) service which emphasizes detection and adaptation in response to jamming. They assume that only a portion of the network is being jammed and attempt to map this area so it can be avoided. Nodes in the affected area switch to low power mode. Information about jammed areas is passed to the network layer so it can successfully route packets around the dead areas. If spread spectrum techniques cannot be incorporated into motes, then detection algorithms such as JAM may be important in defending against jamming attacks.

ii. *Tampering*

A second problematic issue at the physical layer is the relative ease and potential harm of device tampering. This problem is exacerbated by the large-scale, ad-hoc, pervasive nature of sensor networks. Access to thousands of nodes spread over several kilometers cannot be completely controlled [4]. Attackers may very well have greater physical access to nodes than the network administrator. Nodes may be captured, interrogated, and compromised without difficulty.

One defense involves physically tamper-proofing the devices. Nodes should react to tampering by erasing sensitive cryptographic information [Wood and Stankovic 2002]. However, tamper-resistant packaging increases the cost of the devices, thus reducing their economic viability. The preferred solution is algorithmic: algorithms that reduce the effect a single key compromise has on the security of the entire network. The tampering is "one of the most vexing problems in sensor network security",[5].

b) LInk layer

The link and media access control (MAC) layer handles neighbor-to-neighbor communication and channel arbitration. Like the physical layer, the link layer is particularly susceptible to denial of service attacks.

i. *Collision*

If an adversary can generate a collision of even part of a transmission, he can disrupt the entire packet [[15], Stankovic, and Wagner 2004]. A single bit error will cause a CRC mismatch and possibly require retransmission. In some MAC protocols, a corrupted ACK may cause exponential back-off and unnecessarily increase latency. Although error-correcting codes protect against some level of packet corruption, intentional corruption can occur at levels which are beyond the encoding scheme's ability to correct. The advantage, to the adversary, of this MAC level jamming over physical layer jamming is that much less energy is required to achieve the same effect: preventing devices from successfully transmitting packets.

ii. *Exhaustion*

Another malicious goal is the exhaustion of a network's battery power. In addition to the previous types of attacks, exhaustion may also be induced by an interrogation attack. In the IEEE 802.11-based protocols, for example, Request To Send (RTS) and Clear To Send (CTS) packets are used to reserve bandwidth before data transmission. A compromised node could repeatedly send RTS packets in order to elicit CTS packets from a targeted neighbor, eventually consuming the battery power of both nodes.

iii. *Unfairness*

A more subtle goal of the previously described attacks may be unfairness in the MAC layer [3]. A compromised node can be altered to intermittently attack the network in such a way that induces unfairness in the priorities for granting medium access. This weak form of denial of service might, for example, increase latency so that real-time protocols miss their deadlines, [3]. Another form of this attack could target one particular flow of data in order to suppress detection of some event. The use of small frames which prevent a node from capturing the channel for a long period of time has been proposed as a defense against this sort of attack [1].

c) Network Layer

The network layer is responsible for routing packets across multiple nodes. Due to the ad-hoc nature of sensor networks, every node must assume routing responsibilities. WSNs are particularly vulnerable to routing attacks because every node is essentially a router. [3] have identified a variety of routing attacks and have shown them to be effective against every major sensor network routing protocol. Their classifications of attacks are summarized below and are followed by a general discussion of secure routing techniques.

i. *False Routing Information*

The most direct attack on routing is to spoof, alter, or replay routing information. This false information may allow adversaries to create routing loops, attract or repel traffic, shorten or extend route lengths, increase latency, and even partition the network [3]. Clearly, the falsification of routing information can cripple a network. The standard solution is to require authentication for routing information, i.e., routers only accept routing information from valid routers. Not surprisingly, authentication is an important element in the security systems proposed for sensor networks.

ii. *Selective Forwarding*

Selective forwarding is a more subtle attack in which some packets are correctly forwarded but others are silently dropped. A compromised node could be configured to drop all packets, creating a so-called black hole. Since the network is capable of handling node failure it may conclude that the compromised node has failed and find another route. If the compromised node selectively forwards packets, the neighboring nodes will believe that the malicious node is still functioning correctly and continue to route packets to the node. This vulnerability is due to the assumption that nodes will faithfully forward received messages. If an attacker can get in the path of a desired data flow, he can selectively drop packets from that flow.

iii. *Sinkhole Attack*

The adversary's goal is to lure nearly all the traffic from a particular area through a compromised node, creating a metaphorical sinkhole with the adversary at the center. Sinkhole attacks typically work by making a compromised node look especially attractive to surrounding nodes with respect to the routing algorithm. For instance, an adversary could spoof or replay an advertisement for an extremely high quality route to a base station. Due to either the real or imagined high quality route through the compromised node, it is likely each neighboring node of the adversary will forward packets destined for a base station through the adversary, and also propagate the attractiveness of the route to its neighbors. Effectively, the adversary creates a large "*sphere of influence*", attracting all traffic destined for a base station from nodes several hops away from the compromised node.

iv. *The Sybil Attack*

In a Sybil attack, a single node presents multiple identities to other nodes in the network. They pose a significant threat to geographic routing protocols, where location aware routing requires nodes to exchange coordinate information with their neighbors to efficiently route geographically addressed packets. Authentication and encryption techniques can prevent an outsider to launch a Sybil attack on the sensor network. However, an insider cannot be prevented from participating in the network, but (s)he should only be able to do so using the identities of the nodes (s)he has compromised. Using globally shared keys allows an insider to masquerade as any (possibly even nonexistent) node. Public key cryptography can prevent such an insider attack, but it is too expensive to be used in the resource constrained sensor networks. One solution is to have every node share a unique symmetric key with a trusted base station. Two nodes can then use a Needham-Schroeder like protocol to verify each other's identity and establish a shared key. A pair of neighboring nodes can use the resulting key to implement an authenticated, encrypted link between them. An example of a protocol which uses such a scheme is LEAP, which supports the establishment of four types of keys.

v. *Wormhole Attack*

The wormhole attack is used to convince two possibly distant nodes that they are neighbors so that the attacker can place himself on the route between them. Basically, the adversary tunnels messages from one part of the network to another through an out-of-bound channel available only to the attacker. Wormholes typically involve two colluding nodes. This sort of attack is likely to be used in combination with selective forwarding or eavesdropping.

vi. *Hello Flood Attack*

The Hello flood attack, a novel attack proposed by Karlof and Wagner, exploits routing protocols that require periodic HELLO packets be transmitted to announce the presence of a node. Nodes which receive a HELLO packet assume they are within radio range of the sender, i.e., the sender is a neighboring node. This assumption may be false in the case of a laptop-class attacker. An adversary with a powerful transmitter may be able to transmit a single HELLO packet to every node in the network and convince every node that it is a one-hop neighbor. As a result, the network is left in a state of confusion. If, for example, the attacker advertises a very quick route to a base station in the HELLO packet, many non-neighbor nodes will attempt to route packets through the malicious node. In actuality, however, they will be sending packets into oblivion. [3] point out that this attack is actually a "one-way, broadcast wormhole." The simplest solution for this attack is to verify the bidirectionality of a link before acting on its information. Essentially, routing messages from one-way links are ignored. Karlof and Wagner propose an identity verification protocol to defend against the HELLO flood attack.

vii. *Acknowledgement Spoofing*

The last routing attack [3] identify is the acknowledgement spoofing attack. Several routing protocols rely on link layer acknowledgements for determining next-hop reliability. If an adversary can respond for weak or dead nodes, he can deceive the sender about the strength of the link and effectively mount a selective forwarding attack. The artificial reinforcement allows the attacker to manipulate the routing through the weak or dead node.

There have been several approaches to defend against network layer attacks. Authentication and encryption are a first step, but more proactive techniques such as monitoring, probing, and transmitting redundant packets have also been suggested. Secure routing methods protect against some of previous attacks. Proposed techniques are described below.

viii. *Authentication & Encryption*

Link layer authentication and encryption protect against most outsider attacks on a sensor network routing protocol. Even a simple scheme which uses a globally shared key will prevent unauthorized nodes from joining the topology of the network. In addition to preventing selective forwarding and sinkhole attacks, authentication and encryption also make the Sybil attack impossible because nodes will not accept even one identity from the malicious node **[3]**. SPINS and TinySec are two proposed solutions for link level encryption and authentication. They are discussed in greater detail in the next section.

ix. *Monitoring*

A more active strategy for secure routing is for nodes to monitor their neighbors and watch for suspicious behavior **[1].** In this approach, nodes act as "watchdogs" to monitor the next hop transmission of the packet. In the event that misbehavior is detected, nodes will update routing information to avoid the compromised node.

x. *Probing*

Another proactive defense against malicious routers is probing **[1].** This method periodically sends probing packets across the network to detect blackout regions. Since geographic routing protocols have knowledge of the physical topology of the network, probing is especially well-suited to their use. Probes must appear to be normal traffic, however, so that compromised nodes do not intentionally route them correctly in order to escape detection.

xi. *Redundancy*

Redundancy is another strategy for secure routing [**1**]. An inelegant approach, redundancy simply transmits a packet multiple times over different routes. Hopefully, at least one route is uncompromised and will correctly deliver the message to the destination. Despite its inefficiency, this method does increase the difficulty for an attacker to stop a data flow.

IV. Conclusion

While the majority of the research in sensor networks has focused on making them feasible and useful, a few researchers have proposed solutions to the security issues discussed previously. Sensor network security mechanisms can be divided into two categories: communication protocols and key management architectures. Communication protocols deal with the cryptographic algorithms used to achieve availability, confidentiality, integrity, and authentication. Key management architectures handle the complexities of creating and distributing keys used by communication protocols.

References References Referencias

1. A.D. Wood and J.A. Stankovik, (2002) "Denial of service in Sensor Networks", Computer, Vol. 35, No. 10, 2002, pp 54-62.
2. H. Chen, Q-A. Zeng, and D. P. Agrawal, "A Novel Optimal Channel Partitioning Algorithm for Integrated Wireless and Mobile Networks," accepted for the Journal of Mobile Communication, Computation and Information.
3. Karlof and Wagner, Secure Routing in Wireless Sensor Networks, "Attacks and Countermeasures", (SNPA 2003), pages 113-127, May-2003.
4. A.Wood, J. Stankovik, S.Son, "Jam : A mapping service for jammed regionsnin sensor networks, RTTS, Mexico, Dec, 2003.
5. Adrian Perrig, John Stankovik and David Wagner (2004), "Security in wireless Sensor Networks", Communication. ACM, 47(6): 53-57.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 11 Issue 16 Version 1.0 September 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Reengineering of Module for Public Sector & Complexity Measurement

By Ashok Kumar, Anil Kumar
Kurukshetra University, Kurukshetra, India

Abstract - This paper is based on reengineering of module for public sector, it deals with the measurement of complexity as well as effort measurement of module during Reengineering of module at design time. This methodology reduces more than 75% resources as compared to conventional and structural Methodology. It also, enables to reengineering of module faster, high quality, high reliability, and also increases level of reusability & productivity.

Keywords : *S Conventional Methodology [8], Structural Methodology [8], Excel Template [9].*

GJCST Classification : *D.2.8*

Strictly as per the compliance and regulations of:

Reengineering of Module for Public Sector & Complexity Measurement

Ashok Kumar[α], Anil Kumar[Ω]

Abstract - This paper is based on reengineering of module for public sector, it deals with the measurement of complexity as well as effort measurement of module during Reengineering of module at design time. This methodology reduces more than 75% resources as compared to conventional and structural Methodology. It also, enables to reengineering of module faster, high quality, high reliability, and also increases level of reusability & productivity.

Keywords : Conventional Methodology [8], Structural Methodology [8], Excel Template [9]

I. Introduction

Today public sector is an integral part of Govt. and performance of it, has considerable scope for improvement. However, these sector such as govt. hospital, depends on information system, which have been engineered in earlier days, such legacy system using procedural methodology, db handling , GUI etc. As services grow in size or the requirement of public increases continuously, due to this there is accelerating need software maintenance. It has been observed that, the cost of the maintenance is much higher than the cost of reengineering of the software. And also continue maintenance of such system become tedious and cost approach and occurrence of software failure is more due to poor documentation, poorly structured & transparency, and also changes technology infra structured in hardware and software, complexity of module increases continuously, and finally logic code written is outdated hardware and software. Therefore, maintenance is not a good choice. Reengineering, is much better than maintenance. It is an approach to solve problem of legacy system. Its aim is the qualitative improvement of existing software and the extension of its life expectancy. It consists of examination (reverse engineering) and alteration (forward engineering) of legacy system.

II. Purposed Work

The purposed methodology, used to reengineering of module of public sector i.e. more suitable, for available tools and techniques. It will create significant improvement to measure the complexity and effort of module individually by using Excel Template[9]. The Excel Template that is used to measure complexity of each and every modules of hospital are Modified Method Hiding Factor (M-MHF), Modified Attributes Hiding Factor (M-AHF), Modified Method Inheritance Factor (M-MIF), Modified Attributes Inheritance Factor (M-AIF), Modified Coupling Factor (M-CF), and Modified Polymorphism Factor (M-PF); And the Excel Template that are used for measuring effort of each and every modules of hospital during reengineering of the modules are, Modified Weighted Method Per Class (M-WMC), Modified Depth of Inheritance (M-DIT), Modified Number of Children (M-NOC), Modified Coupling Between Object (M-CBO), Modified Response for a class (M-RFC) and Modified Lack of Cohesion in Method (M-LCOM).

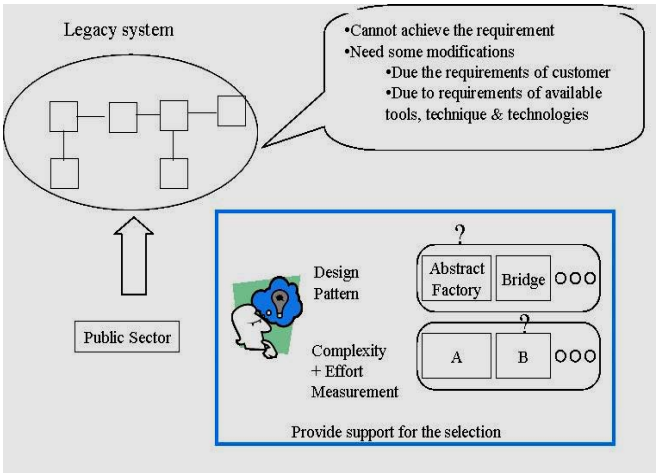
It also specifies causes of errors and the use of the safety design concepts, to prevent minimize errors by detecting them, before undesirable effect takes place. The excel template provides facility to reengineering the modules in such a way that help enables the doctors to better serve their patients, Reducing the time spent by staff filling out forms, Control over the costs incurred by diagnosis – related groups, Increased nursing productivity, Faster and informed decision-making by doctors, Improve decision support for the management, Cost-effective patient transactions.

III. Problem Description

HMS is powerful, flexible, easy to use and has designed & developed to deliver real conceivable benefits to hospitals and clinics. It is designed for multi specialty hospitals, to cover a wide range of hospital administration and management processes.

Hospital Management System is a product of our deep experience in delivering successful solutions to various customers in the healthcare space and our expertise in developing unique Intellectual Property in the form of products and re-usable components for the Healthcare Industry.

Author α : Professor, Department of Computer Science & Application Kurukshetra University, Kurukshetra, India.
E-mail : Anilbest2005@gmail.com
Author Ω : Asst. Professor, Computer Science & Engg. Vaish College of Engineering, Rohtak, India. E-mail : Bestanil2005@rediffmail.com



The legacy system (i.e. hospital) is engineered by using Conventional and Structured Methodology. Conventional methodology, based on SDLC, there is no way to measurement of complexity and effort of module during reengineering of module as well as this methodology not support reusability and also productivity of module not very much effective.

Structured Methodology is slightly improvement of conventional methodology. If we reengineering the module by using this methodology, it help to measure control but not support reusability, but help in productivity and quality of analysis and design. It will provide more effective analysis & more stable or maintainable design. However, both these methodology not support today's available tool and techniques.

IV. Results and Discussion

There are twelve excel template that are used to determine complexity of module that are more efficient as compared to other methodologies. Six excel template [9] such as M-MHF, M-AHF, M-MIF, M-AIF, M-PF & M-CF are used to determine complexity of each and every module of the system, as well as it also provides facilities to hide information, to increase reusability & productivity of modules, measure the degree of method overriding in class inheritance and also measure degree of coupling among different types of modules.

Other six excel template[9], such as M-WMC, M-DIT, M-NOC, M-CBO, M-RFC & M-LCOM, are used to determine effort i.e. required to reengineering of the module during Post Martem Methodology[1 2,3].

Table 1 : 'Complexity measurement values i.e. determined by Complexity Measurement Template'

Sr No	Activity	Post Martem Methodology	Conventional methodology
1	HMS Staff	36.5141	60
2	Emergency	70.47	85
3	Enquiry	46.806	78
4	OPD	52.9121	75
5	Managing Unit Doctor	51.8	77
6	Examination	33.1916	56
7	Nurse Detail	36.914	62
8	Patient Status	20.4	56
9	Pharmacy/Drug	61.2712	80
10	Laundry	33.4272	75
11	Kitchen	33.4272	75

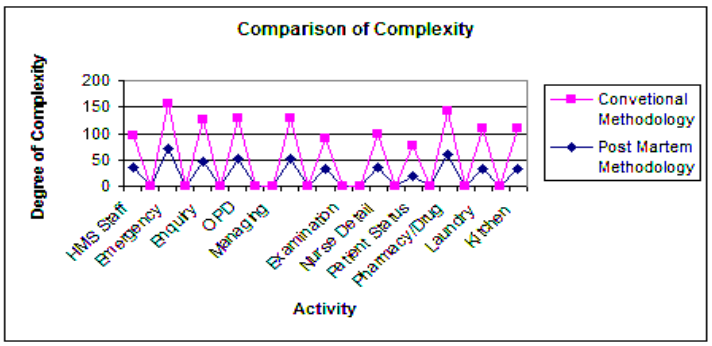


Fig.1 : 'The above graph show complexity of reengineering of module by using values given in table 1

Table 2 : 'Effort Measurement values are given below i.e. determined by using effort measurement template'

Sr No	Activity	M-WMC	M-DIT	M-NOC	M-CB0	M-RFC	M-LCOM
1	HMS Staff	3	4	3	2	2	3
2	Emergency	3	9	3	1	2	4
3	Enquiry	3	3	4	2	2	4
4	OPD	2	20	4	2	2	4
5	Managing Unit	5	20	9	4	4	5
6	Doctor Exam	2	8	3	1	2	4
7	Nurse Detail	2	9	3	1	2	4
8	Patient Status	3	4	5	2	2	2
9	Pharmacy/Drug	4	14	7	3	2	4
10	Laundry	1	2	1	1	2	2
11	Kitchen	1	2	1	1	2	2

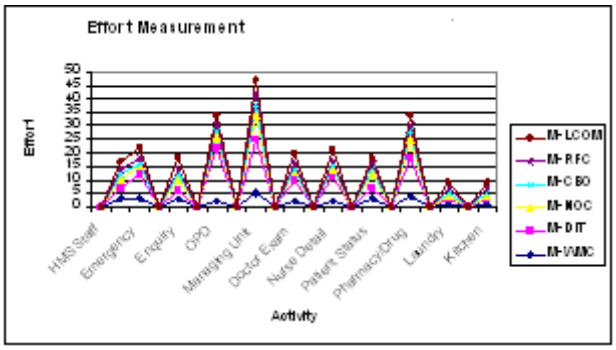


Fig. 2 : 'The above graph show effort required during reengineering of module by using value given in table 2'

The design of modules based on purposed methodology provides facilities such as:

- It enables the doctors to better serve their patients.
- Reducing the time spent by staff filling out forms.
- Control over the costs incurred by diagnosis – related groups.
- Increased nursing productivity.
- Faster and informed decision-making by doctors
- Improve decision support for the management
- Cost-effective patient transactions

The purposed methodology also, allows the developer to communicate using well-known, well understood names for the software interactions. Common design pattern can improved over time, making them more robust than ad-hoc (in-formal or unplanned) design.

V. Conclusion

Overall objective of this paper, is that modules are design in such a way that if any time any where any module need for reengineering in future, it is easily takes place. As well as it provides facilities to determine complexity[3,4] and effort from that module, where reengineering happens. It does not need to determine complexity of entire modules again and again. And it will focus on optimization and increase productivity[7], reusability[7], flexibility[7], understandability and also support reliability of modules[10].

References Références Referencias

1. Wastell G.W., White P. and Kawalek P. (1994). A methodology for business redesign: experience and issues. Journal of Strategic Information Systems **3**(1) 5-22.
2. Kettinger W.J., Teng J.T.C. and Guha S. (1997) Business process change: a study of methodologies, techniques, and tools. MISQ Quarterly March 55-80.
3. Kinny, D; Georgeff, M. and Rao, A. (1996) A methodology and modeling technique for systems of BDI agents. LNAI, vol. 1038, Springer Verlag.
4. Elaine J. Weyuker, *Evaluating Software Complexity Measures*, IEEE Transactions on Software Engineering, Vol. 14, Issue 9, pp. 1357-1365, 1988.
5. Yin-Farn Chen, Michael Y. Nishimoto, and C. V. Ramamoorthy, "The C Information Abstraction System," *IEEE Transactions on Software Engineering*. Vol. 16, No. 3, March 1990, pages 325–334
6. William B. Frakes and Thomas P. Pole, "An Empirical Study of Representation Methods for Reusable Software Components," *IEEE Transactions on Software Engineering*. Vol. 20, No. 8, August 1990, pages 617–630
7. Software Engineering Standards Committee of the IEEE Computer Society, *IEEE Standard for a Software Quality Metrics Methodology*, IEEE Std 1061-1998, 1998
8. Tenner, A. R., and Detoro, I. J., 1992. "The process redesign-the implementation guide for managers", Prentice Hall, New Jersey
9. Lionel C. Briand, Victor R. Basili, and Christopher J. Hetmanski. Developing interpretable models with optimized set reduction for identifying high-risk software components. *IEEE Transactions on Software Engineering*, 19(11):1028–1044, November 1993.
10. Muthu, S., Whitman, L. and Cheraghi, H. S., 1999. "Business process reengineering: a consolidated methodology", Proceedings of the 4th Annual International Conference on Industrial Engineering Theory, Applications and Practice.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 11 Issue 16 Version 1.0 September 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Secure Authentication & Key Establishment protocol with perfect Forward Secrecy for Multi and Broad cast service in IEEE 802.16e

By A.K.M. Nazmus Sakib,Fariha Tasmin Jaigirdar, Samiur Rahman, Tanvir Mahmud ,MuhammadMushfiqur Rahman
Chittagong University of Engineering and Technology

Abstract - Many complicated authentication and encryption techniques have been embedded into WiMAX but it still facing a lot of challenging situations. This paper shows that, GTEK Hash chain algorithm for Multi and Broadcast service of IEEE 802.16e facing a reduced forward secrecy problem. These vulnerabilities are the possibilities to forge key messages in Multi- and Broadcast operation, which are susceptible to forgery and reveals important management information. In this paper, we also propose three UAKE protocols with PFS (Perfect Forward Secrecy) that are efficient and practical for mobile devices.

Keywords : *Multi and Broadcast Service, IEEE 802.16e, Perfect Forward Secrecy, Authentication, Key Establishment, Hash function.*

GJCST Classification : *C.2.0*

Strictly as per the compliance and regulations of:

Secure Authentication & Key Establishment protocol with perfect Forward Secrecy for Multi and Broad cast service in IEEE 802.16e

A.K.M. Nazmus Sakib[α],Fariha Tasmin Jaigirdar[Ω], Samiur Rahman[β], Tanvir Mahmud[Ψ],Muhammad Mushfiqur Rahman[¥]

Abstract - **Many complicated authentication and encryption techniques have been embedded into WiMAX but it still facing a lot of challenging situations. This paper shows that, GTEK Hash chain algorithm for Multi and Broadcast service of IEEE 802.16e facing a reduced forward secrecy problem. These vulnerabilities are the possibilities to forge key messages in Multi- and Broadcast operation, which are susceptible to forgery and reveals important management information. In this paper, we also propose three UAKE protocols with PFS (Perfect Forward Secrecy) that are efficient and practical for mobile devices.**

Keywords : Multi and Broadcast Service, IEEE 802.16e, Perfect Forward Secrecy, Authentication, Key Establishment, Hash function.

I. Introduction

The Multicast and Broadcast service offers the possibility to distribute data to multiple M.S. with one single message. This saves cost and bandwidth. Broadcasted messages in IEEE 802.16e are encrypted symmetrically with a shared key [1]. Every member in the group knows the key & can decrypt the traffic. Message authentication is also based on the same shared key. This algorithm contains the vulnerability that every group member, besides decrypting and verifying broadcast messages, can also encrypt and authenticate messages as if they originate from the legitimate B.S [1, 3, 4, 5]. Another aspect which is much more problematic is the distribution of the traffic encryption keys (GTEKs), when the optional Multicast and Broadcast Rekeying Algorithm (MBRA) is used [6]. To transfer a GTEK to all group members it is broadcasted but encrypted with the key encryption key (GKEK). Due to broadcasting, the GKEK must also be a shared key and every group member knows it [1]. Thus are adversary group member can use it to generate valid encrypted and authenticated GTEK key update command messages & distribute an own GTEK [1]. Every group member would establish the adversary's key as a valid next GTEK. [1] Subsequently all traffic sent by the legitimate B.S can no longer be decrypted by the M.S. From M.Ss point of view only traffic from the adversary is valid. To force M.Ss to establish the adversary's key, there are several possibilities; If the implementation does not work properly, the key from the latter of two subsequently sent GTEK update command messages may overwrite the former one. Hence, the adversary just has to send its GTEK update command message after the B.S broadcasted a key update message. If the implementation follows the standard, the keys of both messages are accepted [1]. To be sure the M.S will not establish the legitimate B.Ss key; an intruder could forge some part of the B.Ss GTEK update command message [1]; Such a changed message would not be verified as correct and discarded by the M.Ss. After this, the adversary can send its own GTEK update command message which will be accepted [1, 7]. In a unicast connection, this different keying material at the mobile station would be detected as the B.S cannot decrypt data sent by the M.S. This result in a TEK invalid message destined to the M.S which subsequently refreshes its keying material [1]. Since the M.Bs is only unidirectional so; the B.S unable to detect that M.S has different GTEKs.

II. Shared Key in Multicast and Broadcast Service

A shared key cannot be used as every group member can forge messages when having the current symmetric keys [1]. Instead the GTEK update command message could be sent to each M.S in a unicast way like the GKEK update command message [1]. The key should then be encrypted with the M.S related KEK which is only known by this individual M.S. The BS sends the GTEK update command message by itself

Author α : A.K.M. Nazmus Sakib completed his BSc in Computer Science & Engineering from Chittagong University of Engineering and Technology. His research area is security issues analysis and solutions. Telephone: +880-1730079790 , +8801917884634

E-mail : sakib425@gmail.com

Author Ω : Fariha Tasmin Jaigirdar completed his M.S from BUET. She is a Lecturer of Stanford University Bangladesh.

E-mail : farihajaigirdar@yahoo.com

Author β : Samiur Rahman completed his B.Sc in Computer Science and Engineering from Chittagong University of Engineering and Technology. His research is in the field of security analysis and solutions. Telephone: +880-1720085936

E-mail : sami_mania@ymail.com

Author Ψ: Tanvir Mahmud completed his B.Sc in C.S.E from Chittagong University of Engineering & Technology.

E-mail : tanvir_cuet@yahoo.com

Author ¥ : Muhammad Mushfiqur Rahman, United International University, Dept. of C.S.E. E-mail : mushfiq.razib@gmail.com

when the current key's lifetime is going to expire [1]. The Fig.1 shows this. Another solution is the use of public key cryptography. Here, the GTEK update command message remains broadcasted and encrypted with the shared key GKEK but is additionally signed by an asymmetric signature [1]. M.Ss receiving a GTEK update command message can verify the signature of the B.S and subsequently decrypt the GTEK with the shared GKEK [1]. The Fig.2 shows this method together with the unicasted GKEK update command message.

A third possibility is to generate GTEKs as part of a one way hash chaining function (Fig. 3). Here the B.S has to generate a random number which represents the initial key GTEK0 [1]. Then the other GTEKs are generated by applying a one way hash function to previous GTEKs respectively. This is iterated n times.

GTEK0 = random ()
GTEK1 = f (GTEK0)
GTEK2 = f (GTEK1)
GTEKn = f (GTEKn-1)

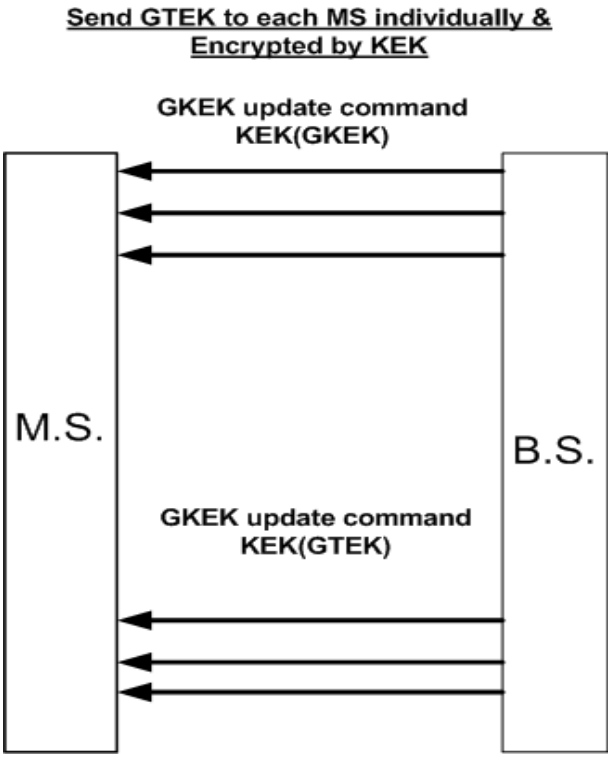


Fig.1 : Possible solution to transmit GTEK in a secure Way

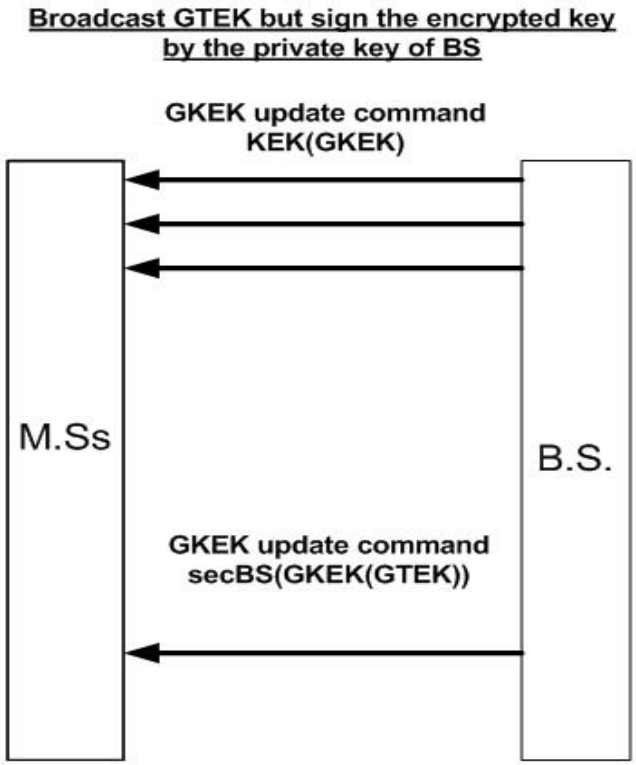


Fig.2 : Possible solution to transmit GTEK in a secure Way

To apply this algorithm, the key GKEK update command message has to be capable of transporting GKEK and GTEK keys together [1]. The design of the key update command message already includes both keys so only a little modification is needed here. Additionally the GTEK state machine at B.S must generate the GTEK hash chain & store all the keys. The GTEK state machine at M.S must add the functionality to authenticate GTEK keys by calculating the hash function and comparing it to the previous key [1]. A drawback of this algorithm is that it has a reduced forward secrecy [1]. This means a M.S joining the group can decrypt all broadcasted data since the last hash chain generation. If forward secrecy is crucial, the hash chain has to be regenerated each time a M.S enters the group [1]. When using an asymmetric signature or a hash chain to authenticate the GTEK transfer, only one message is needed to update the keys of all M.S due to broadcasting [1]. Thus the introduced traffic in these solutions is constant and does not depend on the number of members in the group [1]. Another important fact is that, for unicasting the computing power requirement is very low. Because here the M.S just have to verify the HMAC & save the keys [1]. Also the use of a hash chain does not require much computation. Here the M.S has to calculate the hash function of the received key and compare it with the saved key [1].

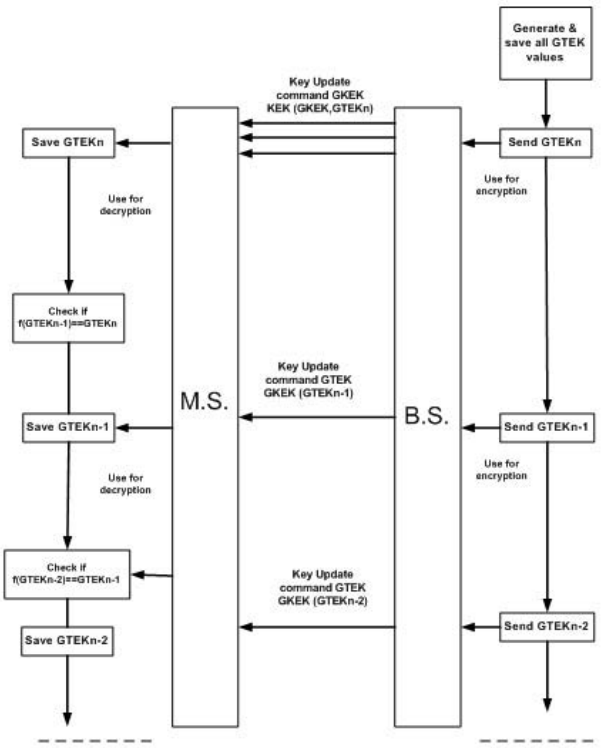


Fig.3 : Avoid key forgery by a GTEK hash chain

III. The Proposed Protocols

In this section, we propose three user authentication with key establishment protocols (UAKE) satisfying: Class-1, Class-3, and Class-7 PFS. The proposed protocols only use one-way hash functions & exclusive-or (XOR) operations. Each proposed protocol involves two phases: 1) the initialization phase 2) the user authentication with key establishment phase. Table I shows the notations used throughout our protocols.

Table 1 : The notations used in our Protocols

Notations	Description
MD	the mobile device
S	the authentication server
AS	the application server
ID_{MD}	the identity of MD
ID_S	the identity of S
ID_{AS}	the identity of AS
x	a secret key held by the
PW_{MD}	the password of MD
S_{AS}	the shared key between S and AS
$h(\cdot)$	a secure one-way hash function
$\|\|$	string concatenation operation
$\oplus$	exclusive-or operation

a) The Proposed UAKE Protocol with Class-1 PFS

In this protocol, an attacker cannot obtain the previous session keys even if PW_{MD} and S_{AS} are both disclosed. Details are given with the following steps.

i. *The initialization phase:*

In this protocol, S computes $A_{MD} = h(ID_{MD} || x)$ and stores it in MD. Moreover, S computes $A_{AS} = h(ID_{AS} || x)$ and sends it to AS via a secure channel.

ii. *User authentication with key establishment phase:*

Step 1. MD generates a random number R_{MD} to compute $M_1 = A_{MD} \oplus R_{MD}$ and $M_{1\_MAC} = h(ID_{MD} || R_{MD}) \oplus PW_{MD}$. Then MD sends (ID_{MD}, ID_{AS}, M_1, M_{1_MAC}) to AS.

Step 2. After receiving (ID_{MD}, ID_{AS}, M_1, M_{1_MAC}), AS generates a random number R_{AS} to compute $M_2 = A_{AS} \oplus R_{AS}$ and $M_{2\_MAC} = h(ID_{AS} || R_{AS}) \oplus S_{AS}$. Then AS sends (ID_{MD}, M_1, M_{1_MAC}, M_2, M_{2_MAC}) to S.

Step 3. S computes $R_{MD} = M_1 \oplus h(ID_{MS} || x)$ and $R_{AS} = M_2 \oplus h(ID_{AS} || x)$ using its secret key x. Then S checks whether M_{1_MAC} and M_{2_MAC} are the same with $h(ID_{MD} || R_{MD}) \oplus PW_{MD}$ and $h(ID_{AS} || R_{AS}) \oplus S_{AS}$, respectively. If both verifies pass, step 4 is then performed. Otherwise, S denies this request.

Step 4. Next, S generates a session key K to compute $M_{MD} = h(R_{MD}) \oplus K$, $M_{MD\_MAC} = h(R_{MD} || K)$, $M_{AS} = h(R_{AS}) \oplus K$ and $M_{AS\_MAC} = h(R_{AS} || K)$. Then, S sends ($ID_{MD}$, M_{MD}, M_{MD_MAC}, ID_{AS}, M_{AS}, M_{AS_MAC}) to AS.

Step 5. As computes $K = MD_{AS} \oplus h(R_{AS})$ and checks whether M_{AS_MAC} is the same with $h(R_{AS} || K)$. If they are the same, AS can obtain the session key K and then sends (ID_{MD}, M_{MD}, M_{MD_MAC}) to MD.

Step 6. After receiving (ID_{MD}, M_{MD}, M_{MD_MAC}), MD computes $K = M_{MD} \oplus h(R_{MD})$ and checks whether M_{MD_MAC} is the same with $h(R_{MD} || K)$. If they are the same, MD also can obtain K.

b) The Proposed UAKE Protocol with Class-7 PFS

In this protocol, an attacker cannot get the previous session keys even if PW_{MD}, S_{AS}, and x are all disclosed. The process is explained below.

i. *The initialization phase:*

Before the protocol begins, S computes $A_{MD} = h(ID_{MD} // x)$ and stores it in MD. Also, S computes $A_{AS} = h(ID_{AS} // x)$ and sends it to AS via a secure channel.

ii. *User authentication with key agreement phase:*

Step 1. MD chooses a large prime p, a primitive

element g in Galois filed GF (p) and a random number d ∈ [1, p-1]. Then, MD computes M1 = AMD⊕ g^d and M_{1_MAC} = h(ID_{MD} || g^d) ⊕ PW_{MD}, and sends

Step 2. After receiving *(ID_{MD}, ID_{AS}, p, g, M_1, M_{1_MAC})*, *AS* chooses a random number a ∈ [1, p-1] to compute M_2 = A_{AS} ⊕ g^a and M_{2_MAC} = h(ID_{AS} || g^d) ⊕ S_{AS}. Then AS sends *(ID_{MD}, p, g ,M_1, M_{1_MAC}, ID_{AS}, M_2, M_{2_MAC})* to *S*.

Step 3. S computes g^d = M_1 ⊕ h(IDMD || x) and g^a = M_2 ⊕ h(IDAS || x) using its secret key x. Then S verifies whether M1_MAC and M2_MAC are equal to h(IDMS || g^d) ⊕ PWMD and h(IDAS || g^a) ⊕ SAS respectively. If they are both equal, step 4 is subsequently carried out. Otherwise, S denies this request.

Step 4. S chooses a random number s∈[1, p-1] to compute k_{CS} = $(g^a)^s$ = g^{as} and k_{AS} = $(g^d)^s$ = g^{ds}. Then S computes M_{MD} = k_{CS} ⊕ g^d, M_{MD_MAC} = h(k_{CS} || g^d), MAS = kAS ⊕ g^a and MAS_MAC = h(kAS || g^a) sends them to AS.

Step 5. After receiving *(ID_{MD}, M_{MD}, M_{MD_MAC}, ID_{AS}, M_{AS}, M_{AS_MAC})*, *AS* computes k_{AS} = M_{AS} ⊕ g^a and verifies whether MAS_MAC equals to *h(k_{AS} || g^a)*. If it holds, AS can compute the session key K from *K* = *(K_{AS})^a* = *(g^{ds})^a* = *g^{ads}*. Then *AS* sends *(ID_{MD}, M_{MD}, M_{MD_MAC})* to *MD*.

Step 6. *MD* computes k_{CS} = M_{MD} ⊕ g^d and verifies whether M_{MD_MAC} equals to *h(k_{CS} || g^d)*. If they are equal, *MD* can compute the session key *K* from *K* = *(k_{CS})^d* = *(g^{as})^d* = *g^{ads}*.

The proposed protocols only use one-way hash functions and XOR operations. Moreover, the proposed protocols also provide three kinds of PFS to meet different requirements. Therefore, compared with Sun and Yeh's protocols, our protocols are more efficient and practical for mobile devices. Wherever Times is specified, Times Roman or Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times. Avoid using bit- mapped fonts if possible. True-Type 1 or Open Type fonts are preferred. Please embed symbol fonts, as well, for math, etc.

IV. Security Analysis and Discussions

In this section, we discuss some potential attacks which might occur on the proposed protocols.

a) Replay attack

The replay attack is an attack in which an attacker can use the previous eavesdropped messages to login the server without being detected [8]. Now, we are going to demonstrate in this subsection that, the proposed protocols can successfully withstand the replay attack.

i. *The proposed UAKE protocol with Class-1 PFS:*

After sending *(ID_{MD}, ID_{AS}, M_1, M_{1_MAC})* to S, an attacker can get M_{MD} in Step 4. However, the attacker can't have A_{MD} = *h(ID_{MD} || x)* that contains a secret key x protected by one-way hashing function. This also means that he cannot extract R_{MD} to obtain *K* or PW_{MD} by computing *K* = M_{MD} ⊕R_{MD} or PW_{MD} = *h(ID_{MD} || R_{MD})* ⊕M_{1_MAC}. Thus, this protocol can prevent the replay attack.

ii. *The proposed UAKE protocol with Class-3 PFS:*

An attacker replays *(ID_{MD}, ID_{AS}, M_1, M_{1_MAC})* to *AS* in Step 1 and receives *(ID_{MD}, M_{MD}, M_{MD_MAC})* in Step 5. Because both A_{MD} and R_{MD} are unknown, the attacker cannot extract *K* or PW_{MD}. As a result, the replay attack cannot *be* mounted in this protocol.

iii. *The proposed UAKE protocol with Class-7 PFS:*

Even if an attacker sends *(ID_{MD}, ID_{AS}, M_1, M_{1_MAC})* to *AS* in Step 1, he cannot obtain *K* or PW_{MD} from *AS*'s reply. Without A_{MD}, the attacker cannot obtain g^d by computing g^d = M_1 ⊕A_{MD}. Also, the attacker faces the discrete logarithm problem in computing d. Thus, it is quite impossible for the replay attack to occur in this protocol.

b) Password guessing attack

This attack refers to an intruder attempts to pass the authentication with certain guessed password [9, 10, 11]. The following discussions show, how the proposed protocols can prevent the password guessing attack.

i. *The proposed UAKE protocol with Class-1 PFS:*

An intruder tries to send the eavesdropped message M_1 and $M^*_{1\_MAC}$ = *h(ID_{MD} || R^*_{MD})* ⊕PW^*_{MD} to *S* in Step 1, where R^*_{MD} and PW^*_{MD} are generated by the intruder. In Step 2, *S* extracts R_{MD} = M_1 ⊕*h(ID_{MD} || R_{MD})* to check whether $M^*_{1\_MAC}$ is the same with *h(ID_{MD} || R_{MD})* ⊕PW_{MD} [9]. The result is S will find the equation is not correct and then refuse the request. Moreover, the intruder has no extra information to verify the guessed password PW^*_{MD}. Therefore, the password guessing attack does not work in this protocol.

ii. *The proposed UAKE protocol with Class-3 PFS:*

Assume that an intruder replays the eavesdropped message M_1 and $M^*_{1\_MAC}$ = *h(ID_{MD} || R^*_{MD})* ⊕PW^*_{MD} to *AS* in Step 1, where R^*_{MD} and PW^*_{MD} are generated by the intruder. If PW^*_{MD} and R^*_{MD} are not correct, *S* will detect this failure and stop the request in Step 3. Thus, the password guessing attack is prevented.

iii. *The proposed UAKE protocol with Class-7 PFS:*

An intruder attempts to send the eavesdropped message M_1, $M^*_{1\_MAC}$ = *h(ID_{MD} || g^*)* ⊕PW^*_{MD} to *AS* in

Step 1, where g^* and PW^*_{MD} are generated by the intruder. However, in Step 3, S will detect the failed login by verifying M_{1_MAC} because g^* and PW^*_{MD} are not correct. Therefore, the intruder has no chance to perform the password guessing attack .

c) Perfect forward secrecy

We show, as follows that the proposed protocols can satisfy Class-1, Class-3 and Class-7 PFS [12].

i. *The proposed UAKE protocol with Class-1 PFS:*

When MD's password PW_{MD} is disclosed, an attacker only can derive $h(ID_{MD} \,||\, R_{MD}) = M_{1\_MAC} \oplus PW_{MD}$. However, the attacker cannot further get the session key K by computing $K = h(R_{MD}) \oplus M_{MD}$ without A_{MD} [12]. Thus, this protocol can provide Class-1 PFS.

ii. *The proposed UAKE protocol with Class-3 PFS:*

When PW_{MD} and S_{AS} are disclosed, an attacker can obtain $h(ID_{MD} \,||\, R_{MD}) = M_{1\_MAC} \oplus PW_{MD}$ and $h(ID_{AS} \,||\, R_{AS}) = M_{2\_MAC} \oplus S_{AS}$. However, the attacker still cannot know A_{MD} and A_{AS}, which are stored in MD and AS respectively [16]. Consequently, the attacker cannot extract R_{MD} and R_{AS} from $M_1 = A_{MD} \oplus R_{MD}$ and $M_2 = A_{AS} \oplus R_{AS}$. That is, the attacker cannot get the session key K by computing $K = M_{MD} \oplus h(R_{MD})$ or $K = M_{AS} \oplus h(R_{AS})$. This protocol can provide Class-3 PFS [16].

iii. *The proposed UAKE protocol with Class-7 PFS:*

When PW_{MD}, S_{AS} and x are all disclosed, an attacker can obtain g^d and g^a by $g^d = M_1 \oplus h(ID_{MD} \,||\, x)$ and $g^a = M_2 \oplus h(ID_{AS} \,||\, x)$. Moreover, the attacker can derive $k_{CS} = M_{MD} \oplus g^d$ and $k_{AS} = M_{AS} \oplus g^a$. To get the session key $K = g^{ads}$, the attacker has to solve Diffie-Hellman problem [16]. Nevertheless, this is hard to be accomplished. Therefore, this protocol can provide Class-7 PFS.

V. Conclusion

Secured data transmission is one of the prime aspects of wireless networks as they are much more vulnerable to security attacks. In this paper, we explore the possibility of key forgery in Multi- and Broadcast service. We proposed three UAKE protocols with PFS based upon one-way hash functions and XOR operations. The computation loads and power supply requirements are less, which make this protocol more efficient and suitable than other.

References Références Referencias

1. "Shared key Vulnerability in IEEE 802.16e: Analysis & Solution"- A.K.M. NAZMUS SAKIB, Mir Md Saki Kawsor, International Conference on Computer & Information Technology 2010 [IEEE].
2. E. Bresson, O. Chevassut, A. Essiari, and D. Pointcheval,"Mutual authentication and group key agreement for low-power mobile devices," Computer Communications, Vol. 27, pp. 1730-1737, 2004.
3. H. Y. Chien and J. K. Jan, "Robust and simple authentication protocol," Computer Journal, Vol. 46, pp. 193-201, 2003.
4. M. Sandirigama, A. Shimizu, and M. T. Noda,"Simple and secure password authentication protocol (SAS)," IEICE Transactions on Communications, Vol. E83-B, pp. 1363-1365, 2000.
5. M. S. Hwang, C. C. Lee, and Y. L. Tang, "A simple remote user authentication protocol," Mathematical and Computer Modelling, Vol. 36, pp. 103-107, 2002.
6. Ju-Yi Kuo, "Analysis of 802.16e Multicast/Broadcast group privacy rekeying protocol", Stanford University, CA, USA, 2006.
7. Tao Han, Ning Zhang, Kaiming Liu, Bihua Tang,"Analysis of Mobile WiMAX security: vulnerabilities and Solutions", Yuan'an Liu Key Lab Of Universal Wireless Communications, Ministry of Education (Beijing University of Posts and Telecommunications)
8. T. Kwon, M. Kang, Jung, and J. Song, "An improvement of the password-based authentication protocol (K1P) on security against replay attacks", IEICE Transactions on Communications, Vol. E82–B, pp. 991-997, 1999.
9. L. Gong, "Optimal authentication protocols resistant to password guessing attacks," Proceedings of The Eighth IEEE Computer Security Foundations Workshop, Country Kerry, Ireland, pp. 24-29, 1995.
10. L. Gong, M. Lomas, R. Needham, and J. Saltzer, "Protecting poorly chosen secrets from guessing attacks," IEEE Journal on Selected Areas in Communications, Vol. 11, pp. 648-656, 1993.
11. H. M. Sun and H. T. Yeh, "Password-based authentication and key distribution protocols with perfect forward secrecy," Journal of Computer and System Sciences, Vol. 72, pp. 1002-1011, 2006.
12. T. Kwon and J. Song, "Authenticated key exchange protocols resistant to password guessing attacks," IEE Proceedings Communications, Vol. 145, pp. 304-308, 1998.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 11 Issue 16 Version 1.0 September 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

A Study on Enhancement of the Security of the Routing Protocols in Adhoc Networks

By C. Chandrasekar, Lt.Dr. S. Santhosh Baboo
Sree Narayana Guru College, Coimbatore , India

Abstract - An ad hoc wireless network is a set of wireless mobile nodes that self-configure to build a network without the requirement for any reputable infrastructure or backbone. Mobile nodes are utilized by the Ad hoc networks to facilitate effective communication beyond the wireless transmission range. As ad hoc networks do not impose any fixed infrastructure, it becomes very tough to handle network services with the available routing approaches, and this creates a number of problems in ensuring the security of the communication. Majority of the existing ad hoc protocols that deal with security issues depends on implicit trust relationships to route packets among participating nodes. The general security objectives like authentication, confidentiality, integrity, availability and nonrepudiation should not be compromised in any circumstances. Thus, security in ad hoc networks becomes an active area of research in the field of networking. There are various techniques available in the literature for providing security to the ad hoc networks. This paper focuses on analyzing the various routing protocols available in the literature for ad hoc network environment and its applications in security mechanisms.

GJCST Classification : C.2, C.2.2

Strictly as per the compliance and regulations of:

A Study on Enhancement of the Security of the Routing Protocols in Adhoc Networks

C. Chandrasekar[α], Lt.Dr. S.Santhosh Baboo[Ω]

Abstract - An ad hoc wireless network is a set of wireless mobile nodes that self-configure to build a network without the requirement for any reputable infrastructure or backbone. Mobile nodes are utilized by the Ad hoc networks to facilitate effective communication beyond the wireless transmission range. As ad hoc networks do not impose any fixed infrastructure, it becomes very tough to handle network services with the available routing approaches, and this creates a number of problems in ensuring the security of the communication. Majority of the existing ad hoc protocols that deal with security issues depends on implicit trust relationships to route packets among participating nodes. The general security objectives like authentication, confidentiality, integrity, availability and nonrepudiation should not be compromised in any circumstances. Thus, security in ad hoc networks becomes an active area of research in the field of networking. There are various techniques available in the literature for providing security to the ad hoc networks. This paper focuses on analyzing the various routing protocols available in the literature for ad hoc network environment and its applications in security mechanisms.

I. Introduction

An ad hoc network [1] is an infrastructureless network in which the nodes themselves are accountable for routing the packets. In the conventional Internet, routers within the central parts of the network are owned by a few well known operators and are therefore assumed to be somewhat trustworthy. This statement cannot hold good in an ad hoc network as all nodes coming into the network are expected to involve in routing. As the links in general are wireless, the security that was obtained because of the difficulty of tapping into a network is lost. Moreover, as the topology in such a network can be extremely dynamic, conventional routing protocols can no be effective.

The routing protocol [2, 3] provides an upper limit to security in any packet network. If routing can be misdirected, the whole network will be affected greatly. The issue is inflated by the fact that routing generally depends on the trustworthiness of all the nodes that are participating in the routing process. It is very tough to differentiate compromised nodes from nodes that are suffering from bad links.

Author α : M.C.A., M.Phil., Assistant Professor, Sree Narayana Guru College, Coimbatore - 641 105, India.
E-mail : Chandrasekar2000@gmail.com
Author Ω : Reader, D.G.Vaishnav College, Chennai - 600 106, India.

Because of self organize and rapidly deploy capability, ad hoc can be used in various applications like battlefield communications, emergency relief scenarios, law enforcement, public meeting, virtual class room and other applications. Though security [4] [5] has long been a vital and active area of research in wired networks, the unique features of Mobile Ad hoc Networks (MANETs) offer a new collection of nontrivial difficulties to security design. These difficulties comprise open network architecture, shared wireless medium, stringent resource constraints, and highly dynamic network topology. It is very tough to maintain security of MANETs in group communication as of multiple senders and multiple receivers.

Previous security research [6] [7] in routing protocol mainly focuses on the use of encryption technology to implement message authentication. These routing protocols rely entirely on a central authority. Moreover, the performance of on demand routing protocols is very less which leads to various attacks. Thus, none of these existing protocols specifies any effective security measures which leads to malicious routing operations.

The main objective of this paper is to discuss ad hoc routing security with respect to the area of security. Various routing protocols available in the literature are analyzed to provide better security to the routing in ad hoc networks.

II. Literature Survey

There is no centralized administration or fixed network infrastructure for the ad hoc network and thus nodes execute routing discovery and routing maintenance in a self-organized way. But, this flexible network topology suffer from various security problems and the existing routing protocols such as AODV has no effective measures to avoid themselves from being attacked. There are various secure routing protocol techniques available in the literature to defend the ad hoc networks. However, majority of these secure routing protocols require certain centralized units or some trusted third parties to provide digital certificates or monitor network traffics, which demolish the self-organization nature of ad hoc networks. In this paper, Zhiyuan et al., [8] propose a secure routing protocol based on the trust mechanism. Each node in this ad hoc network has its views about some other node's

reliability, which are acquired by directly communicating with other nodes or by integrating other node's recommendations. Then the node will determine whether to exchange routing data with another node based on its view about that node's reliability.

The growth and development of tele-communication has increased the need for mobility, wireless or mobile networks and this has given more attention to the wired networks. The upcoming networks has entirely different infrastructure and has various protocols and devices. The main aim of this approach is to assess the two secure routing protocols Ariadne and SAODV in the performance characteristics rather than security features under random way point and Manhattan grid mobility models. Naeem et al., [9] used and implement the extension of AODV that is Secure Ad-hoc On-demand Distance Vector routing protocol (SAODV) and the extension of DSR that is Ariadne in the network simulator 2 (NS-2). In this paper, these protocols are compared with the quality of service parameters like delay, jitter, routing overhead, route acquisition time, throughput, hop count, packet delivery ratio using Manhattan grid and random waypoint mobility models. This paper mainly focuses on finding out the payload a node has to pay to assure the good quality of service.

Communications in MANETs are becoming more malicious in traffic analysis because of the broadcast nature of wireless transmissions. Even though, there are various secure routing protocols, traffic analysis attacks are still not well addressed with those existing techniques. Certainly, these protocols concentrate on security of route maintaining and protecting against modification of routing data, which cannot prevent traffic analysis attack. Anonymity is one of the most vital techniques to resistant against the malicious traffic analysis. In this paper, Sheklabadi et al., [10] described an anonymous version of ARAN, which is one of the significant secure routing protocols, to offer anonymity and maintain security of nodes in MANETs. The proposed protocol is based on the integration of the anonymous communication along with security specifications of ARAN. The main contribution of this protocol is combining several anonymous functionalities such as identity privacy, location privacy and route anonymity together with security features of ARAN

In order to secure the MANET in adversarial environments, it is necessary to possibly detect and defend possible attacks on routing protocols, especially internal attacks, such as a Byzantine attack. Ming Yu et al., [11] proposed a novel technique that identifies internal attacks by using both message and route redundancy during route discovery. The route-discovery messages are secured by pairwise secret keys between a source and destination and some intermediate nodes along a route established by using public key cryptographic mechanisms. An optimal routing technique is also proposed with routing metric integrating both requirements on a node's reliability and performance. A node constructs the reliability on its neighboring node's depending on its observations on the behaviors of the neighbor nodes. These two techniques can be combine into existing routing protocols like Ad hoc On-demand Distance vector routing (AODV) and Dynamic Source Routing (DSR). The author presented an integrated protocol called Secure Routing against Collusion (SRAC), in which a node makes a routing decision depending on its trust of its neighboring nodes and the performance provided by them. The simulation results have shown the advantages of the proposed attack detection and routing algorithm over the existing technique.

MANETs has several kinds of security issues, caused by their nature of collaborative and open systems and by limited availability of resources. In this paper, Cerri et al., [12] consider a Wi-Fi connectivity data link layer as a fundamental technique and concentrates on routing security. The author discusses the implementation of the secure AODV protocol extension, which comprises of alteration policies aimed at enhancing its performance. The author proposed an adaptive technique that adjusts SAODV behavior. Furthermore, the author examined the adaptive technique and another approach that delays the verification of digital signatures. This paper sums up the experimental results collected in the prototype design, implementation, and tuning.

MANETs are a set of wireless mobile devices with limited broadcast range and resources, and no fixed infrastructure. Communication is attained by communicating data along suitable routes that are vigorously identified and maintained through collaboration between the nodes. Determining such routes is a major job, both from efficiency and security points of view. Recently, a security model tailored to the particular needs of MANETs was introduced by Acs, Buttyan, and Vajda. The novel feature of this security system is that it assures security under concurrent executions. A novel route discovery technique called endairA was also proposed, along with a claimed security proof within the same system. In this paper, Burmester et al., [13] described that the security proof for the route discovery algorithm endairA is faulty, and moreover, this approach is susceptible to a hidden channel attack. The author also examined the security framework that was used for route discovery and argued that composability is a vital feature for ever-present applications. Ultimately, some of the major security challenges for route discovery in MANETs are discussed.

Decentralized node admission is a vital and fundamental security service in MANETs. It is required to steadily cope with dynamic membership and topology in addition to bootstrap other considerable security primitives (such as key management) and services (such as secure routing) without the help of any centralized trusted authority. A perfect admission approach should have least interaction among MANET nodes, as connectivity can be unstable. Moreover, as MANETs are frequently consists of weak or resource-limited devices, admission should be capable in terms of computation and communication. Majority of the existing admission protocols are prohibitively costly and need heavy interaction among MANET nodes. In this paper, Saxena et al., [14] concentrates on a general type of MANET that is formed on a temporary basis, and present a secure, efficient, and a fully noninteractive admission technique geared for this type of a network. This admission protocol depends on secret sharing techniques using bivariate polynomials. The author also presents a novel approach that facilitates any pair of MANET nodes to proficiently create an on-the-fly secure communication channel.

Routing in ad hoc networks is different from infrastructure-based wireless networks. In ad hoc networks each node acts as a router and is accountable for organizing topological data and ensuring correct route learning. In spite of various secure routing algorithms, security in ad hoc networks is still a controversial area. In this paper, Afzal et al., [15] first investigate the security issues and attacks in existing routing protocols and then the design and analysis of a new secure on-demand routing protocol, called RSRP is presented which appropriates the problems declared in the existing protocols. Furthermore, unlike Ariadne, RSRP uses a very proficient broadcast authentication technique which does not need any clock synchronization and assists instant authentication.

Routing in ad hoc network is one of the fundamental issues in networking. An opponent can easily hack the information in the network by attacking the routing protocol. There are several techniques available for the security enhancement of ad hoc network. In this paper, Imani et al., [16] argued about the defects in an ad hoc routing protocol that called Ariadne. This paper demonstrates that the security evidence for the route discovery technique Ariadne is defective, and furthermore, this algorithm is susceptible to certain attacks. In order to solve the limitations of this protocol, a novel proposed approach is presented in the route discovery algorithm. The proposed approach in this paper adds the capability of the malevolent node detections to this protocol.

Multipath routing diminishes the penalty of security attacks obtaining from collaborating malevolent nodes in MANET, by increasing the number of nodes that an opponent must negotiate in order to take control of the communication. In this paper, various attacks that cause multipath routing protocols more susceptible to attacks than it is expected, to collaborating malevolent nodes are recognized. Kotzanikolaou et al., [17] proposed a novel On-demand Multipath routing protocol called the Secure Multipath Routing protocol (SecMR) and the author examine its security properties. The SecMR protocol can be easily combined in an extensive variety of on-demand routing protocols, such as DSR and AODV.

Hu et al., [18] propose a more forceful protocol, which is more powerful in terms of security associations. In this approach, it is assumed that security associations are present between all pairs of nodes (through authentic public or Tesla [19] keys, or by shared secret keys). This facilitates both the sender and the receiver to validate all the nodes on the selected routing path.

Papadimitratos et al., [20] assumed that, for effective secure routing, it is enough, if effective security association is established between the sender and the receiver. It is demonstrated that the author's proposal avoids a wide range of attacks, but the proposed protocol is still susceptible to certain active attacks [21]. The author proposed a protocol (SRP) that can be effectively applied to a wide variety of existing routing protocols. This protocol focuses on the security association between source and destination nodes. Intermediate nodes need not require cryptographic validation of the control traffic. It adds an SRP header to the base routing protocol (DSR or AODV) request packet. SRP header has three vital fields namely QSEQ, QID and SRP MAC. QSEQ facilitates to avoid replay of old outdated requests. QID and random number help to prevent fabrication of requests, and SRP MAC guarantees reliability of the packets in communication. In SRP, for every route discovery, it is necessary that the source and destination must have a security association between them. Moreover, this approach does not focus on the route error messages. Hence, they are not protected, and any malevolent node can just counterfeit error messages with other nodes as source.

ARIADNE [22] is based on DSR [23] and TESLA (on which its authentication approach is based). ARIADNE prevents attackers/compromised nodes from troublemaking uncompromised routes that consist of benign nodes. It employs highly effective symmetric key cryptography technique. ARIADNE does not offer effective security against passive attackers eavesdropping on the network traffic. It does not provide security from an attacker from inserting data packets. It is susceptible to active-1-1 attacker that lies along the identified route, which does not forward packets and does not cause error if it meets a broken link. It also needs clock synchronization, which is regarded as an unrealistic necessity for ad hoc networks.

Perlman proposed a link state routing protocol [24] that attains Byzantine strength. Though, the protocol is extremely forceful, it needs a very high operating cost associated with public key encryption. Zhou and Haas [25] chiefly describe key management in their paper to provide security to ad hoc networks. The author devotes a part to secure routing, but in essence concludes that "nodes can defend routing data in the similar way they protect data traffic". They also examine that denial-of-service attacks against routing will be considered as damage and it is routed around. Certain research has been done to secure ad hoc networks by means of misbehavior detection approaches. This technique has two major problems: Initially, it is fairly likely that it will be not possible to discover various kinds of misbehaving; and secondly, it has no real means to assure the integrity and authentication of the routing messages.

Dahill *et al.* [26] proposed ARAN. Managed open environment is considered in this approach, where there is an opportunity for pre-deployment of infrastructure. It consists of two distinctive stages. The first stage is the certification and end-to-end authentication stage. Here the source obtains a certificate from the trusted certification server, and then by means of this certificate, signs the request packet. Each intermediate node consecutively signs the request with its certificate. The destination then validates each of the certificates, hence the source and the intermediate nodes gets authenticated. The destination node then sends the reply through the route reverse to the one in the request; reply signed with the help of the certificate of the destination. The second stage is a non-compulsory stage which is used to identify the shortest path to the destination, but this stage is very costly. It is susceptible to reply attacks using error messages but for the nodes have time synchronization.

III. Problems and Directions

The lack of infrastructure and organizational setting of mobile ad-hoc networks creates unique chances to attackers. MANETs are generally organized without a central control unit; the devices in a MANET depend on other units to route data to their destinations. Moreover, MANET nodes are frequently constrained in power and this makes MANETs susceptible to several malevolent attacks and usage of the routing approaches that work with wired networks is infeasible.

It is the fact that secure ad hoc routing can be achieved at the expense of messages, time and computation power, and that the overhead stems mainly from the computation complexity of the cryptographic techniques employed in frequently repeated routing procedures.

The major factors that should be considered in the establishment of sufficient routing protocols are multi hop, mobility, large network size combined with device heterogeneity, bandwidth and battery power. In order to solve the challenging problem of routing in ad hoc wireless networks, a novel technique is needed. The field of artificial intelligence can provide significant solution to the security problems in routing. Specifically, techniques from Swarm Intelligence (SI) and many Optimization techniques can be taken into account.

IV. Conclusion

To establish a secure MANET routing protocol with multiple metrics is a challenging task, particularly as the network topology and traffic are dynamic and changing all the time. This chapter focuses on the routing protocols in the ad hoc networks. In this paper, the routing algorithms that support communications in mobile ad hoc networks are discussed. The majority of the existing routing protocols suffer from various drawbacks and efficient security is not given to the MANET. This survey on the secure routing protocols is very much useful for the enhancement of the other routing protocol techniques. These routing protocols are the source for the development of new routing protocols with better security and performance.

References References Referencias

1. Ismail M., "Routing Protocols for Ad Hoc Wireless Networks", M. Sc. (ISS) project, Carleton University, Ontario, Canada, 2001.
2. M. Guerrero Zapata and N. Asokan, "Securing Ad hoc Routing Protocols," in Proceedings of the 1st ACM workshop on Wireless security, Atlanta, GA, USA, Sep 2002, pp. 1–10.
3. L. Abusalah, A. Khokhar, and M. Guizani, "A Survey of Secure Mobile Ad Hoc Routing Protocols," IEEE Communications Surveys & Tutorials, vol. 10, no. 4, pp. 78-93, 2009.
4. Keng Seng Ng and Seah W. K. G., "Routing security and data confidentiality for mobile ad hoc networks", The 57th IEEE Semiannual Vehicular Technology Conference, 2003.
5. K. Sanzgiri, B. Dahill, B. Levine, and E. Belding-Royer, "A secure routing protocol for ad hoc networks," in Proceedings of IEEE ICNP, 2002.
6. H. Deng, W. Li, and D. Agrawal, "Routing Security in Wireless Ad Hoc Networks," IEEE Comm. Magazine, vol. 40, no. 10, 2002, pp. 70 75.
7. L. Venkatraman and D.P. Agrawal, "Strategies for enhancing routing security in protocols for mobile ad hoc networks," J. Parallel Distrib. Comp., 2002.
8. Zhiyuan Liu; Shejie Lu; Jun Yan; "Secure Routing Protocol based Trust for Ad Hoc Networks", Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, Vol. 1, Page(s): 279 – 283, 2007.

9. Naeem, M.; Ahmed, Z.; Mahmood, R.; Azad, M.A.; "QOS based performance evaluation of secure on-Demand routing protocols for MANET's", International Conference on Wireless Communication and Sensor Computing, 2010, pages 1-6, ICWCSC 2010.
10. Sheklabadi, E. Berenjkoub, M. "An anonymous secure routing protocol for mobile ad hoc networks", 2011 International Symposium on Computer Networks and Distributed Systems (CNDS), page(s): 142 – 147, 2011.
11. Ming Yu Mengchu Zhou Wei Su "A Secure Routing Protocol against Byzantine Attacks for MANETs in Adversarial Environments", IEEE Transactions on Vehicular Technology, Vol. 58, No. 1, pages 449 – 460, 2009.
12. Cerri, D. Ghioni, A. "Securing AODV: the A-SAODV secure routing prototype", IEEE Communications Magazine, Vol. 46, No. 2, page(s): 120 – 125, 2008.
13. Burmester, M.; de Medeiros, B.; "On the Security of Route Discovery in MANETs", IEEE Transactions on Mobile Computing, Vol. 8, No. 9, Page(s): 1180 – 1188, 2009.
14. Saxena, N.; Tsudik, G.; Jeong Hyun Yi; "Efficient Node Admission and Certificateless Secure Communication in Short-Lived MANETs", IEEE Transactions on Parallel and Distributed Systems, Vol. 20, No. 2, Page(s): 158 – 170, 2009.
15. Afzal, S.R.; Biswas, S.; Jong-bin Koh; Raza, T.; Gunhee Lee; Dong-kyoo Kim; "RSRP: A Robust Secure Routing Protocol for Mobile Ad Hoc Networks", IEEE Wireless Communications and Networking Conference, page(s): 2313 – 2318, 2008. WCNC 2008.
16. Imani, M.; Taheri, M.; Rajabi, M.E.; Naderi, M.; "A secure method on a routing protocol for ad hoc networks", 2010 International Conference on Educational and Network Technology (ICENT), page(s): 482 – 486, 2010.
17. Kotzanikolaou, P.; Mavropodi, R.; Douligeris, C.; "Secure Multipath Routing for Mobile Ad Hoc Networks", WONS 2005. Second Annual Conference on Wireless On-demand Network Systems and Services, Page(s): 89 – 96, 2005.
18. Y.-C. Hu, A. Perrig, and D. B. Johnson. Ariadne: A Secure On-Demand Routing Protocol for Ad Hoc Networks. In Proceedings of MobiCom, September 2002.
19. A. Perrig, R. Canetti, J.D. Tygar, and D. Song. The TESLA Broadcast Authentication Protocol. RSA CryptoBytes, 5(Summer), 2002.
20. P. Papadimitratos and Z.J. Haas. Secure Routing for Mobile Ad Hoc Networks. In Proceedings of CNDS, January 2002.
21. Yih-Chun Hu, Adrian Perrig, and David B. Johnson. Packet leashes: A defense against wormhole attacks in wireless networks. In Proceedings of IEEE Infocom, April 2003.
22. Y. C. Hu, A. Perrig, and D. Johnson, "ARIADNE: A secure on-demand routing protocol for ad-hoc networks", Technical Report TR01-383, Rice University (2001).
23. D. B. Johnson et al., "The dynamic source routing protocol for mobile ad-hoc networks (DSR)", Internet draft, MANET Working Group (2002).
24. R. Perlman, Fault-tolerant broadcast of routing information, Computer Networks, 7, 395–405 (1983).
25. L. Zhou, and Z. J. Haas, Securing ad-hoc networks, IEEE Network Mag., 13, 24–30 (1999).
26. B. Dahill, B. N. Levine, E. Royer, and C. Shields, A secure routing protocol for ad-hoc networks, Technical Report UM-CS-2001-037, Department of Computer Science, University of Massachusetts (2001).

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 11 Issue 16 Version 1.0 September 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

An Effective XML Keyword Search with User Search Intention over XML Documents

By Pradeep Kumar Reddy Gade, N Prasanna Balaji, U Sreenivasulu
Ibrahimpatnam, Andhra Pradesh, India

Abstract - The extreme success of web search engines makes keyword search the most popular search model for ordinary users. Keyword search on XML is a user friendly way to query XML databases since it allows users to pose queries without the knowledge of complex query languages and the database schema. The three main challenges faces in XML keyword search: 1) Identify the user search intention, i.e., identify the XML node types that users want to search for and search via. 2) Resolve keyword ambiguity problems: a keyword can appear as both a tag name and a text value of some node; a keyword can appear as the text values of different XML node types and carry different meanings; a keyword can appear as the tag name of different XML node types with different meanings. 3) As the search results are sub trees of the XML documents, new scoring function is needed to estimate its relevance to a given query. However, existing methods cannot resolve these challenges, thus return low result quality in term of query relevance. In this paper, we propose an IR-style approach which basically utilizes the statistics of underlying XML data to address these challenges. We first propose specific guidelines that a search engine should meet in both search intention identification and relevance oriented rankingfor search results over XML documents. Then, based on theseguidelines, we design novel formulae to identify the search fornodes and search via nodes of a query, and present a novelXML TF*IDF ranking strategy to rank the individual matches ofall possible search intentions over XML documents.

Keywords : *XML, keyword search, ranking.*

GJCST Classification : *H.3.3, F.2.2*

Strictly as per the compliance and regulations of:

An Effective XML Keyword Search with User Search Intention over XML Documents

Pradeep Kumar Reddy Gade[α], N Prasanna Balaji[Ω], U Sreenivasulu[β]

Abstract - The extreme success of web search engines makes keyword search the most popular search model for ordinary users. Keyword search on XML is a user friendly way to query XML databases since it allows users to pose queries without the knowledge of complex query languages and the database schema. The three main challenges faces in XML keyword search: 1) Identify the user search intention, i.e., identify the XML node types that users want to search for and search via. 2) Resolve keyword ambiguity problems: a keyword can appear as both a tag name and a text value of some node; a keyword can appear as the text values of different XML node types and carry different meanings; a keyword can appear as the tag name of different XML node types with different meanings. 3) As the search results are sub trees of the XML documents, new scoring function is needed to estimate its relevance to a given query. However, existing methods cannot resolve these challenges, thus return low result quality in term of query relevance. In this paper, we propose an IR-style approach which basically utilizes the statistics of underlying XML data to address these challenges. We first propose specific guidelines that a search engine should meet in both search intention identification and relevance oriented ranking for search results over XML documents. Then, based on these guidelines, we design novel formulae to identify the search for nodes and search via nodes of a query, and present a novel XML TF*IDF ranking strategy to rank the individual matches of all possible search intentions over XML documents.

Keywords : XML, keyword search, ranking.

I. Introduction

The extreme success of web search engines makes keyword search the most popular search model for ordinary users. In the real world, computer systems and databases contain data in incompatible formats. XML data is stored in plain text format. This provides a software- and hardware-independent way of storing data. AsXML is becoming a standard in data representation,it is desirable to support keyword search in XML database. It is a user friendly way to query XML databases since it allows users to pose queries without the knowledge of complex query languages and the database schema.

Author [α] : 2nd year M-tech, GuruNanak Engineering College, Ibrahimpatnam, Andhra Pradesh, India. Email : reddys_gp@yahoo.com
Author [Ω]: Professor & HOD(IT), GuruNanak Engineering College, Ibrahimpatnam, Andhra Pradesh, India.
Email : gneccsebalaji@gmail.com
Author [β] : Asst. Professor, GuruNanak Engineering College, Ibrahimpatnam, Andhra Pradesh, India.
Email : ulsa535@gmail.com

Effectiveness in terms of result relevance is the most crucial part in keyword search, which can be summarized as the following three issues in XML field:

Issue 1&2 : Capture user's search intention.
i) Identify the target that user intends to search for.
ii) Infer the predicate constraint that user intends to search via.

Issue 3 : Result ranking.
i) Ranking the query results according to their objective relevance to user search intention.

Issues 1&2 addresses the search intention problem, while the third one addresses the relevance-based ranking problem w.r.t. the search intention. The search intention for a keyword query is not easy to determine and can be ambiguous, because the search via condition is not unique. While performing keyword search on XML database, three Ambiguities arises. They are:

- Ambiguity 1: A keyword can appear both as an XML tag name and as a text value of some other nodes.
- Ambiguity 2: A keyword can appear as the text values of different types of XML nodes and carry different meanings.
- Ambiguity 3 : A keyword can appear as an XML tag name in different contexts and carry different meanings.

Although many research efforts have been conducted in XML keyword search [8] [10] [29] [22][23], none of them has been addressed and resolved the above three issues in the presence of ambiguities. So far some efforts have been conducted to satisfy the user search intention but none of them addressed relevance oriented result ranking in depth.

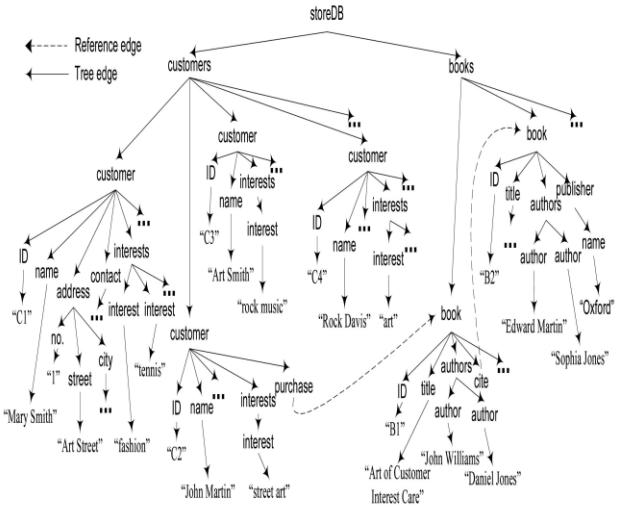


Fig. 1 : Portion of data tree for an online bookstore XML database.

Consider a keyword query "Customer name martin". The user search intention is to find the customers whose name is martin. By XML keyword search we will get two results C2 and B2 who has the keyword martin.

Even though B2 contains the name martin the XML search engine XReal give only C2 because we are searching for customer whose name is martin not the author name. So, C2 is relevant data and B2 is irrelevant data. Finally the main objective of this paper is to catch the user search intention and ranking the results in the presence of keyword ambiguities over multiple XML databases.

II. Related Work

Although many efforts have been conducted to find smallest substructures in XML data that each contains all query keywords in tree data or digraph data model. In tree data model, at first lowest common ancestor [17] (LCA) semantics is proposed to find XML nodes, each of which contains all query keywords within their subtree. Subsequently, Smallest LCA (SLCA [13], [20]) is proposed to find the smallest LCAs that do not contain other LCAs in their subtrees. GDMCT (minimum connecting trees) [7] excludes the subtrees rooted at the LCAs that do not contain query keywords. Sun et al. [18] generalize SLCA to support keyword search involving combinations of AND and OR Boolean operators. XSEEK [14] generates the return nodes which can be inferred by keyword match pattern and the concept of entities in XML data which neither addresses the ranking problem nor keyword ambiguity problem. However, it causes the multivalued attribute to be mistakenly identified as an entity, causing the inferred return node not as intuitive as possible. For example, phone and interest are not intuitive as entities. In fact, the identification of entity is highly dependent on the semantics of underlying database rather than its DTD, so it usually requires the verification and decision from database administrator. In digraph data model, previous approaches are heuristics based, as the reduced tree problem on graph is as hard as NP-complete. BANKS [6] uses bidirectional expansion heuristic algorithms to search as small portion of graph as possible. BLINKS [9] propose a bilevel index to prune and accelerate searching for top-k results in digraphs. Cohen et al. [3] study the computation complexity of interconnection semantics. XKeyword [8] provides keyword proximity search that conforms to an XML schema; however, it needs to compute candidate networks and, thus, is constrained by schemas. On the issue of result ranking, XRank[4] also extends the notion of PageRank to XML data, but no empirical study is done to show the effectiveness of its ranking function. XSearch adopts a variant of LCA, and combines a simple tf*idf IR ranking with size of the tree and the node relationship to rank results; but it requires users to know the XML schema information, causing limited query flexibility. EASE [12] combines IR ranking and structural compactness based DB ranking to fulfill keyword search on heterogeneous data. Regarding to ranking methods, TF*IDF similarity [16] which is originally designed for flat document retrieval is insufficient for XML keyword search due to XML's hierarchical structure and the presence of Ambiguity 1-3.Several proposals for XML information retrieval suggest to extend the existing XML query languages [4], [1], [19] or use XML fragments [2] to explicitly specify the search intention for result retrieval and ranking.

III. Preliminaries

a) *Your TF*IDF Cosine Similarity*

TF*IDF(Term Frequency * Inverse Document Frequency) similarity is one of the most widely used approaches to measure the relevance of keywords and document in keyword search over flat documents. We first review its basic idea, then address its limitations for keyword search in XML. The main idea of TF*IDF is summarized in the following three rules:

Rule 1 : A keyword appearing in many documents should not be regarded as being more important than a keyword appearing in a few.

Rule 2 : A document with more occurrences of a query keyword should not be regarded as being less important for that keyword than a document that has less.

Rule 3 : A normalization factor is needed to balance between long and short documents, as Rule 2 discriminates against short documents which may have less chance to contain more occurrences of keywords.

b) Data Model

The data model for XML is very simple - or very abstract, depending on one's point of view. XML provides no more than a baseline on which more complex models can be built.

We model XML document as a rooted, labeled tree plus a set of directed IDRef edges between XML nodes, such as the one in Fig. 1. In contrast to general directed graph model, the containment edge and IDRef edge are distinguished in our model.

Definition 3.1 (Node Type) : The type of a node n in an XML document is the prefix path from root to n. Two nodes are of the same node type if they share the same prefix path.

Definition 3.2(Data Node) : The text values that are contained in the leaf node of XML data and have no tag name are defined as data node.

Definition 3.3(Structural Node) : An XML node labeled with a tag name is called a structural node. A structural node that contains other structural nodes as its children is called an internal node; otherwise, it is called a leaf node.

Definition 3.4 (Single-Valued Type): A structural node t is of single-valued type if each node of type t has at most one occurrence within its parent node.

Definition 3.5 (Multivalued Type) : A structural node t is of multivalued type if some node of type t has more than one occurrence within its parent node.

Definition 3.6 (Grouping Type) : An internal node t is defined as a grouping type if each node of type t contains child nodes of only one multivalued type.

Single-valued type and multivalued type of XML nodes can be easily identified when parsing the data. Every multivalued node has a grouping node as its parent and a grouping node is also a single-valued node. Thus, the children of an internal node are either of same multivalued type or of different single-valued types. An internal node n contains both data nodes and structural nodes.

c) Capturing Keyword Co-Occurrence

In this section, we discuss the search via confidence for a data node. Although statistics provide a macro way to compute the confidence of a structural node type to search via, it alone is not adequate to infer the likelihood of an individual data node to search via for a given keyword in the query. Example 6. Consider a query "customer name Rock interest Art" searching for customers whose name includes "Rock" and interest includes "Art." Based on statistics, we can infer that name typed and interest-typed nodes have high confidence to search via by (7), as the frequency of keywords "name" and "interest" are high in node types name and interest, respectively. However, statistics is not adequate to help the system infer that the user wants "Rock" to be a value of name and "Art" to be a value of interest, which is intuitive with the help of keyword co-occurrence captured. Thus, if purely based on statistics, it is difficult for a search engine to differ customer C4 (with name "Art" and interest "Rock") from C3 (with name "Rock" and interest "Art") in Fig. 1.

IV. Inferring Keyword Search Intention

In this section, we discuss how to interpret the search intentions of keyword query according to the statistics in XML data and the pattern of keyword co-occurrence in a query.

a) Inferring the Node Type to Search for

The desired node type to search for is the first issue that a search engine needs to address in order to retrieve the relevant answers, as the search target in a keyword query may not be specified explicitly like in structured query language. Given a keyword query q, a node type T is considered as the desired node to search for only if the following three guidelines hold:

Guideline 1 : T is intuitively related to every query keyword in q, i.e., for each keyword k, there should be some (if not many) T-typed nodes containing k in their subtrees.

Guideline 2 : XML nodes of type T should be informative enough to contain enough relevant information.

Guideline 3 : XML nodes of type T should not be overwhelming to contain too much irrelevant information.

b) Inferring the Node Types to Search via

Similar to inferring the desired search for node, Intuition 1 is also useful to infer the node types to search via. However, unlike the search for case which requires a node type to be related to all keywords, it is enough for a node type to have high confidence as the desired search via node if it is closely related to some (not necessarily all) keywords, because a query may intend to search via more than one node type. For example, we can search for customer(s) named "Smith" and interested in "fashion" with query "name smith interest fashion." In this case, the system should be able to infer with high confidence that name and interest are the node types to search via, even if keyword "interest" is probably not related to name nodes.

V. Relevance Oriented Ranking

a) Principles of Keyword Search in XML

Compared with flat documents, keyword search in XML has its own features. In order for an IR-style ranking approach to smoothly apply to it, we present three principles that the search engine should adopt.

Principle 1 : When searching for XML nodes of desired type D via a single-valued node type V , ideally, only the values and structures nested in V -typed nodes can affect the relevance of D-typed nodes as answers,

whereas the existence of other typed nodes nested in D-typed nodes should not. In other words, the size of the subtree rooted at a D-typed node d (except the subtree rooted at the search via node) shouldn't affect d's relevance to the query.

Principle 2 : When searching for the desired node type D via a multivalued node type V 0, if there are many V 0-typed nodes nested in one node d of type D, then the existence of one query-relevant node of type V 0 is usually enough to indicate, d is more relevant to the query than another node d0 also of type D but with no nested V 0-typed nodes containing the keyword(s). In other words, the relevance of a D-typed node which contains a query-relevant V 0-typed node should not be affected (or normalized) too much by other query irrelevant V 0-typed nodes.

Principle 3: The proximity of keywords in a query is usually important to indicate the search intention.

*b) Advantages of XML TF*IDF*

Compatibility : The XML TF*IDF similarity can work on both semi-structured and unstructured data, because unstructured data is a simpler kind of semi-structured data with no structure, and XML TF*IDF ranking (9a) for data node can be easily simplified to the original TF*IDF (1) by ignoring the node type.

Robustness : Unlike existing methods which require a query result to cover all keywords [14], [20], [7], we adopt a heuristic-based approach that does not enforce the occurrence of all keywords in a query result; instead, we rank the results according to their relevance to the query. In this way, more relevant results can be found, because a user query may often be an imperfect description of his real information need [5]. Users never expect an empty result to be returned even though no result can cover all keywords; fortunately, our approach is still able to return the most relevant results to users.

c) XML keyword search over xml documents

The main objective of XReal search engine is to capture users search intention and relevance ranking the results in the presence of keyword ambiguity problems mentioned above. In these paper, an algorithms is used for searching a keyword in folder (having recursive folders containing xml databases) containing different xml databases.

For example, an xml database maintaining particular database for each academic year, then XReal search engine is used.

The important steps followed are:

Step 1 : Searching for keywords in every database and collecting list of databases containing the keywords.

Step 2 : keyword search by applying search for and search via node for an individual database.

Step 3 : Appling XML TF*IDF similarity on the results obtained for an individual database.

Algorithm. RecurrsivePath()

1. *Let FolderSearch = True, Result[] = Null, RecursiveSearch= True*
2. *If (FolderSearch)*
3. *ScanDir(FolderPath, RecursiveSearch)*

Function ScanDir(FolderPath, RecursiveSearch)

1. *Files = GetFiles(StartingPath)*
2. *foreach f ∈ Files*
3. *If (KWSearch(Q[m], IL[m], F[m]))*
4. *Result = XMLFileListItem(filename)*
5. *If (RecursiveSearch)*
6. *Folders = GetDirectories(StartingPath)*
7. *foreach f ∈ Folders*
8. *ScanDir (f, RecurresiveSearch)*

Algorithm. KWSearch(Q[m], IL[m], F[m]) [21] is used for keyword search in individual xml keywords.

VI. Conclusion

In this paper, we study the problem of effective XML keyword search which includes the identification of user search intention and result ranking in the presence of keyword ambiguities. We utilize statistics to infer user search intention and rank the query results. In particular, we define XML TF and XML DF, based on which we design formulae to compute the confidence level of each candidate node type to be a search for/search via node, and further propose a novel XML TF*IDF similarity ranking scheme to capture the hierarchical structure of XML data. Lastly, the popularity of a query result (captured by IDRef relationships) is considered to handle the case that multiple results have comparable relevance scores. In future, we would like to extend our approach to handle the XML document conforming to a highly recursive schema as well.

Refrences Refrences Refrencias

1. S. Amer-Yahia, L.V.S. Lakshmanan, and S.Pandit, "Flexpath: Flexible Structure and Full-Text Querying for XML," Proc. ACM SIGMOD Conf., 2004.
2. D. Carmel, Y.S. Maarek, M. Mandelbrod, Y.Mass, and A. Soffer, "Search XML Documents via XML Fragments," Proc. ACM SIGIR, pp.151-158, 2003.
3. S. Cohen, Y. Kanza, B. Kimelfeld, and Y. Sagiv, "Interconnection Semantics for Keyword Search in XML," Proc. ACM Int'l Conf. Information and Knowledge Management (CIKM), pp. 389-396, 2005.
4. N. Fuhr and K. Großjohann, "XIRQL: A Query Language for Information Retrieval in XML Documents," Proc. ACM SIGIR, pp. 172-180, 2001.
5. R. Jones, B. Rey, O. Madani, and W. Greiner, "Generating Query Substitutions," Proc. Int'l Conf. World Wide Web (WWW), 2006.

6. V. Kacholia, S. Pandit, S. Chakrabarti, S.Sudarshan, R. Desai, and H. Karambelkar, 034 International Journal of Current Research, Vol. 33, Issue, 4, pp.030-035, April, 2011 "Bidirectional Expansion for Keyword Search on Graph Databases," Proc. Int'l Conf. Very Large Data Bases (VLDB), pp. 505-516, 2005.
7. V. Hristidis, N. Koudas, Y. Papakonstantinou, and D. Srivastava, "Keyword Proximity Search in XML Trees," IEEE Trans. Knowledge and Data Eng., vol. 18, no. 4, pp. 525-539, Apr. 2006.
8. V. Hristidis, Y. Papakonstantinou, and A.Balmin, "Keyword Proximity Search on XML Graphs," Proc. IEEE Int'l Conf. Data Eng.(ICDE), pp. 367-378, 2003.
9. H. He, H. Wang, J. Yang, and P.S. Yu, "Blinks: Ranked Keyword Searches on Graphs," Proc. ACM SIGMOD Conf., pp. 305-316, 2007.
10. M. Ley DBLP, http://www.informatik.unitrier.de/ley/db/, 2009.
11. G. Li, J. Feng, J. Wang, and L. Zhou, "Effective Keyword Search for Valuable LCAs over XML Documents," Proc. ACM Int'l Conf. Information and Knowledge Management (CIKM), pp. 31-40, 2007.
12. G. Li, B.C. Ooi, J. Feng, J. Wang, and L. Zhou, "Ease: Efficient and Adaptive Keyword Search on Unstructured, Semi-Structured and Structured Data," Proc. ACM SIGMOD Conf., 2008.
13. Y. Li, C. Yu, and H.V. Jagadish, "Schema-Free XQuery," Proc. Int'l Conf. Very Large Data Bases (VLDB), 2004.
14. Z. Liu and Y. Chen, "Identifying Meaningful Return Information for XML Keyword Search," Proc. ACM SIGMOD Conf., 2007.
15. Z. Liu and Y. Chen, "Reasoning and Identifying Relevant Matches for XML Keyword Search," Proc. Int'l Conf. Very Large Data Bases (VLDB) vol. 1, no. 1, pp. 921-932, 2008.
16. G. Salton and M.J. McGill, Introduction to Modern Information Retrieval. McGraw-Hill, Inc., 1986.
17. A. Schmidt, M.L. Kersten, and M.Windhouwer, "Querying XML Documents Made Easy: Nearest Concept Queries," Proc. IEEE Int'l Conf. Data Eng. (ICDE), pp. 321-329, 2001.
18. C. Sun, C.Y. Chan, and A.K. Goenka, "Multiway SLCA-Based Keyword Search in XML Data," Proc. Int'l Conf. World Wide Web (WWW), pp. 1043-1052, 2007.
19. A. Theobald and G. Weikum, "The Index-Based XXL Search Engine for Querying XML Data with Relevance Ranking," Proc. Int'l Conf. Extending Database Technology (EDBT), 2002.
20. Y. Xu and Y. Papakonstantinou, "Efficient Keyword Search for Smallest LCAs in XML Databases," Proc. ACM SIGMOD, pp. 537-538, 2005. 035 International Journal of Current Research, Vol. 33, Issue, 4, pp.030-035, April, 2011.
21. Zhifeng Bao, Jiaheng Lu, Tok Wang Ling, Senior Member, IEEE, and Bo Chen, "Towards an Effective XML Keyword Search," Proc. IEEE Transactions on Knowledge and Data Engineering, Vol. 22, No. 8, August-2010.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 11 Issue 16 Version 1.0 September 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Energy Efficient Network Generation for Application Specific NoC

By Naveen Choudhary, M. S. Gaur , V. Laxmi
Maharana Pratap University of Agriculture and Technology, Udaipur, Rajasthan, India

Abstract - Networks-on-Chip is emerging as a communication platform for future complex SoC designs, composed of a large number of homogenous or heterogeneous processing resources. Most SoC platforms are customized to the domainspecific requirements of their applications, which communicate in a specific, mostly irregular way. The specific but often diverse communication requirements among cores of the SoC call for the design of application-specific network of SoC for improved performance in terms of communication energy, latency, and throughput. In this work, we propose a methodology for the design of customized irregular network architecture of SoC. The proposed method exploits priori knowledge of the application's communication characteristic to generate an energy optimized network and corresponding routing tables.

Keywords : *SoC, on-chip networks, application specific NoC, design methodologies, Mesh topology, interconnection network.*

GJCST Classification : *C.2.1*

Strictly as per the compliance and regulations of:

Energy Efficient Network Generation for Application Specific NoC

Naveen Choudhary [α], M. S. Gaur [Ω], V. Laxmi [β]

Abstract - Networks-on-Chip is emerging as a communication platform for future complex SoC designs, composed of a large number of homogenous or heterogeneous processing resources. Most SoC platforms are customized to the domain-specific requirements of their applications, which communicate in a specific, mostly irregular way. The specific but often diverse communication requirements among cores of the SoC call for the design of application-specific network of SoC for improved performance in terms of communication energy, latency, and throughput. In this work, we propose a methodology for the design of customized irregular network architecture of SoC. The proposed method exploits priori knowledge of the application's communication characteristic to generate an energy optimized network and corresponding routing tables.

Keywords : *SoC, on-chip networks, application specific NoC, design methodologies, Mesh topology, inter-connection network.*

I. Introduction

The shrinking feature sizes in silicon technologies is making possible the integration of complex systems-on Chip (SoC), offering a remarkable amount of computational power. In order to address the design complexity and assist reuse, these systems are usually built from predesigned and preverified building blocks like general-purpose processor, a DSP, a memory subsystem, etc. Functionality of these systems is generally captured by a set of communicating tasks at a high level of abstraction. These tasks are mapped to computational resources which are interconnected by an underlying communication infrastructure.

NoC (Dally & Towles, 2001; Benini & DeMicheli, 2002; Kumar, Jantsch, Soininen, Forsell, Millberg, Oberg, Tiensyrja & Hemani, 2002; Ogras, Hu & Marculescu, 2005) has been recently proposed by academia and industry as the preferred choice for the communication infrastructure for the on-chip communication challenges of future SoC architectures. NoC is characterized by packet switching based communication mechanism that is enabled by on-chip routers. NoC architectures can be classified as custom or regular based on their underlying communication infrastructure / topology. This communication infrastructure or topology impacts both performance and implementation costs of the system in terms of silicon area and energy consumption to a substantial extent.

A large number of NoC architectures have been proposed based on regular building patterns (Benini & DeMicheli, 2002; Kumar, Jantsch, Soininen, Forsell, Millberg, Oberg, Tiensyrja & Hemani, 2002; Natvig, 1997) like meshes, tori, k-ary n-cubes or fat trees for the implementation of on-chip networks to overcome conventional bus-based designs. However regular topologies may not be appropriate where communication requirement are not uniformly distributed across cores and links. Moreover most application specific SoCs are designed with static (or semi-static) mapping of tasks to processors or hardware cores and consequently the communication requirements of the SoC can be well characterized at design time. Therefore, the NoCs with irregular topology customized to the application's requirements is expected to be the preferred choice for application specific SoC platforms.

The routing function in NoC based systems is tightly coupled to the underlying topology defining the set of allowed paths on which packets may be sent from a sender to the destination core. The proper selection of the adequate topology and routing function form a key decision in the design of the overall NoC architecture. Conventionally, the proof of deadlock-freedom has mostly been carried out on the assumption of the regular topology (Dally & Seitz, 1987; Glass & Ni, 1992; Duato, Yalamanchili & Ni, 2003) and is far more complicated for NoC with underlying irregular topology. However in the NoC research domain some routing functions based on turn prohibition (Glass & Ni, 1992) methodology are proposed for irregular topology based NoCs such as prefix routing (Wu & Sheng, 1999), up*/down* (Schroeder et al. 1991), Left-Right (Jouraku, Funahashi, Amano & Koibuchi, 2001), L-turn (Jouraku, Funahashi, Amano & Koibuchi, 2001) and down/up (Sun, Yang, Chung & Hang, 2004).

In this paper, two genetic algorithm based heuristics are proposed for the design of energy efficient customized irregular topology Networks-on-Chip based on the applied routing function for application having IP cores with varying communication bandwidth

Author α : Department of Computer Science & Engineering, College of Technology and Engineering, Maharana Pratap University of Agriculture and Technology, Udaipur, Rajasthan, India.
E-mail : naveenc121@yahoo.com
Author Ω : Department of Computer Engineering, MNIT, Jaipur, Rajasthan, India. E-mail : gaurms@gmail.com
Author β : Department of Computer Engineering, MNIT, Jaipur, Rajasthan, India. E-mail : vlaxmi@mnit.ac.in

requirements. The presented methodologies exploit the predefined communication requirements of the application to generate energy efficient customized NoC along with the routing tables for supporting deadlock free communication. It is worth mentioning here that the topology and routing table generation are tightly coupled aspects of the NoC design and therefore optimization of only one aspect or one after another may lead to suboptimal solutions. The paper is organized as follows. A brief account of related work is presented in Section II. Communication model and architecture for Irregular NoC are defined in Section III. The proposed genetic algorithm based energy efficient NoC design methodologies are presented in Section IV. The Genetic Algorithm used in the proposed methodologies is described in Section V. Experimental results are presented in Section VI followed by a brief conclusion in Section VII.

II. Related Work

Methods to collect and analyze traffic information that can be fed as input to the bus and NoC design processes have been presented in (Lahiri et al. 2004) and (Murali & De Micheli, 2005). Mappings of cores onto standard NoC topologies have been explored in (Murali & DeMicheli, 2004; Hansson et al. 2005; Hu & Marculescu, 2003; Murali et al. 2005). In (Murali & DeMicheli, 2004; Murali et al. 2005) a floorplanner is used during the mapping process to get area and wire-length estimates. These works only select from a library of standard topologies, and cannot generate a fully customized topology. In (Hansson et al. 2005), a unified approach to mapping, routing and resource reservation has been presented.

However, the work does not explore the topology design process. Important research in macro networks has considered the topology generation problem (Ravi et al. 2001). As the traffic patterns on these networks are difficult to predict most approaches are tree-based (like spanning or Steiner trees) and only ensure connectivity with node degree constraints. These techniques cannot be directly extended to address the NoC synthesis problem.

Application-specific custom topology design has been explored in (Pinto et al. 2003; Ho & Pinkston, 2003; Ahonen et al. 2004; Srinivasan et al. 2005). The works from (Pinto et al. 2003; Ho & Pinkston, 2003), do not consider the floorplanning information during the topology design process. In (Ahonen et al. 2004), a floorplanner is used during topology design to reduce power consumption on wires. It does not consider the area and power consumption of switches in the design. Also, the number and size of network partitions are manually fed. In (Srinivasan et al. 2005), a slicing tree based floorplanner is used during the topology design process. This work assumes that the switches are located at the corners of the cores and does not consider the network components (switches, network interfaces) during the floorplanning process. Actual sizes of the cores in (Srinivasan et al. 2005; Srinivasan, & Chatha 2005) are considered only after generating their relative positions. The resulting floorplan can be extremely area inefficient when compared to the standard floorplanning process. In (Choudhary, N et al. 2010), a methodology to generate Bandwidth Aware NoC topology according to the application requirement is proposed. This methodology does floorplanning as the first step with high priority and later accomplishes topology generation with better traffic load distribution across the channels of the NoC leading to reduced congestion as well as hot spots in the topology. A range of issues in the design methods and tools for efficient synthesis of application specific Network-on-Chip interconnect for 3D SoC were addressed in (Seiculescu, Murali, Benini & De Micheli, 2009; Murali, Seiculescu, Benini & De Micheli, 2009).

In addition to the above, one of the major challenges for successful adoption of the Network-on-Chip paradigm is in reducing the energy consumed during the interaction between the IP cores. In (Hu & Marculescu, 2003; Hu & Marculescu, 2005), Hu and Marculescu have presented an energy-aware mapping algorithm to minimize the total communication energy cost for a 2-D mesh NoC architecture under real-time performance constraints. Similarly in (Choudhary, N., Gaur, M. S., Laxmi, V., Singh, V. (2010)) a deterministic methodology of order O(n2) to generate energy efficient NoC topology is proposed. This methodology also does floorplanning as the first step with highest priority as in (Choudhary, N et al. 2010). However due to its deterministic nature the methodology is not capable to generate energy optimized NoC topology for all the given applications. The work in (Choudhary, N., Gaur, M. S., Laxmi, V., Singh, V. (2010)) is extended in this paper by incorporating a genetic algorithm based heuristic in the methodology for improved and optimized NoC topology generation. The methodology proposed in this work address the issue of topology/network design with deadlock free communication for application specific homogenous or heterogeneous NoC according to communication requirements. The proposed methodology accepts application's communication characteristics and floorplanning information as input. Therefore this methodology is especially suitable for applications where optimized placement of cores in chip layout during floorplanning based on metric such as area is done in advance with highest priority.

III. Irregular NoC Communication Model Architecture

The basic platform for the proposed methodology including the basic communication model assumed along with the associated NoC architecture and routing function are described in this section. The mapping of tasks in Task graphs (Hu & Marculescu, 2003; Dick, Rhodes & Wolf, 1998) to the actual physical IP cores in the NoC topology graph (NoC) can be done with the help of intermediate mapping to Core Graph as exhibited in Figure 1. The Core Graph and NoC topology graph can be defined as follows.

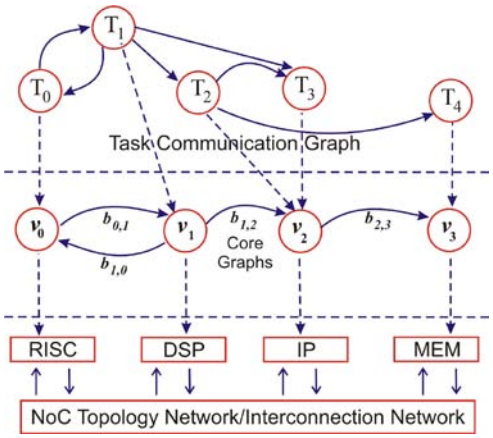


Fig.1 : Application specific communication model for NoC.

Definition 1 : Core Graph is a directed graph, G (V, E) with each vertex ν i ∈V representing an IP core and a directed edge ei,j ∈ E, representing the communication between the cores ν i and ν j. The weight of the edge ei,j denoted by bi,j, represents the desired average bandwidth requirement of the communication from ν i and ν j.

Definition 2 : NoC topology graph is a directed graph N (U, F) with each vertex υ i ∈U representing a node/tile in the topology and a directed edge fi,j ∈F represents direct communication channel between vertices υ i and υ j. Weight of the edge fi,j denoted by bi,j represents the available link/channel bandwidth across the edge fi,j.

The energy model (Hu & Marculescu, 2003) for the regular Network-on-Chip can be defined as

$$E_{bit}(t_i, t_j) = n_{hops} \times Er_{bit} + (n_{hops} - 1) \times El_{bit}$$

Follows :

(1)

Where E_{bit} (t_i, t_j) is the average dynamic energy consumption for sending one bit of data from tile ti to tile t_j, n_{hops} is the number of routers the bit traverses from tile t_i to tile t_j, Er_{bit} is the energy consumed by router for transporting one bit of data and El_{bit} is the energy consumed by link/channel for transporting one bit of data. In case of Irregular NoC with unequal length channels for transporting data, the equation (1) can be modified as follows.

$$E_{bit}(t_i, t_j) = n_{hops} \times Er_{bit} + \sum_{k=1}^{n_{hops}-1} El_{bit}^{k} \qquad (2)$$

Where the 2nd term of the summation in equation (2) represent the bit energy consumed by each channel in the route, the bit follows from communication source core to the intended destination cores.

For optimized chip layout, floorplanning according to desired metric like area can be done as a first step with the help of available floorplannning tools such as B*-Trees (Chang, Chang, Wu & Wu, 2000; Lin & Chang, 2005).

The presented work uses the escape path based routing function as proposed by (Silla & Duato, 2000). To provide deadlock free communication in the NoC, the up*/down* routing (Schroeder et al. 1991; Silla et al. 1997) and Left-Right routing (Jouraku, Funahashi, Amano & Koibuchi, 2001) were used. These routing functions assign direction to the channels of the NoC with the help of a spanning tree of the give NoC topology.

In (Silla & Duato, 2000), a generic methodology for designing adaptive routing function for Irregular NoC was proposed. The proposed methodology allow messages to follow minimal paths, in most cases, reducing message latency and increasing network throughput (Duato, Yalamanchili & Ni 2003). Moreover the methodology enforces the deadlock free route to be followed only when the minimal path is occupied by other traffic/packet. This methodology assumes that all the physical channels in the NoC can be split into two virtual channels i.e. original virtual channel and the new virtual channel. Moreover the presence of a given deadlock free routing functions based on turn prohibition (Glass & Ni, 1992) for the given irregular NoC is also assumed. The methodology further proposes to extend the given routing function in such a way that newly injected messages can use new channels without any restriction as long as the original channels are used exactly in the same way as in the original routing function. In this paper original channels are made to use deadlock free paths based on up*/down* (Left-Right) deadlock free routing functions and new channels are allowed to follow the shortest available path to the destination. The modified routing function allows a packet arriving on a new channel following shortest path to be routed to any channel without any restrictions but preferably with higher priority to new channels as new channel assure shorter paths and higher adaptively (flexibility). If no new channels are available due to congestion, one of the original channels following up*/down* (Left-Right) must be provided. However, once a packet acquires an original channel following

up*/down* (Left-Right) path, it is not allowed to do transition to a new channel anymore to avoid deadlock situation.

IV. Design Methodologies for Energy Efficient Noc Generation

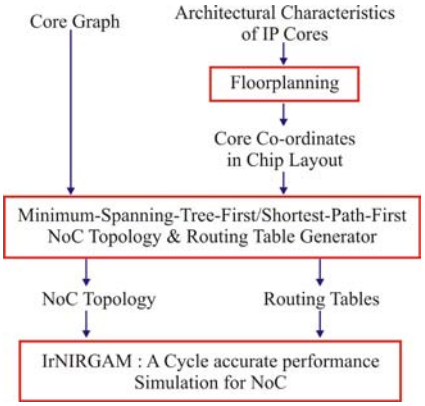


Fig.2 : Network construction using GA based

Based on the routing scheme presented by Silla et. al. (Silla & Duato, 2000), two novel genetic algorithm based methodologies referred as MSTF (minimum-spanning-tree-first) & SPF (shortest-paths-first) for energy efficient NoC topology generation are presented in this section. The presented methodologies generate an energy efficient customized NoC topology along with the required routing tables to provide deadlock free communication according to the communication requirement of the application under consideration. In both the presented methodologies, information from the floorplan and Core Graph exhibiting the chiplayout and traffic characteristics respectively are taken as inputs as exhibited in Figure 2.

Assuming over the cell routing (Srinivasan & Chatha, 2006), the link length among the nodes in the chip layout can be taken according to Manhattan distance. In both the proposed methodologies, the link/channel length is not allowed to exceed the maximum permitted channel length (e_{max}) due to constraint of physical signaling delay. This also prevents the algorithm from inserting wires that span long distances across the chip. Also, the nodes of the generated topology are not allowed to exceed a given maximum permitted node-degree (nd_{max}). This constraint prevents the algorithm from instantiating slow routers with a large number of I/O-channels that would otherwise decrease the achievable clock frequency due to internal routing and scheduling delay of the router.

a) Minimum Spanning Tree First (MSTF) Methodology

In this proposed methodology to generate the energy efficient customized topology, first a minimum spanning tree (MST) using Prim's algorithm (Cormen, Leiserson & Rivest, 1990) is generated on the nodes of the Core Graph according to information regarding the Manhattan distance from the floorplan with the constraints on nd_{max} and e_{max}. The node/core with maximum bandwidth requirement is assumed as the root of the tree. The minimum spanning tree in the topology helps us in classifying all the channels/links of the topology as "up" ("Left") or "down" ("Right"). The following phases of MSTF methodology helps in extending the network/topology for energy efficient deadlock free communication.

Energy Aware Topology Extension Phase : While keeping the constraints on nd_{max} and e_{max}, the topology is further extended by laying the shortest energy path for each traffic characteristics (edges corresponding to pair of nodes in the Core Graph). Due to constraints on nd_{max} and e_{max}, the order in which such shortest energy paths are generated basically decides the total communication energy requirement of the generated topology. The optimized order of traffic characteristics of the application is found using a genetic algorithm (refer next section). The routing tables of nodes/routers in the discovered shortest energy path are updated with the routing table entry type tag as shortest path.

Deadlock Avoidance Phase : Lastly the proposed methodology uses the modified Dijkstra's algorithm (Cormen, Leiserson & Rivest, 1990) according to up*/down* (Left Right) rule for finding deadlock free escape routing paths from each node in the shortest energy path to the corresponding destination in the generated NoC and tags them as up*/down* (Left-Right).

While taking routing decision the output channels tagged as shortest path are selected with higher priority and up*/down* (Left Right) tagged channels are selected only when no output channel corresponding to shortest path is free.

b) Shortest Path First (SPF) Methodology

SPF is similar to MSTF methodology with the exception that in SPF the topology generation is initiated by first finding the shortest energy path and later the topology is extended by constructing the MST. As in Energy Aware Topology Extension Phase of MSTF, a genetic algorithm is used to find the optimized energy-efficient traffic characteristics order of the application. Since in MSTF, MST is constructed first, it is possible that a large number of links for a number of nodes/cores in the topology are the links pertaining to MST. As maximum links emanating from a node is limited to ndmax, this phenomenon can lead to increased value of hop count in the shortest energy paths generated later leading to increased communication energy. However the SPF overcomes this drawback by creating the links pertaining to shortest energy path before the links pertaining to MST. As shortest energy paths in the topology are generated first in SPF and so there can be

a possibility that not enough number of free ports is available to construct the MST in the topology later. In such case a minimum number of ports per node/core need to be reserved before finding the shortest energy paths. However experiments showed that if communication requirement are uniformly distributed over the Core Graph then such problems are rare if any. Algorithm 1 briefly presents the proposed methodologies.

Algorithm 1 : MSTF & SPF Design Methodology

Require :

1. ℭ =CG = Core Graph = {E edges (i.e. traffic characteristics), V vertices}
2. V = {v_i | v_i is i^{th} IP core}
3. E = {e_{ij} : $v_i \rightarrow v_j$ with weight bw_{ij} | v_i (source), v_j (destination) ∈ V}
4. ℵ = NoC = {T (Topology), R (Set of routing tables), S (set of shortest path)}
5. TC_Array = {Array of traffic characteristic (i.e. ordered set of E)}
6. nd_{max} = Maximum permitted node degree in the topology T
7. e_{max} = The maximum permitted length of a link(channel) in topology T
8. Manhattan Distance = Δ = {d_{ij} | d_{ij} = $|v_i - v_j|$, v_i, v_j ∈ V}
9. Manhattan Distance greater than e_{max} are not considered.

Ensure : Energy Aware NoC Topology for CG

Procedure Minimum-Spanning-Tree-First ()

ℵ • .NoC_{EA} ;

// initialize the energy aware NoC (i.e. NoC_{EA})

• NoC_{EA}.T = Φ; NoC_{EA}.R = Φ; NoC_{EA}.S = Φ;

• Γ = {minimum spanning tree as per Δ with constraint nd_{max} & e_{max} , root is node with maximum communication in ℭ }

• NoC_{EA}.T = NoC_{EA}.T ∪ { Γ }

• (NoC_{EA}, TC_Array) = GeniticAlgo(NoC_{EA}, Γ)

• **for** each path s_i ∈ S in NoC_{EA}.S

o N = {set of nodes in path s_i}

o **for** n_j ∈ N

▪ NoC_{EA}.R =NOC_{EA}. R ∪ {update routing tables in NOC_{EA}. R for nodes ∈ V in the root followed by the shortest up*/down* (Left–Right) escape path from node n_j to the destination node of path s_i. The routing Table entry type tag is set as up*/down* (Lef –Right) for these nodes}

o endfor

• endfor

Endprocedure

Procedure Shortest-Paths-First ( )

ℵ • .NoC_{EA} ;

// initialize the energy aware NoC (i.e. NoC_{EA})

• NoC_{EA}.T = Φ; NoC_{EA}.R = Φ; NoC_{EA}.S = Φ;

• Γ = Φ ;

• (NoCEA, TC_Array) = GeniticAlgo(NoC_{EA}, Γ)

• Γ = {minimum spanning tree as per Δ with constraint nd_{max} & e_{max} , root is node with maximum communication in ℭ }

• NoC_{EA}.T = NoC_{EA}.T ∪ { Γ }

• **for** each path s_i ∈ S in NoC_{EA}.S

o N = {set of nodes in path s_i}

o **for** nj ∈ N

▪ NoC_{EA}.R =NOC_{EA}. R ∪ {update routing tables in NOC_{EA}. R for nodes ∈ V in the root followed by the shortest up*/down* (Left–Right) escape path from node nj to the destination node of path s_i. The routing table entry type tag is set as up*/down* (Lef –Right) for these nodes}

o **endfor**

• endfor

endprocedure

VII. Genetic Algorithm

A genetic algorithm (Eiben & Smith, 2003) based heuristic is used to find the best order of the traffic characteristics to generate the shortest energy paths in topology such that the communication energy requirement of the application is optimized. Genetic algorithm is a search technique used in determining exact or approximate solutions to optimization and search problems. Genetic algorithms are a particular class of evolutionary algorithms which uses techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover. The proposed genetic algorithm explores the search space extensively to generate an irregular topology with optimized communication energy requirement for the given application. The proposed genetic algorithm formulation is as follows.

a) Solution Space

In formulation of the proposed methodology, each chromosome is represented as an array of genes. Maximum size of the gene array is equal to the number of edges in the Core Graph. Each gene of the chromosome represents a traffic characteristic (an edge corresponding to a pair of nodes in the Core Graph)

b) Initial Population

A large population (i.e. 500 chromosomes) of chromosome is initially generated. The chromosomes of the initial population are generated by assigning traffic characteristics of the application to the chromosome's gene array in some random order. The initial population is later sorted according to the increasing order of total communication energy requirement of the generated topology (chromosome). It is worth highlighting here that the communication energy consumption by a chromosome varies depending on the traffic characteristics order (order of elements in gene array) of the chromosome.

c) Crossover

In each generation, crossover is performed on 50% of the population with the bias towards the Best Class of the chromosome population. For achieving crossover of two chromosomes, a random crossover point is selected. Two new chromosomes are created by

the crossover operation. The new chromosomes are created by copying the traffic characteristics (genes) from their respective parents till crossover point or from crossover point to the end of the chromosome and then the remaining traffic characteristics (genes) are copied according to the order of traffic characteristics (genes) in the other chromosome such that there are no duplicate traffic characteristics in the created chromosomes.

d) Mutation

In each generation, mutation is performed on 40% of the population to avoid the solution from getting stuck up in the local minima. Two types of mutations with probability of 50% each are performed in each generation. In first type of mutation a gene in the gene array of the chromosome with highest energy requirement is swapped with a randomly selected gene of the chromosome. In second type two randomly selected genes in the gene array of the chromosome are swapped.

e) Measure of Fitness

The cost function used to measure the fitness of the chromosomes in the population can be formulated as under.

$$Cost = Ec_i / X$$

Where X is maximum chromosome energy requirement among all the chromosomes in the population, Ec_i is the energy requirement for chromosome c_i. Fitness of chromosome is regarded as high if its cost approaches 0. It may be noted that, the best 10% chromosomes (referred as Best Class) in any generation are directly transferred to the next generation so as not to degrade the solution between the generations.

Algorithm 2 briefly presents the proposed genetic algorithm formulation. After genetic algorithm methodology is made to run for a required number of generations, the NoC topology and routing tables corresponding to the best output chromosome are accepted as the customized energy optimized application specific NoC.

Algorithm 2 : Genetic Algorithm (GA) formulation of energy aware application specific NoC generator

procedure GeniticAlgo(ℵ NoCEA, T Γ)
- μ = % of chromosomes for mutation
- ξ = % of chromosomes for crossover
- λ = % of chromosomes retained in next generation
- G = {gene array[] | size(gene array[]) = | E | (i.e. | traffic characteristics |)}
- C = chromosome = {G (set of genes), ℵ (corresponding NoC)}
- Chromosome Population = CSet = {C_i | C_i is i^{th} chromosome with gene array G_i and associated NoC ℵi}
- CSet C_{Set} = Generate_Initial_Population(NoC_{EA}, C_{Set})
- **while**(number of generations not attained)
  - Sort C_{Set} in ascending order of cost (i.e. total communication energy)
  - Keep first λ fraction chromosomes of C_{Set} for next generation as Best Class
  - Generate next ξ fraction chromosomes for next generation with crossover operations on C_{Set}
  - Select a random pair (C_1, C_2) of chromosomes from C_{Set} with bias towards Best Class
  - (C_1', C_2') = CrossOver(C_1, C_2, Γ)
  - Generate the remaining μ fraction of chromosomes for next generation with mutation operations on C_{Set}
  - Randomly Select a chromosome C_i from C_{set}
  - Select random r ∈ {1, 2}
  - Mutation(C_i, r, Γ)
- endwhile
- Sort C_{Set} in ascending order of cost (i.e. total communication energy)
- C C_{best} = C_{Set} [0]
- **return**($\aleph_{best}$ (NoC), G_{best} (Gene Array) corresponding to C_{best})

endprocedure

VIII. Experimental Results

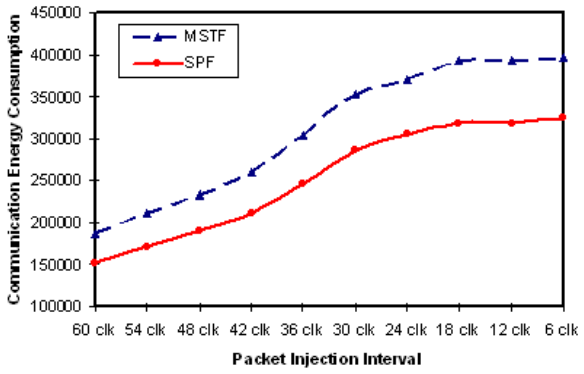
The generated energy aware application specific topology was evaluated with respect to the communication energy consumption with applied traffic load on the NoC simulation framework. In order to obtain a broad range of different irregular traffic scenarios, multiple Core Graphs using TGFF (Dick, Rhodes & Wolf, 1998) were randomly generated with diverse bandwidth requirement of the IP Cores. For performance comparison, a NoC simulator IrNIRGAM, the extended version of NIRGAM (Jain, Al-Hashimi, Gaur, Laxmi & Narayanan, 2007; Jain 2007) supporting irregular topology with the provision of supporting escape path routing for avoiding deadlock condition, was deployed. IrNIRGAM is a discrete event, cycle accurate simulator. IrNIRGAM supports irregular topology framework with source and table based routing in a wormhole switching based architecture wherein an IP Core is directly connected to a dedicated router. In IrNIRGAM, input buffered routers can have multiple virtual channels (VCs) and uses wormhole switching for flow control. The packets are split into an arbitrary number of flits (flow control units) and forwarded through the network in a pipelined fashion. A Round-Robin scheme for switch arbitration is used in the router nodes to provide fair bandwidth allocation while effectively preventing scheduling anomalies like starvation. For performance comparison on experimental set, the IrNIRGAM was run for 10000 clock cycles with applied packet injection interval to evaluate the network performance with varying traffic load. The energy consumption by the flits reaching their corresponding destination and flit latency were used as performance metric. The energy consumption by router

in transmitting a bit is evaluated using the power simulator orion (Kahng, Li, Peh & Samadi, 2009) for 0.18 μm technology. Similarly the dynamic bit energy consumption for inter-node links (Elbit) can be calculated using the following equation.

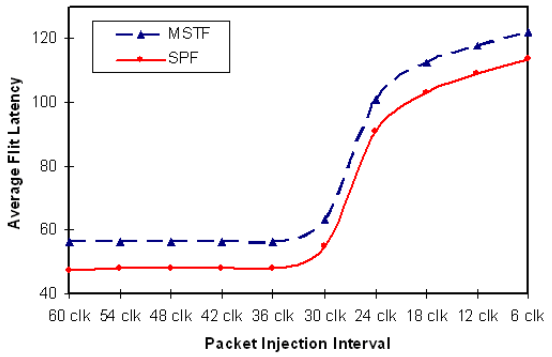
$$El_{bit} = (1/2) \times \alpha \times C_{phy} \times V_{DD}^2$$

Where α is the average probability of a 1 to 0 or 0 to 1 transition between two successive samples in the stream for a specific bit. The value of α can be taken as 0.5 assuming data stream to be purely random. Cphy is the physical capacitance of inter-node wire under consideration for the given technology and VDD is the supply voltage.

a) Experiments on SPF and MSTF with Random Benchmarks



(a)

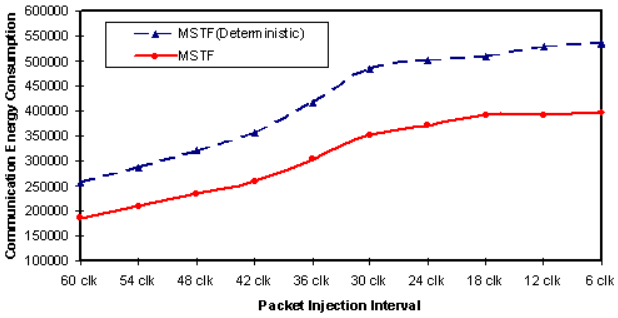


(b)

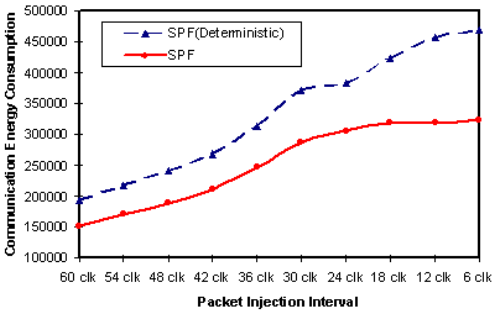
Fig.3 : Performance comparison with varying packet injection interval of (a) communication energy consumption (in pico joules) and (b) Average flit latency (in clock cycles) of the proposed Minimum-Spanning-Tree-First (MSTF) and Shortest-Path-First methodologies (SPF).

The performance of the proposed Shortest-Path-First Methodology (SPF) and Minimum-Spanning-Tree-First Methodology (MSTF) were compared on the IrNIRGAM simulation framework with varying packet injection interval (i.e. varying communication traffic load). Figure 3 shows performance results averaged over 50 generated energy efficient irregular topologies generated based on up*/down* routing function with varying number of cores from 16 to 81, ndmax = 4 and permitted channel length (emax) was taken as 1.5 times the length of the core/node with largest length among all the cores in the NoC. The proposed shortest-path-First (SPF) methodology's total dynamic communication energy consumption was on average 18.5% lesser in comparison to minimum-spanning-tree-first (MSTF) methodology in addition to reduced latency (in the range of 7.5 clocks to 10 clocks) for equivalent throughput.

b) Experiments on SPF/MSTF and SPF (Deterministic)/ MSTF (Deterministic) with Random Benchmarks



(a)



(b)

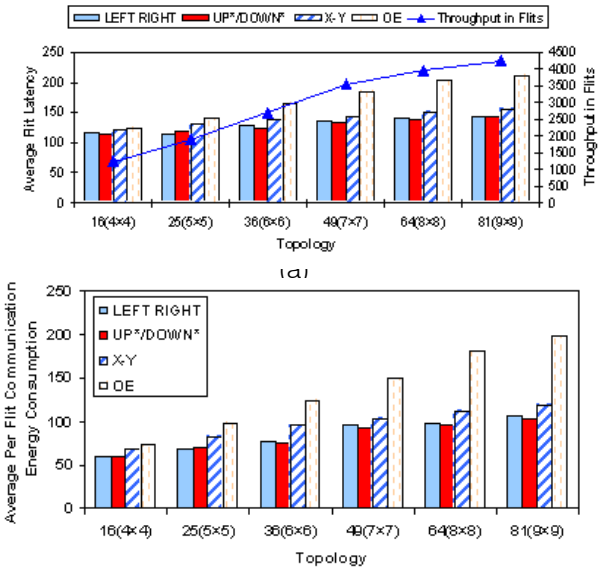
Fig.4 : Performance comparison with varying packet injection interval of dynamic communication energy consumption (in pico joules) of the (a) MSTF and MSTF (Deterministic) and (b) SPF and SPF (Deterministic).

The performance of the proposed Genetic algorithm based Shortest-Path-First Methodology (SPF) and Minimum-Spanning-Tree-First Methodology (MSTF) were compared with deterministic methodologies MSTF (Deterministic) and SPF (Deterministic) proposed in (Choudhary, N., Gaur, M. S., Laxmi, V., Singh, V., 2010) of order O(n2). IrNIRGAM simulation framework was run for 10000 clock cycles with varying packet injection interval (i.e. varying communication traffic load) . Figure 4 shows comparison of the dynamic communication energy consumption by the proposed methodologies and the work proposed in (Choudhary, N., Gaur, M. S., Laxmi, V., Singh, V., 2010). The experimental results were averaged over 50 generated customized irregular topologies generated based on up*/down* routing function with varying number of cores from 16 to 81,

nd_{max} = 4 and permitted channel length (e_{max}) was assumed as 1.5 times the length of the core/node with largest length among all the cores in the NoC. The proposed MSTF methodology's total dynamic communication energy consumption was on average 26.2% lesser in comparison to MSTF (Deterministic) whereas for the proposed SPF the total dynamic communication energy consumption was on average 24.3% lesser in comparison to SPF (Deterministic) for equivalent throughput.

c) Experiments on SPF, MSTF and Regular NoC with Random Benchmarks

To compare the performance of the proposed methodologies with regular NoC, the performance of the proposed methodologies with up*/down* and Left-Right routing function were compared with 2D-Mesh NoC with XY and OE routing for the packet injection intervals according to the application's traffic characteristics. The sizes of the tiles are kept same in the proposed methodologies as in regular 2D-Mesh. Figure 5 shows the performance comparison of MSTF with 2D-Mesh averaged over 50 generated energy efficient irregular topologies with varying number of cores from 16 to 81, nd_{max} = 4 and e_{max} was taken as 2 times the length of the core/node. The MSTF with up*/down* (Left-Right) routing shows reduced average flit latency in the range of 5.8 (4.4) clocks to 13.3 (15.2) clocks and 9.6 (8.2) clocks to 68 (67) clocks in comparison to 2D-Mesh with XY and OE routing respectively. The average per flit communication energy comparison of MSTF with 2D-Mesh shows reduction in the range of 10% (8%) to 21% (19%) and 18% (17%) to 46% (46%) in comparison to XY and OE routing respectively for up*/down* (Left-Right) routing.

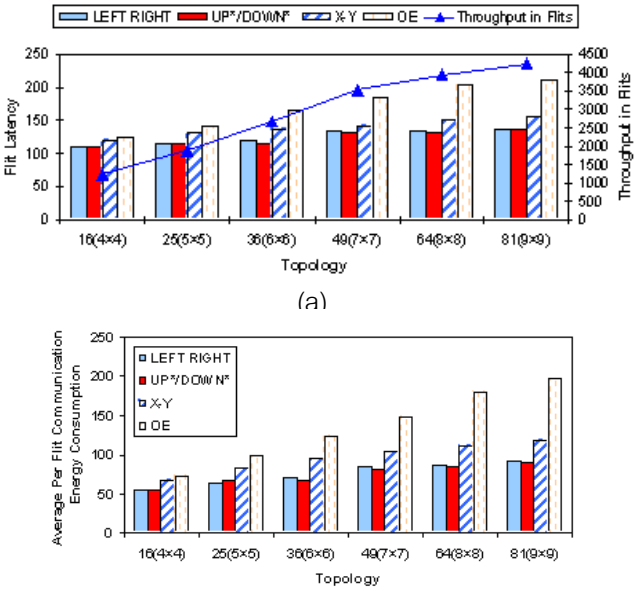


(a)

(b)

Fig.5 : MSTF performance comparison with 2D-Mesh (a) Average flit latency (in clock cycles) and (b) Average communication energy consumption per flit (in pico joules)

The average per flit communication energy comparison of SPF with 2D-Mesh shows reduction in the range of 18.8% (18.5%) to 29.2% (25.8%) and 25.2% (24.6%) to 54.7% (53%) in comparison to XY and OE routing respectively for up*/down* (Left-Right) routing.



(a)

(b)

Fig.6 : SPF performance comparison with 2D-Mesh (a) Average flit latency (in clock cycles) and (b) Average communication energy consumption per flit (in pico joules).

The above mentioned results shows that the performance of Left-Right and up*/down* routing function for MSTF and SPF depends on the traffic characteristics and the corresponding generated topology i.e. one routing function performs better than other depending on the traffic characteristic and the corresponding generated topology. However we have observed that up*/down* routing tends to perform better in most of the cases. Moreover the performance comparison between MSTF and SPF clearly shows that in most cases the SPF methodology performs reasonably better than MSTF methodology.

Figure 6 shows the SPF performance results. The SPF with up*/down* (Left-Right) routing shows reduced average flit latency in the range of 10 (9.4) clocks to 20.9 (18.4) clocks and 13.8 (13.2) clocks to 76 (69) clocks in comparison to 2D-Mesh with XY and OE routing respectively.

D. Experiments on SPF, 2D-Mesh, and BA-TGM

The per flit dynamic communication energy consumption for proposed SPF and Bandwidth Aware Topology Generation Methodology (referred as BA-TGM) presented in (Choudhary, N. et al., 2010) are compared for 50 generated customized irregular topologies with cores having varying sizes and ndmax of 4 for number of cores varying between 16 to 81. For BA-TGM, up*/down* routing was assumed whereas for SPF escape path based up*/down* routing was used. The

emax was taken as 1.5 times the length of the core having maximum length among all the cores of the NoC.

Figure 7 shows that SPF consistently performs better in comparison to BA-TGM as far as average dynamic communication energy consumption by flits reaching their destination is concerned. The SPF showed on average a reduction of 38.6% for the communication energy per flit in comparison to BA-TGM.

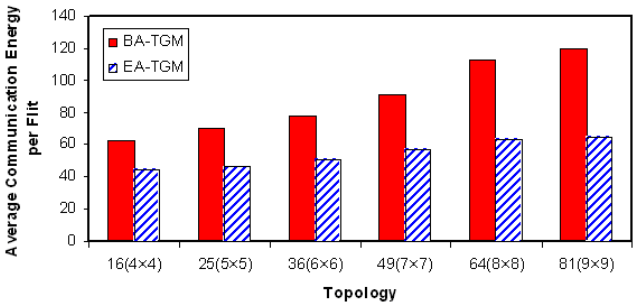


Fig.7 : Comparison of average communication energy consumed by flits in reaching their destination for BA-TGM and SPF with ndmax = 4

IX. Conclusion

In this paper, the energy efficient customized Irregular topology generation problem for NoC was addressed. Two genetic algorithm based novel methodologies are proposed for generating the NoC topology with optimized communication energy requirements according to the traffic characteristics of the given application. Although in this paper up*/down* and Left-Right routing were used as escape path for deadlock prevention, we argue that the proposed methodologies can be adapted with any of the topology agnostic routing algorithms where generic routing rules based on turn prohibition can be enforced. It is believed that the combined treatment of the routing and topology generation as done in the presented methods offers a huge potential of optimization for future application-specific NoC architectures.

Some interesting extensions of the proposed design can be to combine the topology generation with the task partitioning/scheduling into the presented framework to make the design more adaptable to the dynamic communication requirement of the application in such a way that the computation and communication energy consumption can be optimized at the same time.

References References Referencias

1. Ahonen, T. et al. (2004), Topology optimization for application specific networks on chip. In Proceedings SLIP.
2. Benini, L., & DeMicheli, G. (2002). Networks on Chips: a new SoC paradigm. In IEEE Comput. 35, 70–78.
3. Benini, L., & DeMicheli, G. (2002). Networks on Chips: a new SoC paradigm. In IEEE Comput. 35, 70–78.
4. Chang, Y. C., Chang, Y. W., Wu, G. M. Wu, S. W. (2000). B*-Trees : a new representation for non-slicing floorplans. In Proceeding of 37th Design Automation Conference, 458-463.
5. Cormen, T., Leiserson, C. & Rivest, R. (1990). Introduction to algorithms, Prentice Hall International.
6. Dally, W. J., & Towles, B. (2001). Route packets, not wires: on-chip interconnection networks. In IEEE Proceedings of the 38th Design Automation Conference (DAC), 684-689.
7. Dally, W., & Seitz, C. (1987). Deadlock-free message routing in multiprocessor interconnection networks. In IEEE Transactions on Computers, 547–553.
8. Dick, R. P., Rhodes, D. L., & Wolf, W. (1998). TGFF: task graphs for free. In Proceeding of the International Workshop on Hardware/Software Codesign.
9. Duato, J., Yalamanchili, S. & Ni, L. (2003). Interconnection networks: an engineering approach, Elsevier.
10. Eiben, A. E., & Smith, J. E. (2003). Introduction to evolutionary computing, Berlin, Heidelberg. Springer-Verlag.
11. Glass, C. & Ni, L. (1992). The turn model for adaptive routing. In Proceeding of 19¬th International Symposium on Computer Architecture, 278– 287.
12. Hansson, A. et al. (2005). A unified approach to constrained mapping and routing on network-on-chip architectures. In Proceeding of ISSS, 75-80.
13. Ho, W. H., & Pinkston, T. M. (2003). A methodology for designing efficient on-chip interconnects on well-behaved communication patterns. In HPCA, 377-388.
14. Hu, J. & Marculescu, R. (2003). energy-aware mapping for tile-based NoC architectures under performance constraints. In ASP-DAC.
15. Hu, J., & Marculescu, R. (2005). Energy- and performance-aware mapping for regular NoC architectures. In IEEE Trans. on CAD of Integrated Circuits and Systems, 24(4).
16. Jain, L., (2007) Network on Chip simulator: NIRGAM. Retrieved October 17, 2010, from http://www.nirgam.ecs.soton.ac.uk
17. Jain, L., Al-Hashimi, B. M., Gaur, M. S., Laxmi, V., & Narayanan, A. (2007). NIRGAM: a simulator for NoC interconnect routing and application modelling. In proceedings of DATE.
18. Jouraku, A., Funahashi, A., Amano, H., & Koibuchi, M. (2001). L-turn routing: an adaptive routing in

irregular networks. In Proceeding of the International Conference on Parallel Processing, 374-383.
19. Kahng, A. B., Li, B. L., Peh, S., & Samadi, K. (2009). Orion 2.0: a fast and accurate NoC power and area model for early-stage design space exploration. In Proceedings DATE, 423–428.
20. Kumar, S., Jantsch, A., Soininen, J. P., Forsell, M., Millberg, M., Oberg, J., Tiensyrja, K., & Hemani, A (2002). A network on chip architecture and design methodology. In Proceedings of VLSI Annual Symposium (ISVLSI 2002), 105–112.
21. Lahiri, K. et al. (2004). Design space exploration for optimizing on-chip communication architectures. In IEEE TCAD, 23(6), 952- 961.
22. Lin, J. M. & Chang, Y. W. (2005). TCG : A transitive closure graph-based representation of general floorplans. In IEEE Transactions on VLSI Systems, 288-292.
23. Murali, S. & De Micheli, G. (2005). An application-specific design methodology for STbus crossbar generation. In Proceedings DATE. 1176-1181.
24. Murali, S. et al. (2005). Mapping and physical planning of networks on chip architectures with quality-of-service guarantees. In Proceedings ASPDAC.
25. Murali, S., & DeMicheli, G. (2004). SUNMAP: a tool for automatic topology selection and generation for NoCs. In Proceeding of DAC.
26. Murali, S., Seiculescu, C., Benini, L., & De Micheli, G. (2009). Synthesis of networks on chips for 3d systems on chips. In Asian and South Pacific Design Automation Conference (ASPDAC), 242-247.
27. Natvig, L. (1997). High-level architectural simulation of the torus routing chip. In Proceedings of the International Verilog HDL Conference, California, 48–55.
28. Choudhary, N., Gaur, M. S., Laxmi, V., Singh, V. (2010). Fast Energy Aware Application Specific Network-on-Chip Topology Generator. In Proceeding of the IEEE International Conference IACC, Patiala, India, 250-255
29. Choudhary, N. et al. (2010). Genetic Algorithm Based Topology Generation for Application Specific Network-on-Chip. In Proceeding of the IEEE International Conference ISCAS, Paris, France, 3156-3159
30. Ogras , U., Hu, J., & Marculescu, R. (2005). Key research problems in NoC design: a holistic perspective. In IEEE CODES+ISSS, 69-74.
31. Pinto A. et al. (2003). Efficient Synthesis of Networks on Chip. In ICCD, 46-150.
32. Ravi, R. et al. (2001). Approximation algorithms for degree-constrained minimum cost network design problems. In Algorithmica, 31(1), 58-78.
33. Schroeder, M. D. et al., (1991). Autonet: a high-speed self-configuring local area network using point-to-point links. In Journal on Selected Areas in Communications, I(9).
34. Seiculescu, C., Murali, S., Benini, L., & De Micheli, G. (2009). SunFloor 3D: a tool for networks on chip topology synthesis for 3d systems on chip. In Proceedings DATE, 9-14.
35. Silla, F. et al. (1997). Efficient adaptive routing in networks of workstations with irregular topology. In Proceedings of the Workshop on Communications and Architectural Support for Network-Based Parallel Computing, 46-60.
36. Silla, F., & Duato, J. (2000). High-performance routing in networks of workstations with irregular topology. In IEEE Transactions on Parallel and Distributed Systems, I(11), 699-719.
37. Srinivasan, K. & Chatha, K. S. (2006). Layout aware design of mesh based NoC architectures. In Proceedings of 4th International Conference on Hardware Software Codesign and System Synthesis. Seoul, Korea, 36-141.
38. Srinivasan, K. et al. (2005). An automated technique for topology and route generation of application specific on-chip interconnection networks. In Proceedings ICCAD.
39. Srinivasan, K. & Chatha, K. S. (2005). ISIS: A genetic algorithm based technique for custom on-chip interconnection network synthesis. In Proceedings of 18th International Conference on VLSI Design, Kolkata, India, 623-628.
40. Sun, Y. M., Yang, C. H., Chung, Y. C., & Hang, T. Y. (2004). An efficient deadlock-free tree-based routing algorithm for irregular wormhole-routed networks based on turn model. In Proceeding of International Conference on Parallel Processing, I(1), 343-352.
41. Wu, J. & Sheng, L. (1999). Deadlock-free routing in irregular networks using prefix routing. DIMACS (Tech. Rep.), 99-19.

Global Journal of Computer Science and Technology
Volume 11 Issue 16 Version 1.0 September 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Multimodal Biometric Authentication System: Challenges and Solutions

By Shyam Sunder Yadav, Jitendra Kumar Gothwal , Prof. (Dr.) Ram Singh
Maharana Pratap University of Agriculture and Technology, Udaipur, Rajasthan, India

Abstract - Biometric technologies are automated methods for measuring and analyzing biological data, extracting a feature set from acquired data and comparing this set against to the templates set in the database. Unimodal biometric system have variety of problems such as noisy data, spool attacks etc. Multimodal biometrics refers the combination of two or more biometric modalities in a single identification. Most biometric verification systems are done based on knowledge base and token based identification these are prone to fraud. Biometric authentication employs unique combinations of measurable physical characteristics- fingerprint, facial features , iris of the eye, voice print and so on- that cannot be readily imitated or forged by others. This paper discuss the various scenarios that are possible in multi model biometric system , the level of fusion that are plausible and the integration strategic that can be adopted to consolidate information. Fusion methods include processing biometric madalitics sequential until an acceptable match is obtained.

Keywords : *Multimodal Biometrics, Authentication, Templates, Fusion, Fingerprint.*

GJCST Classification : *D.4.6, K.6.5*

Strictly as per the compliance and regulations of:

Multimodal Biometric Authentication System: Challenges and Solutions

Shyam Sunder Yadav[α], Jitendra Kumar Gothwal[Ω], Prof. (Dr.) Ram Singh[β]

Abstract - Biometric technologies are automated methods for measuring and analyzing biological data, extracting a feature set from acquired data and comparing this set against to the templates set in the database. Unimodal biometric system have variety of problems such as noisy data, spool attacks etc. Multimodal biometrics refers the combination of two or more biometric modalities in a single identification. Most biometric verification systems are done based on knowledge base and token based identification these are prone to fraud. Biometric authentication employs unique combinations of measurable physical characteristics- fingerprint, facial features , iris of the eye, voice print and so on- that cannot be readily imitated or forged by others. This paper discuss the various scenarios that are possible in multi model biometric system , the level of fusion that are plausible and the integration strategic that can be adopted to consolidate information. Fusion methods include processing biometric madalitics sequential until an acceptable match is obtained.

Keywords : Multimodal Biometrics, Authentication, Templates, Fusion, Fingerprint.

I. Introduction

The need for reliable user authentication techniques has increased in the wake of heightened concerns about security and rapid advancements in networking, communication and mobility. Biometrics, described as the science of recognizing an individual based on her physiological or behavioral traits, is beginning to gain acceptance as a legitimate method for determining an individual's identity. Biometric systems have now been deployed in various commercial, civilian and forensic applications as a means of establishing identity. These systems rely on the evidence of fingerprints, hand geometry, iris, retina, face, hand vein, facial thermo gram, signature, voice, etc. to either validate or determine an identity [2]. Most biometric systems deployed in real-world applications are unimodal, i.e., they rely on the evidence of a single source of information for authentication (e.g., single fingerprint *or* face). These systems have to contend with a variety of problems such as:

Author [α] : Research Scholar ,Department of Computer Sci. & Engg., NIMS University Jaipur ,Rajasthan, INDIA. Telephone : +91-9992443747 E-mail :ssyadav78@gmail.com

Author [Ω] : Research Scholar ,Department of Computer Sci. & Engg., NIMS University Jaipur ,Rajasthan, INDIA. Telephone : +91-941449128 8 E-mail : jkgothwal@rediffmail.com

Author [β] : Principal & Professor Department of Computer Science & Engg., MIT, Bikaner (Raj.), INDIA Rajasthan, INDIA. Telephone:+91-9460191291 E-mail : dr_ramsingh@yahoo.co.in

(a) Noise in sensed data : A fingerprint image with a scar, or a voice sample altered by cold are examples of noisy data. Noisy data could also result from defective or improperly maintained sensors (e.g., accumulation of dirt on a fingerprint sensor) or unfavorable ambient conditions (e.g., poor illumination of a user's face in a face recognition system). (b) Intra-class variations : These variations are typically caused by a user who is incorrectly interacting with the sensor (e.g., incorrect facial pose), or when the characteristics of a sensor are modified during authentication (e.g., optical versus solid-state fingerprint sensors). (c) Inter-class similarities : In a biometric system comprising of a large number of users, there may be inter-class similarities (overlap) in the feature space of multiple users. (d) Non-universality : The biometric system may not be able to acquire meaningful biometric data from a subset of users. A fingerprint biometric system, for example, may extract incorrect minutiae features from the fingerprints of certain individuals, due to the poor quality of the ridges. (e) Spoof attacks : This type of attack is especially relevant when behavioral traits such as signature or voice are used.

Some of the limitations imposed by unimodal biometric systems can be overcome by including multiple sources of information for establishing identity [5]. Such systems, known as *multimodal biometric systems*, are expected to be more reliable due to the presence of multiple, (fairly) independent pieces of evidence [7]. These systems are able to meet the stringent performance requirements imposed by various applications. In this paper we examine the levels of fusion that are plausible in a multimodal biometric system, the various scenarios that are possible, the different modes of operation, the integration strategies that can be adopted and the issues related to the design and deployment of these systems.

Biometrics refers to the automatic recognition of individuals based on their physiological and/or behavioral characteristics. Biometric technologies are becoming the foundation of an extensive array of highly secure identification and personal verification solutions. This technology acts as a front end to a system that requires precise identification before it can be accessed or used .Utilizing biometrics for personal authentication is becoming more accurate than current methods (such as the utilization of passwords or Personal Identification Number - PINs) and more convenient (nothing to carry

or remember). Thus, Biometrics is not just about security, it's also about convenience. The need for biometrics can be found in a wide range of commercial and military applications.

II. Biometric Identification System

A biometric system have five important modules: i) sensor module – which captures the trait in the form of raw biometric data, ii) feature extraction modules- which process the data to extract a feature set that is a compact representation of the trait, iii) matching module- which employs a classifier to compare the extract feature set with the stored templets to generate the matching scores, iv) decision module- which uses the matching score to either determine an identity or validate a claimed identity, v) system database module- which uses database pattern using pattern matching technique .

The main working operation that the system can perform are enrolment and testing. During enrolment biometric information of individual are stored, during test biometric information are dedected and compared with the stored ones. The sensor module the interface between real world an our system. We can say it is an image acquisition but it can change according to the characteristics we want to consider. The feature extraction module performs all the necessary preprocessing- it removes artifacts from the sensor, to enhance the input and use some kind of normalization. In the matching module we extract the features we need and choose which features to extract how to do it, with certain efficiency to create a template. After this in the matching module we are match the input pattern and the database pattern with the pattern matching technique. In the last module authentication occurs based on pattern matching technique.

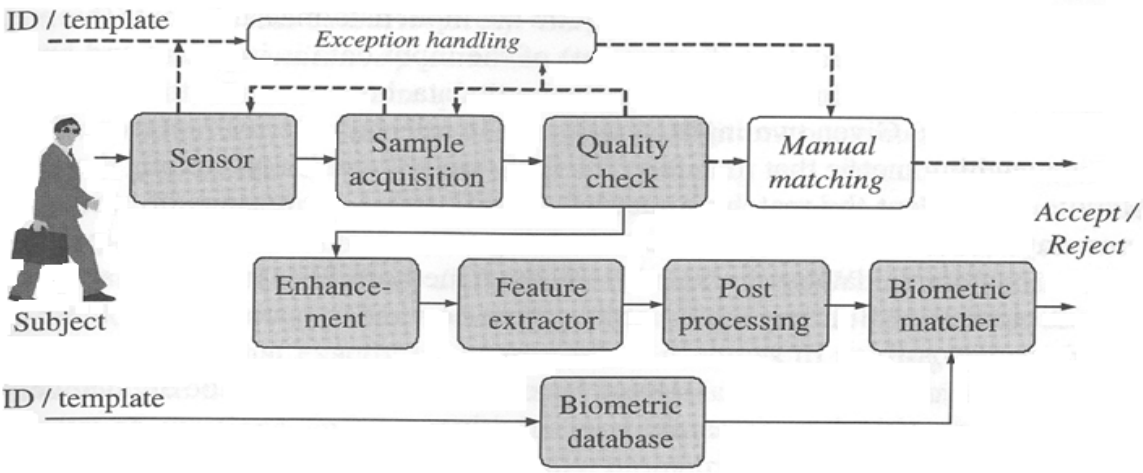


Figure 1

III. Proposed Multimodal Approach

Multimodal Biometrics System (MBS) strongly depend on the application scenario and refers to the use of a combination of two or more biometric modalities in a verification / identification system. The proposed system adopts identification based on multiple biometrics represents an emerging trend of an individual, to established the identity. The most compelling reason to combine different modalities is to improve the recognition rate. This can be done when biometric features of different biometrics are statistically independent. There are other reasons to combine two or more biometrics. One is that different biometric modalities might be more appropriate for the different applications. Another reason is simply customer preference.

The proposed system operates on five stages - stage-1 : the multiple sensor capture the raw biometric data and can be processed and integrate to generate a new data from which feature can be extracted, shown fig 2; stage-2: the preprocessor extract the necessary features that are subject to interest; stage-3: template will be generated for the extract features; stage-4: decision fusion integrate multiple cues ; stage-5: the input data will be compared with database data for matching. Finally a matching is genuine authentication is accepted, if not authentication is rejected

a) Proposed MBS Performance

The proposed system's performance is determined its accuracy. The main widely used standard metrics to determine the accuracy of a system are :

- False accept rate (FAR)
- False reject rate (FRR)
- Failure to enroll rate (FTE)
- Susceptibility to artifacts or mimics

IV. Multimodal Biometric System Architecture

Here we discussed some of the existing architectures. A Multimodal biometric system using Face & Fingerprint, they have proposed various levels of combinations of the fusion this system is shown in Fig. 2.

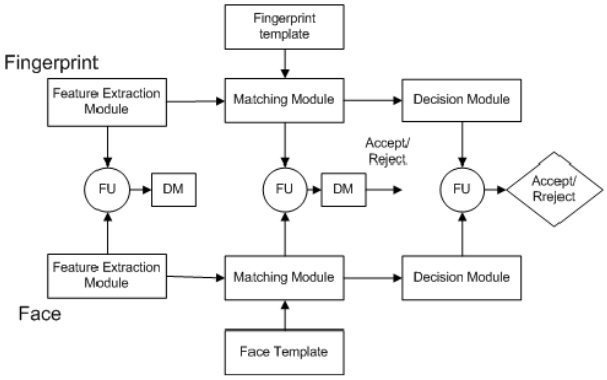


Figure 2 : Multimodal Biometric System using Face & Fingerprint

The promise of biometric technology for countering security threats Biometric authentication employs unique combinations of measurable physical characteristics--fingerprint, facial features, iris of the eye, voice print, hand geometry, vein patterns, and so on—that cannot be readily imitated or forged by others to determine or verify a person's identity. Initially the raw biometric data pertaining to multiple sensors are obtained. In our proposed system since we are using multiple biometric characters of an individual to establish identity. Here, we employ multiple sensors to Fig. 2 Proposed system an overview acquire data pertaining to different characters. The independence of the characters ensures good and reliable performance. Provide high level security by integrating the patterns by Decision level fusion.

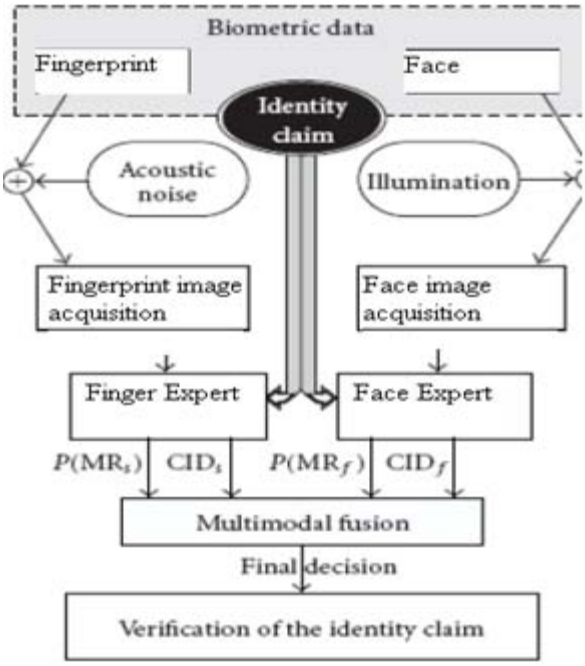


Figure 3 : Multimodal Biometric System with reliability information

V. Results

We took 09 combination sets of face images and fingerprint images from 80 users, to evaluate the performance of the proposed technique. By plotting the False Rejection Rate (FRR) against the False Accept Rate at various thresholds that summarizes the matching performance using ROC (Receiver Operating System). Using match score level fusion is 4.0 & 3.5 respectively with respect to Table i & ii, as per the databases shown in Figure 4 & 5. As expected, likelihood ratio based fusion leads to significant improvement in the performance. At a false accept rate of 0:001%, the improvement in the genuine Acceptance is achieved. FAR & FRR exits when the threshold level is >0.1

Result analysis of acceptance - Table (i)

Threshold	Finger	Face	Finger & Face
0.0	2	3	2
0.5	2	8	2
1.0	2	10	2
1.5	5	11	5
2.0	5	13	5
2.5	6	14	6
3.0	9	14	9
3.5	10	14	10
4.0	10	14	10

Receiver Operating Characteristics (ROC) Curve

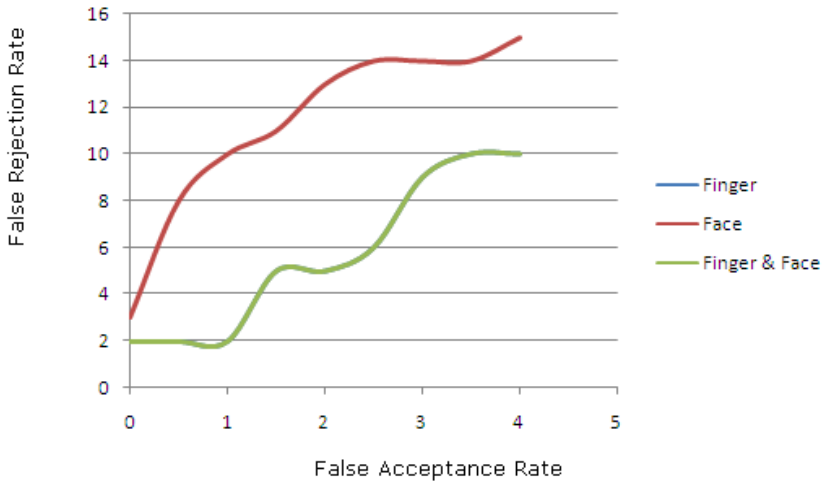


Figure 4

Result analysis of imposter

Threshold	Face	Finger	Finger & Face
0.0	4	2	2
0.5	8	3	3
1.0	14	5	5
1.5	14	8	8
2.0	14	8	8
2.5	14	9	9
3.0	14	10	10
3.5	14	10	10
4.0	14	10	10

Table (ii)

Receiver Operating Characteristics (ROC) Curve

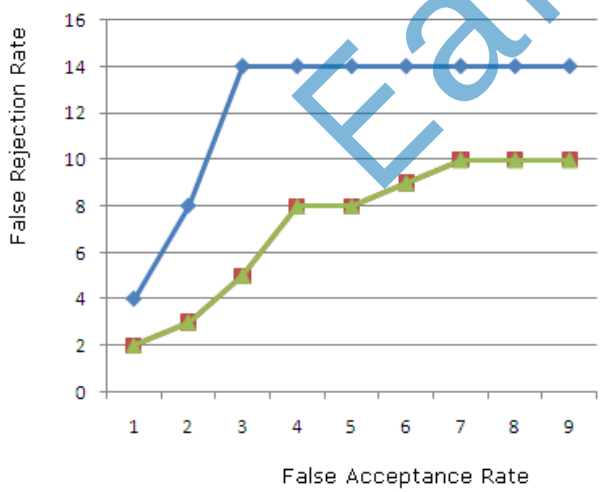


Figure 5

VI. Conclusions

Multimodal biometric systems elegantly address several of the problems present in ununimodal systems. By combining multiple sources of information, these systems improves matching performance, increase population coverage , deter spoofing and facilitate indexing . Various fusion levels and scenarios are possible in multimodal systems. Fusion at the match score level is most popular due to easy in accessing and consolidating matching scores, performance gain is pronounced when uncorrelated traits are use in multimodal system. With the wide spread deployment of biometric systems in several civilian and government applications. In applications such as border entry/exit, access control, civil identification, and network security, multi-modal biometric systems are looked to as a means of (a) reducing false acceptance and false rejection, (b) providing a secondary means of enrollment, verification, and identification if sufficient data cannot be acquired from a given biometric sample, and (c) combating attempts to spoof biometric systems through non-live data sources such as fake fingers. The performance of multimodal biometric system shows great promise to personal identity in the biometric authentication society.

References References Referencias

1. J.Wayman , A Jain, D. Maltoni, D.Maio, Biometric systems , Technology ,Degign Performance evaluation, Springer 2005.
2. A.K.Jain, A.Ross and S.Prabhakar, "An introduction to biometric recognition ", IEEE Trans. On Circuits and Systems for Video Technology, vol 14, pp. 4-20, Jan 2004.
3. Bounkong, S., Toch, B., Saad, D. and Lowe, D. (2003) ICA for watermarking digital images, Journal of Machine Learning Research, Pp. 1471-1498.
4. A.Ross , K.Nandakumar, and A.K.Jain, Handbook of Multibiometrics. Springer, 2006.

5. A.Ross and A.K. Jain, "Information fusion in biometrics", Pattern Recognition Letters, vol. 24, pp. 2115-2125, Sep 2003.
6. K. Jain and U. Uludag, "Hiding biometric data," IEEE Trans. Pattern Anal. Mach. Intelligence, vol. 25, no. 11, pp. 1493–1498, 2003.
7. L. I. Kuncheva, C.J. Whitaker, C.A. Shipp and R.P.W. Duin, "Is independence good for combining classifiers?", in Proc. of International Conf. on Pattern Recognition, vol. 2, pp. 168-171, 2000.
8. D. Maltoni, D.Maio, A.K.Jain , S.Prabahakar, Handbook of finger print recognition , Springer 2003
9. M. Indovina, U. Uludag, R.Snelick, A. Mink and A.Jain, "Multimodal Biometric Authentication methods: A COTS Approach".
10. R.M.Bolle, S.Pankati and N.K.Ratha, "Evaluation Techniques for Biometrics-Based Authentication Systems (FRR), "Proc. 15th Int'1 Conf.Pattern Recognition, vol 2, pp. 831-837, Sept.2000.
11. Teddy Ko,"Multimodal Biometric Identification for large user population using fringer print, face and iris recognition ", Proceedings of the 34th Applied Imagery and Pattern Recognition Workshop (AIPR05), 2005.
12. C.Soutar, "Biometric System Security", White paper, Bioscrypt, http://www.bioscrypt.com.
13. N.Ratha, J.Connell, and R.Bolle,"Enhancing security and Privacy in biometrics-based Authentication Systems", IBM Systems Journal, vol-40, no-3, pp-614-634, 2001.
14. R.W.Frischholz and U.Dieckmann,"Bioid: A Multimodal Biometric Identification System," IEEE Computer, vol-33,no-2, pp. 64-68, 2000.
15. Monrose, F.,Rubin,A.D.,"Keystroke Dyanamics as a Biometric for Authentication" Future Generation computer systems, vol-16, no-4(2000) 351-359.
16. A.K.Jain and A.Ross, "Learning User-Specific Parameters in a Multibiometric System",Proc. IEEE Int'1 conf. Image Processing , PP. 57-60, Sept. 2002.
17. A.K.Jain, K.Nandakumar, A.Ross, "Score normalization in multimodal biometric systems", Pattern Recognition, 2005.
18. Richard W. Hamming. Error Detecting and Error Correcting Codes Bell System Technical Journal 26(2): 147-160, 1950
19. Y. Sutcu, Q.Li, and N.Memon, "Secure Biometric Template from fingerprint-face features", in proceedings of CVPR Workshop on Biometrics , Minneapolis, USA, June 2007
20. Vetrro and N.Memon, "Biometric system security", tutorial presented at second International Conference on Biometrics, Seoul, South Korea, August 2007.

Global Journals Inc. (US) Guidelines Handbook 2011

www.GlobalJournals.org

FELLOW OF INTERNATIONAL CONGRESS OF COMPUTER SCIENCE AND TECHNOLOGY (FICCT)

- FICCT' title will be awarded to the person after approval of Editor-in-Chief and Editorial Board. The title 'FICCT" can be added to name in the following manner e.g. **Dr. Andrew Knoll, Ph.D., FICCT, Er. Pettor Jone, M.E., FICCT**
- FICCT can submit two papers every year for publication without any charges. The paper will be sent to two peer reviewers. The paper will be published after the acceptance of peer reviewers and Editorial Board.
- Free unlimited Web-space will be allotted to 'FICCT 'along with subDomain to contribute and partake in our activities.
- A professional email address will be allotted free with unlimited email space.
- FICCT will be authorized to receive e-Journals - GJCST for the Lifetime.
- FICCT will be exempted from the registration fees of Seminar/Symposium/Conference/Workshop conducted internationally of GJCST (FREE of Charge).
- FICCT will be an Honorable Guest of any gathering hold.

ASSOCIATE OF INTERNATIONAL CONGRESS OF COMPUTER SCIENCE AND TECHNOLOGY (AICCT)

- AICCT title will be awarded to the person/institution after approval of Editor-in-Chef and Editorial Board. The title 'AICCTcan be added to name in the following manner:
  eg. **Dr. Thomas Herry, Ph.D., AICCT**
- AICCT can submit one paper every year for publication without any charges. The paper will be sent to two peer reviewers. The paper will be published after the acceptance of peer reviewers and Editorial Board.
- Free 2GB Web-space will be allotted to 'FICCT' along with subDomain to contribute and participate in our activities.
- A professional email address will be allotted with free 1GB email space.
- AICCT will be authorized to receive e-Journal GJCST for lifetime.
- A professional email address will be allotted with free 1GB email space.
- AICHSS will be authorized to receive e-Journal GJHSS for lifetime.

AUXILIARY MEMBERSHIPS

ANNUAL MEMBER

- Annual Member will be authorized to receive e-Journal GJCST for one year (subscription for one year).
- The member will be allotted free 1 GB Web-space along with subDomain to contribute and participate in our activities.
- A professional email address will be allotted free 500 MB email space.

PAPER PUBLICATION

- The members can publish paper once. The paper will be sent to two-peer reviewer. The paper will be published after the acceptance of peer reviewers and Editorial Board.

Process of submission of Research Paper

The Area or field of specialization may or may not be of any category as mentioned in 'Scope of Journal' menu of the GlobalJournals.org website. There are 37 Research Journal categorized with Six parental Journals GJCST, GJMR, GJRE, GJMBR, GJSFR, GJHSS. For Authors should prefer the mentioned categories. There are three widely used systems UDC, DDC and LCC. The details are available as 'Knowledge Abstract' at Home page. The major advantage of this coding is that, the research work will be exposed to and shared with all over the world as we are being abstracted and indexed worldwide.

The paper should be in proper format. The format can be downloaded from first page of 'Author Guideline' Menu. The Author is expected to follow the general rules as mentioned in this menu. The paper should be written in MS-Word Format (*.DOC,*.DOCX).

The Author can submit the paper either online or offline. The authors should prefer online submission.Online Submission: There are three ways to submit your paper:

(A) (I) First, register yourself using top right corner of Home page then Login. If you are already registered, then login using your username and password.

(II) Choose corresponding Journal.

(III) Click 'Submit Manuscript'. Fill required information and Upload the paper.

(B) If you are using Internet Explorer, then Direct Submission through Homepage is also available.

(C) If these two are not convenient, and then email the paper directly to dean@globaljournals.org.

Offline Submission: Author can send the typed form of paper by Post. However, online submission should be preferred.

Preferred Author Guidelines

MANUSCRIPT STYLE INSTRUCTION (Must be strictly followed)

Page Size: 8.27" X 11'"

- Left Margin: 0.65
- Right Margin: 0.65
- Top Margin: 0.75
- Bottom Margin: 0.75
- Font type of all text should be Swis 721 Lt BT.
- Paper Title should be of Font Size 24 with one Column section.
- Author Name in Font Size of 11 with one column as of Title.
- Abstract Font size of 9 Bold, "Abstract" word in Italic Bold.
- Main Text: Font size 10 with justified two columns section
- Two Column with Equal Column with of 3.38 and Gaping of .2
- First Character must be three lines Drop capped.
- Paragraph before Spacing of 1 pt and After of 0 pt.
- Line Spacing of 1 pt
- Large Images must be in One Column
- Numbering of First Main Headings (Heading 1) must be in Roman Letters, Capital Letter, and Font Size of 10.
- Numbering of Second Main Headings (Heading 2) must be in Alphabets, Italic, and Font Size of 10.

You can use your own standard format also.
Author Guidelines:

1. General,

2. Ethical Guidelines,

3. Submission of Manuscripts,

4. Manuscript's Category,

5. Structure and Format of Manuscript,

6. After Acceptance.

1. GENERAL

Before submitting your research paper, one is advised to go through the details as mentioned in following heads. It will be beneficial, while peer reviewer justify your paper for publication.

Scope

The Global Journals Inc. (US) welcome the submission of original paper, review paper, survey article relevant to the all the streams of Philosophy and knowledge. The Global Journals Inc. (US) is parental platform for Global Journal of Computer Science and Technology, Researches in Engineering, Medical Research, Science Frontier Research, Human Social Science, Management, and Business organization. The choice of specific field can be done otherwise as following in Abstracting and Indexing Page on this Website. As the all Global

Journals Inc. (US) are being abstracted and indexed (in process) by most of the reputed organizations. Topics of only narrow interest will not be accepted unless they have wider potential or consequences.

2. ETHICAL GUIDELINES

Authors should follow the ethical guidelines as mentioned below for publication of research paper and research activities.

Papers are accepted on strict understanding that the material in whole or in part has not been, nor is being, considered for publication elsewhere. If the paper once accepted by Global Journals Inc. (US) and Editorial Board, will become the copyright of the Global Journals Inc. (US).

Authorship: The authors and coauthors should have active contribution to conception design, analysis and interpretation of findings. They should critically review the contents and drafting of the paper. All should approve the final version of the paper before submission

The Global Journals Inc. (US) follows the definition of authorship set up by the Global Academy of Research and Development. According to the Global Academy of R&D authorship, criteria must be based on:

1) Substantial contributions to conception and acquisition of data, analysis and interpretation of the findings.

2) Drafting the paper and revising it critically regarding important academic content.

3) Final approval of the version of the paper to be published.

All authors should have been credited according to their appropriate contribution in research activity and preparing paper. Contributors who do not match the criteria as authors may be mentioned under Acknowledgement.

Acknowledgements: Contributors to the research other than authors credited should be mentioned under acknowledgement. The specifications of the source of funding for the research if appropriate can be included. Suppliers of resources may be mentioned along with address.

Appeal of Decision: The Editorial Board's decision on publication of the paper is final and cannot be appealed elsewhere.

Permissions: It is the author's responsibility to have prior permission if all or parts of earlier published illustrations are used in this paper.

Please mention proper reference and appropriate acknowledgements wherever expected.

If all or parts of previously published illustrations are used, permission must be taken from the copyright holder concerned. It is the author's responsibility to take these in writing.

Approval for reproduction/modification of any information (including figures and tables) published elsewhere must be obtained by the authors/copyright holders before submission of the manuscript. Contributors (Authors) are responsible for any copyright fee involved.

3. SUBMISSION OF MANUSCRIPTS

Manuscripts should be uploaded via this online submission page. The online submission is most efficient method for submission of papers, as it enables rapid distribution of manuscripts and consequently speeds up the review procedure. It also enables authors to know the status of their own manuscripts by emailing us. Complete instructions for submitting a paper is available below.

Manuscript submission is a systematic procedure and little preparation is required beyond having all parts of your manuscript in a given format and a computer with an Internet connection and a Web browser. Full help and instructions are provided on-screen. As an author, you will be prompted for login and manuscript details as Field of Paper and then to upload your manuscript file(s) according to the instructions.

To avoid postal delays, all transaction is preferred by e-mail. A finished manuscript submission is confirmed by e-mail immediately and your paper enters the editorial process with no postal delays. When a conclusion is made about the publication of your paper by our Editorial Board, revisions can be submitted online with the same procedure, with an occasion to view and respond to all comments.

Complete support for both authors and co-author is provided.

4. MANUSCRIPT'S CATEGORY

Based on potential and nature, the manuscript can be categorized under the following heads:

Original research paper: Such papers are reports of high-level significant original research work.

Review papers: These are concise, significant but helpful and decisive topics for young researchers.

Research articles: These are handled with small investigation and applications

Research letters: The letters are small and concise comments on previously published matters.

5.STRUCTURE AND FORMAT OF MANUSCRIPT

The recommended size of original research paper is less than seven thousand words, review papers fewer than seven thousands words also.Preparation of research paper or how to write research paper, are major hurdle, while writing manuscript. The research articles and research letters should be fewer than three thousand words, the structure original research paper; sometime review paper should be as follows:

Papers: These are reports of significant research (typically less than 7000 words equivalent, including tables, figures, references), and comprise:

(a)Title should be relevant and commensurate with the theme of the paper.

(b) A brief Summary, "Abstract" (less than 150 words) containing the major results and conclusions.

(c) Up to ten keywords, that precisely identifies the paper's subject, purpose, and focus.

(d) An Introduction, giving necessary background excluding subheadings; objectives must be clearly declared.

(e) Resources and techniques with sufficient complete experimental details (wherever possible by reference) to permit repetition; sources of information must be given and numerical methods must be specified by reference, unless non-standard.

(f) Results should be presented concisely, by well-designed tables and/or figures; the same data may not be used in both; suitable statistical data should be given. All data must be obtained with attention to numerical detail in the planning stage. As reproduced design has been recognized to be important to experiments for a considerable time, the Editor has decided that any paper that appears not to have adequate numerical treatments of the data will be returned un-refereed;

(g) Discussion should cover the implications and consequences, not just recapitulating the results; conclusions should be summarizing.

(h) Brief Acknowledgements.

(i) References in the proper form.

Authors should very cautiously consider the preparation of papers to ensure that they communicate efficiently. Papers are much more likely to be accepted, if they are cautiously designed and laid out, contain few or no errors, are summarizing, and be conventional to the approach and instructions. They will in addition, be published with much less delays than those that require much technical and editorial correction.

The Editorial Board reserves the right to make literary corrections and to make suggestions to improve briefness.

It is vital, that authors take care in submitting a manuscript that is written in simple language and adheres to published guidelines.

Format

Language: The language of publication is UK English. Authors, for whom English is a second language, must have their manuscript efficiently edited by an English-speaking person before submission to make sure that, the English is of high excellence. It is preferable, that manuscripts should be professionally edited.

Standard Usage, Abbreviations, and Units: Spelling and hyphenation should be conventional to The Concise Oxford English Dictionary. Statistics and measurements should at all times be given in figures, e.g. 16 min, except for when the number begins a sentence. When the number does not refer to a unit of measurement it should be spelt in full unless, it is 160 or greater.

Abbreviations supposed to be used carefully. The abbreviated name or expression is supposed to be cited in full at first usage, followed by the conventional abbreviation in parentheses.

Metric SI units are supposed to generally be used excluding where they conflict with current practice or are confusing. For illustration, 1.4 l rather than 1.4 × 10-3 m3, or 4 mm somewhat than 4 × 10-3 m. Chemical formula and solutions must identify the form used, e.g. anhydrous or hydrated, and the concentration must be in clearly defined units. Common species names should be followed by underlines at the first mention. For following use the generic name should be constricted to a single letter, if it is clear.

Structure

All manuscripts submitted to Global Journals Inc. (US), ought to include:

Title: The title page must carry an instructive title that reflects the content, a running title (less than 45 characters together with spaces), names of the authors and co-authors, and the place(s) wherever the work was carried out. The full postal address in addition with the e-mail address of related author must be given. Up to eleven keywords or very brief phrases have to be given to help data retrieval, mining and indexing.

Abstract, used in Original Papers and Reviews:

Optimizing Abstract for Search Engines

Many researchers searching for information online will use search engines such as Google, Yahoo or similar. By optimizing your paper for search engines, you will amplify the chance of someone finding it. This in turn will make it more likely to be viewed and/or cited in a further work. Global Journals Inc. (US) have compiled these guidelines to facilitate you to maximize the web-friendliness of the most public part of your paper.

Key Words

A major linchpin in research work for the writing research paper is the keyword search, which one will employ to find both library and Internet resources.

One must be persistent and creative in using keywords. An effective keyword search requires a strategy and planning a list of possible keywords and phrases to try.

Search engines for most searches, use Boolean searching, which is somewhat different from Internet searches. The Boolean search uses "operators," words (and, or, not, and near) that enable you to expand or narrow your affords. Tips for research paper while preparing research paper are very helpful guideline of research paper.

Choice of key words is first tool of tips to write research paper. Research paper writing is an art.A few tips for deciding as strategically as possible about keyword search:

- One should start brainstorming lists of possible keywords before even begin searching. Think about the most important concepts related to research work. Ask, "What words would a source have to include to be truly valuable in research paper?" Then consider synonyms for the important words.
- It may take the discovery of only one relevant paper to let steer in the right keyword direction because in most databases, the keywords under which a research paper is abstracted are listed with the paper.
- One should avoid outdated words.

Keywords are the key that opens a door to research work sources. Keyword searching is an art in which researcher's skills are bound to improve with experience and time.

Numerical Methods: Numerical methods used should be clear and, where appropriate, supported by references.

Acknowledgements: Please make these as concise as possible.

References

References follow the Harvard scheme of referencing. References in the text should cite the authors' names followed by the time of their publication, unless there are three or more authors when simply the first author's name is quoted followed by et al. unpublished work has to only be cited where necessary, and only in the text. Copies of references in press in other journals have to be supplied with submitted typescripts. It is necessary that all citations and references be carefully checked before submission, as mistakes or omissions will cause delays.

References to information on the World Wide Web can be given, but only if the information is available without charge to readers on an official site. Wikipedia and Similar websites are not allowed where anyone can change the information. Authors will be asked to make available electronic copies of the cited information for inclusion on the Global Journals Inc. (US) homepage at the judgment of the Editorial Board.

The Editorial Board and Global Journals Inc. (US) recommend that, citation of online-published papers and other material should be done via a DOI (digital object identifier). If an author cites anything, which does not have a DOI, they run the risk of the cited material not being noticeable.

The Editorial Board and Global Journals Inc. (US) recommend the use of a tool such as Reference Manager for reference management and formatting.

Tables, Figures and Figure Legends

Tables: Tables should be few in number, cautiously designed, uncrowned, and include only essential data. Each must have an Arabic number, e.g. Table 4, a self-explanatory caption and be on a separate sheet. Vertical lines should not be used.

Figures: Figures are supposed to be submitted as separate files. Always take in a citation in the text for each figure using Arabic numbers, e.g. Fig. 4. Artwork must be submitted online in electronic form by e-mailing them.

Preparation of Electronic Figures for Publication

Even though low quality images are sufficient for review purposes, print publication requires high quality images to prevent the final product being blurred or fuzzy. Submit (or e-mail) EPS (line art) or TIFF (halftone/photographs) files only. MS PowerPoint and Word Graphics are unsuitable for printed pictures. Do not use pixel-oriented software. Scans (TIFF only) should have a resolution of at least 350 dpi (halftone) or 700 to 1100 dpi (line drawings) in relation to the imitation size. Please give the data for figures in black and white or submit a Color Work Agreement Form. EPS files must be saved with fonts embedded (and with a TIFF preview, if possible).

For scanned images, the scanning resolution (at final image size) ought to be as follows to ensure good reproduction: line art: >650 dpi; halftones (including gel photographs) : >350 dpi; figures containing both halftone and line images: >650 dpi.

Color Charges: It is the rule of the Global Journals Inc. (US) for authors to pay the full cost for the reproduction of their color artwork. Hence, please note that, if there is color artwork in your manuscript when it is accepted for publication, we would require you to complete and return a color work agreement form before your paper can be published.

Figure Legends: Self-explanatory legends of all figures should be incorporated separately under the heading 'Legends to Figures'. In the full-text online edition of the journal, figure legends may possibly be truncated in abbreviated links to the full screen version. Therefore, the first 100 characters of any legend should notify the reader, about the key aspects of the figure.

6. AFTER ACCEPTANCE

Upon approval of a paper for publication, the manuscript will be forwarded to the dean, who is responsible for the publication of the Global Journals Inc. (US).

6.1 Proof Corrections

The corresponding author will receive an e-mail alert containing a link to a website or will be attached. A working e-mail address must therefore be provided for the related author.

Acrobat Reader will be required in order to read this file. This software can be downloaded

(Free of charge) from the following website:

www.adobe.com/products/acrobat/readstep2.html. This will facilitate the file to be opened, read on screen, and printed out in order for any corrections to be added. Further instructions will be sent with the proof.

Proofs must be returned to the dean at dean@globaljournals.org within three days of receipt.

As changes to proofs are costly, we inquire that you only correct typesetting errors. All illustrations are retained by the publisher. Please note that the authors are responsible for all statements made in their work, including changes made by the copy editor.

6.2 Early View of Global Journals Inc. (US) (Publication Prior to Print)

The Global Journals Inc. (US) are enclosed by our publishing's Early View service. Early View articles are complete full-text articles sent in advance of their publication. Early View articles are absolute and final. They have been completely reviewed, revised and edited for publication, and the authors' final corrections have been incorporated. Because they are in final form, no changes can be made after sending them. The nature of Early View articles means that they do not yet have volume, issue or page numbers, so Early View articles cannot be cited in the conventional way.

6.3 Author Services

Online production tracking is available for your article through Author Services. Author Services enables authors to track their article - once it has been accepted - through the production process to publication online and in print. Authors can check the status of their articles online and choose to receive automated e-mails at key stages of production. The authors will receive an e-mail with a unique link that enables them to register and have their article automatically added to the system. Please ensure that a complete e-mail address is provided when submitting the manuscript.

6.4 Author Material Archive Policy

Please note that if not specifically requested, publisher will dispose off hardcopy & electronic information submitted, after the two months of publication. If you require the return of any information submitted, please inform the Editorial Board or dean as soon as possible.

6.5 Offprint and Extra Copies

A PDF offprint of the online-published article will be provided free of charge to the related author, and may be distributed according to the Publisher's terms and conditions. Additional paper offprint may be ordered by emailing us at: editor@globaljournals.org .

the search? Will I be able to find all information in this field area? If the answer of these types of questions will be "Yes" then you can choose that topic. In most of the cases, you may have to conduct the surveys and have to visit several places because this field is related to Computer Science and Information Technology. Also, you may have to do a lot of work to find all rise and falls regarding the various data of that subject. Sometimes, detailed information plays a vital role, instead of short information.

2. Evaluators are human: First thing to remember that evaluators are also human being. They are not only meant for rejecting a paper. They are here to evaluate your paper. So, present your Best.

3. Think Like Evaluators: If you are in a confusion or getting demotivated that your paper will be accepted by evaluators or not, then think and try to evaluate your paper like an Evaluator. Try to understand that what an evaluator wants in your research paper and automatically you will have your answer.

4. Make blueprints of paper: The outline is the plan or framework that will help you to arrange your thoughts. It will make your paper logical. But remember that all points of your outline must be related to the topic you have chosen.

5. Ask your Guides: If you are having any difficulty in your research, then do not hesitate to share your difficulty to your guide (if you have any). They will surely help you out and resolve your doubts. If you can't clarify what exactly you require for your work then ask the supervisor to help you with the alternative. He might also provide you the list of essential readings.

6. Use of computer is recommended: As you are doing research in the field of Computer Science, then this point is quite obvious.

7. Use right software: Always use good quality software packages. If you are not capable to judge good software then you can lose quality of your paper unknowingly. There are various software programs available to help you, which you can get through Internet.

8. Use the Internet for help: An excellent start for your paper can be by using the Google. It is an excellent search engine, where you can have your doubts resolved. You may also read some answers for the frequent question how to write my research paper or find model research paper. From the internet library you can download books. If you have all required books make important reading selecting and analyzing the specified information. Then put together research paper sketch out.

9. Use and get big pictures: Always use encyclopedias, Wikipedia to get pictures so that you can go into the depth.

10. Bookmarks are useful: When you read any book or magazine, you generally use bookmarks, right! It is a good habit, which helps to not to lose your continuity. You should always use bookmarks while searching on Internet also, which will make your search easier.

11. Revise what you wrote: When you write anything, always read it, summarize it and then finalize it.

12. Make all efforts: Make all efforts to mention what you are going to write in your paper. That means always have a good start. Try to mention everything in introduction, that what is the need of a particular research paper. Polish your work by good skill of writing and always give an evaluator, what he wants.

13. Have backups: When you are going to do any important thing like making research paper, you should always have backup copies of it either in your computer or in paper. This will help you to not to lose any of your important.

14. Produce good diagrams of your own: Always try to include good charts or diagrams in your paper to improve quality. Using several and unnecessary diagrams will degrade the quality of your paper by creating "hotchpotch." So always, try to make and include those diagrams, which are made by your own to improve readability and understandability of your paper.

15. Use of direct quotes: When you do research relevant to literature, history or current affairs then use of quotes become essential but if study is relevant to science then use of quotes is not preferable.

16. Use proper verb tense: Use proper verb tenses in your paper. Use past tense, to present those events that happened. Use present tense to indicate events that are going on. Use future tense to indicate future happening events. Use of improper and wrong tenses will confuse the evaluator. Avoid the sentences that are incomplete.

17. Never use online paper: If you are getting any paper on Internet, then never use it as your research paper because it might be possible that evaluator has already seen it or maybe it is outdated version.

18. **Pick a good study spot:** To do your research studies always try to pick a spot, which is quiet. Every spot is not for studies. Spot that suits you choose it and proceed further.

19. Know what you know: Always try to know, what you know by making objectives. Else, you will be confused and cannot achieve your target.

20. Use good quality grammar: Always use a good quality grammar and use words that will throw positive impact on evaluator. Use of good quality grammar does not mean to use tough words, that for each word the evaluator has to go through dictionary. Do not start sentence with a conjunction. Do not fragment sentences. Eliminate one-word sentences. Ignore passive voice. Do not ever use a big word when a diminutive one would suffice. Verbs have to be in agreement with their subjects. Prepositions are not expressions to finish sentences with. It is incorrect to ever divide an infinitive. Avoid clichés like the disease. Also, always shun irritating alliteration. Use language that is simple and straight forward. put together a neat summary.

21. Arrangement of information: Each section of the main body should start with an opening sentence and there should be a changeover at the end of the section. Give only valid and powerful arguments to your topic. You may also maintain your arguments with records.

22. Never start in last minute: Always start at right time and give enough time to research work. Leaving everything to the last minute will degrade your paper and spoil your work.

23. Multitasking in research is not good: Doing several things at the same time proves bad habit in case of research activity. Research is an area, where everything has a particular time slot. Divide your research work in parts and do particular part in particular time slot.

24. Never copy others' work: Never copy others' work and give it your name because if evaluator has seen it anywhere you will be in trouble.

25. Take proper rest and food: No matter how many hours you spend for your research activity, if you are not taking care of your health then all your efforts will be in vain. For a quality research, study is must, and this can be done by taking proper rest and food.

26. Go for seminars: Attend seminars if the topic is relevant to your research area. Utilize all your resources.

27. Refresh your mind after intervals: Try to give rest to your mind by listening to soft music or by sleeping in intervals. This will also improve your memory.

28. Make colleagues: Always try to make colleagues. No matter how sharper or intelligent you are, if you make colleagues you can have several ideas, which will be helpful for your research.

29. **Think technically:** Always think technically. If anything happens, then search its reasons, its benefits, and demerits.

30. Think and then print: When you will go to print your paper, notice that tables are not be split, headings are not detached from their descriptions, and page sequence is maintained.

31. Adding unnecessary information: Do not add unnecessary information, like, I have used MS Excel to draw graph. Do not add irrelevant and inappropriate material. These all will create superfluous. Foreign terminology and phrases are not apropos. One should NEVER take a broad view. Analogy in script is like feathers on a snake. Not at all use a large word when a very small one would be

sufficient. Use words properly, regardless of how others use them. Remove quotations. Puns are for kids, not grunt readers. Amplification is a billion times of inferior quality than sarcasm.

32. Never oversimplify everything: To add material in your research paper, never go for oversimplification. This will definitely irritate the evaluator. Be more or less specific. Also too, by no means, ever use rhythmic redundancies. Contractions aren't essential and shouldn't be there used. Comparisons are as terrible as clichés. Give up ampersands and abbreviations, and so on. Remove commas, that are, not necessary. Parenthetical words however should be together with this in commas. Understatement is all the time the complete best way to put onward earth-shaking thoughts. Give a detailed literary review.

33. Report concluded results: Use concluded results. From raw data, filter the results and then conclude your studies based on measurements and observations taken. Significant figures and appropriate number of decimal places should be used. Parenthetical remarks are prohibitive. Proofread carefully at final stage. In the end give outline to your arguments. Spot out perspectives of further study of this subject. Justify your conclusion by at the bottom of them with sufficient justifications and examples.

34. After conclusion: Once you have concluded your research, the next most important step is to present your findings. Presentation is extremely important as it is the definite medium though which your research is going to be in print to the rest of the crowd. Care should be taken to categorize your thoughts well and present them in a logical and neat manner. A good quality research paper format is essential because it serves to highlight your research paper and bring to light all necessary aspects in your research.

Informal Guidelines of Research Paper Writing

Key points to remember:

- Submit all work in its final form.
- Write your paper in the form, which is presented in the guidelines using the template.
- Please note the criterion for grading the final paper by peer-reviewers.

Final Points:

A purpose of organizing a research paper is to let people to interpret your effort selectively. The journal requires the following sections, submitted in the order listed, each section to start on a new page.

The introduction will be compiled from reference matter and will reflect the design processes or outline of basis that direct you to make study. As you will carry out the process of study, the method and process section will be constructed as like that. The result segment will show related statistics in nearly sequential order and will direct the reviewers next to the similar intellectual paths throughout the data that you took to carry out your study. The discussion section will provide understanding of the data and projections as to the implication of the results. The use of good quality references all through the paper will give the effort trustworthiness by representing an alertness of prior workings.

Writing a research paper is not an easy job no matter how trouble-free the actual research or concept. Practice, excellent preparation, and controlled record keeping are the only means to make straightforward the progression.

General style:

Specific editorial column necessities for compliance of a manuscript will always take over from directions in these general guidelines.

To make a paper clear

· Adhere to recommended page limits

Mistakes to evade

- Insertion a title at the foot of a page with the subsequent text on the next page

- Separating a table/chart or figure - impound each figure/table to a single page
- Submitting a manuscript with pages out of sequence

In every sections of your document

· Use standard writing style including articles ("a", "the," etc.)

· Keep on paying attention on the research topic of the paper

· Use paragraphs to split each significant point (excluding for the abstract)

· Align the primary line of each section

· Present your points in sound order

· Use present tense to report well accepted

· Use past tense to describe specific results

· Shun familiar wording, don't address the reviewer directly, and don't use slang, slang language, or superlatives

· Shun use of extra pictures - include only those figures essential to presenting results

Title Page:

Choose a revealing title. It should be short. It should not have non-standard acronyms or abbreviations. It should not exceed two printed lines. It should include the name(s) and address (es) of all authors.

Abstract:

The summary should be two hundred words or less. It should briefly and clearly explain the key findings reported in the manuscript--must have precise statistics. It should not have abnormal acronyms or abbreviations. It should be logical in itself. Shun citing references at this point.

An abstract is a brief distinct paragraph summary of finished work or work in development. In a minute or less a reviewer can be taught the foundation behind the study, common approach to the problem, relevant results, and significant conclusions or new questions.

Write your summary when your paper is completed because how can you write the summary of anything which is not yet written? Wealth of terminology is very essential in abstract. Yet, use comprehensive sentences and do not let go readability for briefness. You can maintain it succinct by phrasing sentences so that they provide more than lone rationale. The author can at this moment go straight to

shortening the outcome. Sum up the study, with the subsequent elements in any summary. Try to maintain the initial two items to no more than one ruling each.

- Reason of the study - theory, overall issue, purpose
- Fundamental goal
- To the point depiction of the research
- Consequences, including definite statistics - if the consequences are quantitative in nature, account quantitative data; results of any numerical analysis should be reported
- Significant conclusions or questions that track from the research(es)

Approach:

- Single section, and succinct
- As a outline of job done, it is always written in past tense
- A conceptual should situate on its own, and not submit to any other part of the paper such as a form or table
- Center on shortening results - bound background information to a verdict or two, if completely necessary
- What you account in an conceptual must be regular with what you reported in the manuscript
- Exact spelling, clearness of sentences and phrases, and appropriate reporting of quantities (proper units, important statistics) are just as significant in an abstract as they are anywhere else

Introduction:

The **Introduction** should "introduce" the manuscript. The reviewer should be presented with sufficient background information to be capable to comprehend and calculate the purpose of your study without having to submit to other works. The basis for the study should be offered. Give most important references but shun difficult to make a comprehensive appraisal of the topic. In the introduction, describe the problem visibly. If the problem is not acknowledged in a logical, reasonable way, the reviewer will have no attention in your result. Speak in common terms about techniques used to explain the problem, if needed, but do not present any particulars about the protocols here. Following approach can create a valuable beginning:

- Explain the value (significance) of the study
- Shield the model - why did you employ this particular system or method? What is its compensation? You strength remark on its appropriateness from a abstract point of vision as well as point out sensible reasons for using it.
- Present a justification. Status your particular theory (es) or aim(s), and describe the logic that led you to choose them.
- Very for a short time explain the tentative propose and how it skilled the declared objectives.

Approach:

- Use past tense except for when referring to recognized facts. After all, the manuscript will be submitted after the entire job is done.
- Sort out your thoughts; manufacture one key point with every section. If you make the four points listed above, you will need a least of four paragraphs.
- Present surroundings information only as desirable in order hold up a situation. The reviewer does not desire to read the whole thing you know about a topic.
- Shape the theory/purpose specifically - do not take a broad view.
- As always, give awareness to spelling, simplicity and correctness of sentences and phrases.

Procedures (Methods and Materials):

This part is supposed to be the easiest to carve if you have good skills. A sound written Procedures segment allows a capable scientist to replacement your results. Present precise information about your supplies. The suppliers and clarity of reagents can be helpful bits of information. Present methods in sequential order but linked methodologies can be grouped as a segment. Be concise when relating the protocols. Attempt for the least amount of information that would permit another capable scientist to spare your outcome but be cautious that vital information is integrated. The use of subheadings is suggested and ought to be synchronized with the results section. When a technique is used that has been well described in another object, mention the specific item describing a way but draw the basic

principle while stating the situation. The purpose is to text all particular resources and broad procedures, so that another person may use some or all of the methods in one more study or referee the scientific value of your work. It is not to be a step by step report of the whole thing you did, nor is a methods section a set of orders.

Materials:

- Explain materials individually only if the study is so complex that it saves liberty this way.
- Embrace particular materials, and any tools or provisions that are not frequently found in laboratories.
- Do not take in frequently found.
- If use of a definite type of tools.
- Materials may be reported in a part section or else they may be recognized along with your measures.

Methods:

- Report the method (not particulars of each process that engaged the same methodology)
- Describe the method entirely
- To be succinct, present methods under headings dedicated to specific dealings or groups of measures
- Simplify - details how procedures were completed not how they were exclusively performed on a particular day.
- If well known procedures were used, account the procedure by name, possibly with reference, and that's all.

Approach:

- It is embarrassed or not possible to use vigorous voice when documenting methods with no using first person, which would focus the reviewer's interest on the researcher rather than the job. As a result when script up the methods most authors use third person passive voice.
- Use standard style in this and in every other part of the paper - avoid familiar lists, and use full sentences.

What to keep away from

- Resources and methods are not a set of information.
- Skip all descriptive information and surroundings - save it for the argument.
- Leave out information that is immaterial to a third party.

Results:

The principle of a results segment is to present and demonstrate your conclusion. Create this part a entirely objective details of the outcome, and save all understanding for the discussion.

The page length of this segment is set by the sum and types of data to be reported. Carry on to be to the point, by means of statistics and tables, if suitable, to present consequences most efficiently.You must obviously differentiate material that would usually be incorporated in a study editorial from any unprocessed data or additional appendix matter that would not be available. In fact, such matter should not be submitted at all except requested by the instructor.

Content

- Sum up your conclusion in text and demonstrate them, if suitable, with figures and tables.
- In manuscript, explain each of your consequences, point the reader to remarks that are most appropriate.
- Present a background, such as by describing the question that was addressed by creation an exacting study.
- Explain results of control experiments and comprise remarks that are not accessible in a prescribed figure or table, if appropriate.
- Examine your data, then prepare the analyzed (transformed) data in the form of a figure (graph), table, or in manuscript form.

What to stay away from

- Do not discuss or infer your outcome, report surroundings information, or try to explain anything.
- Not at all, take in raw data or intermediate calculations in a research manuscript.

- Do not present the similar data more than once.
- Manuscript should complement any figures or tables, not duplicate the identical information.
- Never confuse figures with tables - there is a difference.

Approach

- As forever, use past tense when you submit to your results, and put the whole thing in a reasonable order.
- Put figures and tables, appropriately numbered, in order at the end of the report
- If you desire, you may place your figures and tables properly within the text of your results part.

Figures and tables

- If you put figures and tables at the end of the details, make certain that they are visibly distinguished from any attach appendix materials, such as raw facts
- Despite of position, each figure must be numbered one after the other and complete with subtitle
- In spite of position, each table must be titled, numbered one after the other and complete with heading
- All figure and table must be adequately complete that it could situate on its own, divide from text

Discussion:

The Discussion is expected the trickiest segment to write and describe. A lot of papers submitted for journal are discarded based on problems with the Discussion. There is no head of state for how long a argument should be. Position your understanding of the outcome visibly to lead the reviewer through your conclusions, and then finish the paper with a summing up of the implication of the study. The purpose here is to offer an understanding of your results and hold up for all of your conclusions, using facts from your research and generally accepted information, if suitable. The implication of result should be visibly described. Infer your data in the conversation in suitable depth. This means that when you clarify an observable fact you must explain mechanisms that may account for the observation. If your results vary from your prospect, make clear why that may have happened. If your results agree, then explain the theory that the proof supported. It is never suitable to just state that the data approved with prospect, and let it drop at that.

- Make a decision if each premise is supported, discarded, or if you cannot make a conclusion with assurance. Do not just dismiss a study or part of a study as "uncertain."
- Research papers are not acknowledged if the work is imperfect. Draw what conclusions you can based upon the results that you have, and take care of the study as a finished work
- You may propose future guidelines, such as how the experiment might be personalized to accomplish a new idea.
- Give details all of your remarks as much as possible, focus on mechanisms.
- Make a decision if the tentative design sufficiently addressed the theory, and whether or not it was correctly restricted.
- Try to present substitute explanations if sensible alternatives be present.
- One research will not counter an overall question, so maintain the large picture in mind, where do you go next? The best studies unlock new avenues of study. What questions remain?
- Recommendations for detailed papers will offer supplementary suggestions.

Approach:

- When you refer to information, differentiate data generated by your own studies from available information
- Submit to work done by specific persons (including you) in past tense.
- Submit to generally acknowledged facts and main beliefs in present tense.

Administration Rules Listed Before Submitting Your Research Paper to Global Journals Inc. (US)

Please carefully note down following rules and regulation before submitting your Research Paper to Global Journals Inc. (US):

Segment Draft and Final Research Paper: You have to strictly follow the template of research paper. If it is not done your paper may get rejected.

- The **major constraint** is that you must independently make all content, tables, graphs, and facts that are offered in the paper. You must write each part of the paper wholly on your own. The Peer-reviewers need to identify your own perceptive of the concepts in your own terms. NEVER extract straight from any foundation, and never rephrase someone else's analysis.

- Do not give permission to anyone else to "PROOFREAD" your manuscript.

- Methods to avoid Plagiarism is applied by us on every paper, if found guilty, you will be blacklisted by all of our collaborated research groups, your institution will be informed for this and strict legal actions will be taken immediately.)
- To guard yourself and others from possible illegal use please do not permit anyone right to use to your paper and files.

Criterion for Grading a Research Paper (Compilation) by Global Journals Inc. (US)

Please note that following table is only a Grading of "Paper Compilation" and not on "Performed/Stated Research" whose grading solely depends on Individual Assigned Peer Reviewer and Editorial Board Member. These can be available only on request and after decision of Paper. This report will be the property of Global Journals Inc. (US).

Topics	Grades		
	A-B	**C-D**	**E-F**
Abstract	Clear and concise with appropriate content, Correct format. 200 words or below	Unclear summary and no specific data, Incorrect form Above 200 words	No specific data with ambiguous information Above 250 words
Introduction	Containing all background details with clear goal and appropriate details, flow specification, no grammar and spelling mistake, well organized sentence and paragraph, reference cited	Unclear and confusing data, appropriate format, grammar and spelling errors with unorganized matter	Out of place depth and content, hazy format
Methods and Procedures	Clear and to the point with well arranged paragraph, precision and accuracy of facts and figures, well organized subheads	Difficult to comprehend with embarrassed text, too much explanation but completed	Incorrect and unorganized structure with hazy meaning
Result	Well organized, Clear and specific, Correct units with precision, correct data, well structuring of paragraph, no grammar and spelling mistake	Complete and embarrassed text, difficult to comprehend	Irregular format with wrong facts and figures
Discussion	Well organized, meaningful specification, sound conclusion, logical and concise explanation, highly structured paragraph reference cited	Wordy, unclear conclusion, spurious	Conclusion is not cited, unorganized, difficult to comprehend
References	Complete and correct format, well organized	Beside the point, Incomplete	Wrong format and structuring

INDEX

save our planet

Global Journal of Computer Science and Technology

Visit us on the Web at www.GlobalJournals.org | www.ComputerResearch.org
or email us at helpdesk@globaljournals.org

ISSN 9754350