Global Journal of Computer Science and Technology

discovering thoughts and inventing future

Volume 10 Issue 2 Version 1.0 Online ISSN: 0975-4172 Print ISSN: 0975-4350

14 Technology Changing Ideas

highlights

- Locomotion and Devices Control
- Compiling Mechanical Nanocomputer
- Software Development Life Cycle
- Recognition of the Making on IC

April 2010

ENG

Global Journal of Computer Science and Technology

Global Journal of Computer Science and Technology

Volume 10 Issue 2 (Ver. 1.0)

Global Academy of Research and Development

Editorial Board Members

Dr. Carlos García Pont
Associate Professor of Marketing
IESE Business School, University of Navarra
Doctor of Philosophy (Management),
Massachussetts 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 Regina
Ph.D., M.Sc. in Mathematics
B.A. (Honors) in Mathematics
University of Windsor

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. 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 Neurosciences
Northwestern University Feinberg School of Medicine

Chief Author

Dr. R.K. Dixit (HON.)
M.Sc., Ph.D., FICCT
Chief Author, India
Email: authorind@computerresearch.org

Dean & Editor-in-Chief (HON.)

Vivek Dubey(HON.)
MS (Industrial Engineering),
MS (Mechanical Engineering)
University of Wisconsin
FICCT
Editor-in-Chief, USA
editorusa@computerresearch.org

Sangita Dixit
M.Sc., FICCT
Dean and Publisher, India
deanind@computerresearch.org

Er. Suyog Dixit
BE (HONS. in Computer Science),
FICCT
SAP Certified Consultant
Technical Dean, India
Website: www.suyogdixit.com
Email:suyog@suyogdixit.com,
dean@computerresearch.org

Contents of the Volume

From the Chief Author's Desk

The research activities among different disciplines of natural science are backbone of system. The deep and strong affords are the demands of today. Sincere afford must be exposed worldwide. Which, in turns, require international platform for rapid and proper communication among similar and interdisciplinary research groups.

The Global Journal of Computer Science and Technology is to fulfill all such demands and requirements, and functions also as an international platform. Of course, the publication of research work must be reviewed to establish its authenticity. This helps to promote research activity also. We know, great scientific research have been worked out by philosopher seeking to verify quite erroneous theories about the nature of things.

The research activities are increasing exponentially. These great increments require rapid communication, also to link up with others. The balanced communication among same and interdisciplinary research groups is major hurdle to aware with status of any research field.

The Global Journals is proving as milestone of research publication. In view of whole spectrum of Knowledge, the research work of different streams may be considered as branches of big tree. Every branch is of great importance. Thus, we look after the complete spectrum as whole. Global Journals let play all the instruments simultaneously. We hope, affords of Global Journals will sincerely help to build the world in new shape.

Dr. R. K. Dixit
Chief Author
chiefauthor@globaljournals.org

Drop Out Feature of Student Data for Academic Performance Using Decision Tree Techniques

Mr. M. N. Quadri1

Dr. N.V. Kalyankar[2]

GJCST Computing Classification
H.2.8 & K.3.m

Abstract- **Students' academic performance is critical for educational institutions because strategic programs can be planned in improving or maintaining students' performance during their period of studies in the institutions. The academic performance in this study is measured by their cumulative grade point average (CGPA) upon graduating. This study presents the work of data mining in predicting the drop out feature of students. This study applies decision tree technique to choose the best prediction and analysis. The list of students who are predicted as likely to drop out from college by data mining is then turned over to teachers and management for direct or indirect intervention.**

Keywords- Intruder; hacker; cracker; Intrusion detection; anomaly detection; verification; validation.

I INTRODUCTION

Data mining is the process of analyzing data from different perspectives and summarizing the results as useful information. It is defined as "The nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data.

Due to rapid advancement in the field of information technology, the amount of information stored in educational databases is rapidly increasing. These huge databases contain a wealth of data and constitute a potential goldmine of valuable information. As new courses and new colleges emerge in environment, the structure of the educational database changes. Finding the valuable information hidden in those databases and identifying and constructing appropriate models is a difficult task. Data mining techniques play an important role at each stop of the information discovery process.

Nowadays, higher educational organizations are placing in a very high competitive environment and are aiming to get more competitive advantages over the other business competitions. These organizations should improve the quality of their services and satisfy their customers. They consider students and teachers as their main assets and they want to.

Manuscript received "Thu, Feb 11, 2010 at 9:05 PM GMT"

Mr. M.N. Quadri is working as Lecturer with Dept. of Computer Science Yeshwant College Nanded
E-Mail : mnq_1977@yahoo.com
Dr. N.V. Kalyankar is working as Principal, Yeshwant College Nanded.
E-Mail: drkalyankarnv@rediffmail.com

improve their key process indicators by effective and efficient use of their assets

The most striking features of data mining techniques are clustering and prediction. The clustering aspect of data mining offers comprehensive characteristics analysis of students, while the predicting function estimate the likelihood for a variety of outcomes of them, such as transferability, persistence, retention and success in classes.

This study makes use of decision tree analysis to analyze the problem of drop outs in any educational institution.

Decision tree analysis is a popular data mining technique that can be used in many areas of education. In this study, decision trees are used to make important design decisions and explain the interdependencies among the properties of drop out students. This study also provides examples of how data mining technique can be used to improve the effectiveness and efficiency of the modeling process.

This study is an extension of the educational model developed and published in the information technology journal[1] . The main contribution in this study is addressing the capabilities and strengths of data mining technology in identifying drop out students and to guide the teachers to concentrate on appropriate features associated and counsel the students or arrange for financial aid to them.

II APPLICATIONS OF DATA MINING IN EDUCATIONAL INDUSTRY

Identify risk factors that predict results: One critical question in any educational institution is the following "What are the risk factors or variables that are important for predicting the results (pass/fail) of students?". Although many risk factors that affect results are obvious, subtle and non-intuitive relationships can exist among variable that are difficult, for not impossible to identify without applying more sophisticated analysis.

Modern data mining models such as decision trees can more accurately predict risk than current models, educational institutions can predict the results more accurately, which in turn can result in quality education.

Student Level Analysis: Successfully training the student requires analyzing the data at the student level. Using the associated discovery data mining technique, educational institutions can more accurately select the kind of training to offer to different kinds of students. With the help of this technique, educational institutions can.

i. Segment the student database to create student profiles.

ii. Conduct analysis on a single student segment for a single factor. For example, "the institution can perform in-depth analysis of the relationship between attendance and academic achievement".

iii. Analyze the student segments for multiple factors using group processing and multiple target variables. For example, "What are the characters shared by students who drop out from colleges?".

iv. Perform sequential (over time) basket analysis on student segments. For example, "What percentage of high attendance holders also achieved in academic side also?".

Developing new strategies: Teachers can increase the pass percentage by identifying the most lucrative student segments and organize the training sessions accordingly. The results may be affected, if teachers do not offer the "right" kind of training to the "right" student segment at the "right" time. With data mining operations such as segmentation or association analysis, institutions can now utilize all of their available information for betterment of students.

III DROP OUT

Graduation, especially timely graduation is an increasingly important policy issue[2]. College graduates earn twice as much as high school graduates and six times as much as college dropouts[3]. In addition to the financial rewards, the spouses of college graduates are more educated and their children do better in schools and colleges. Graduation rates are considered as one of the institutional effectiveness[4]. Students drop out due to different reasons; academic trouble, academic preferences, marriage (girls) and their financial position.

i. Students are unable to get into the major they prefer when they matriculate and therefore they find it difficult to carry on with the course and may leave the institution due to academic trouble.

ii. Students also drop out due to academic preferences. Generally, students choose majors offering the greatest stream of future earnings.

iii. In Indian society, girls are expected to get married at the age of 18 and they may drop out when they are married.

iv. Financial position of the students plays an important role in drop out percentage.

It is important to understand the determinants of successful and timely degree completion. Most studies of student departure focus on the characteristics of students as determinants of success. The study considers the features such as gender, attendance, previous semester grade, parent education, parent income, scholarship, first child, and part time job.

Parental income: Is an important determinant of the demand for education. Students from higher-income families are less likely to have to drop out to work to finance their education and are most likely to have aspirations that promote persistence. Empirical studies indicate a strong positive correlation between family income and other family background measures on educational attainment enrollment, persistence and graduation[5,6].

Parental education: Plays an important role. Children of college graduates fare well in their exams and are less likely to drop out. A student's previous semester grade and attendance are also included in the study. Grades and attendance may have some tangible value that can be used for future educational and career mobility. Grades may also be considered as an indication of realized academic potential.

Financial aid/scholarship plays an important role in higher education by lowering the costs of attendance. The study measures the effect of financial aid/scholarship on student departure. The study also investigates about other information such as whether the student is the 'first child' in the family and he/she is doing part time job to support the family. Both these variables are expected to be positively correlated with graduation.

IV PREDICTIVE DATA MINING

Decision Trees: A tree diagram contains the following

i. Root node-top node in the tree that contains all observations

ii. Internal nodes-non-terminal nodes (including the root node) that contain the splitting nodes.

iii. Leaf nodes-terminal does that contain the final classification for a set of observations.

Decision trees are part of the induction class of data mining techniques[7]. An empirical tree represents a segmentation of the data that is created by applying a series of simple rules. Each rule assigns an observation to a segment based on the value of one input. One rule is applied after another, resulting in a hierarchy of segments within segments. The hierarchy is called a tree and each segment is called a node.

The original segment contains the entire data set and is called the root node of the tree. A node with all its successors forms a branch of the node that created it. The final nodes are leaves. For each leaf, a decision is made and applied to all observations in the leaf. The type of decision depends on the context. In predictive modeling, the decision is simply the predicted value.

The tree techniques provide insights into the decision making process[8]. The decision tree is efficient and is thus suitable for large/small data sets. They are perhaps the most successful exploratory method for uncovering deviant data structure. Trees recursively partition the input data space in order to identify segments where the records are homogeneous.

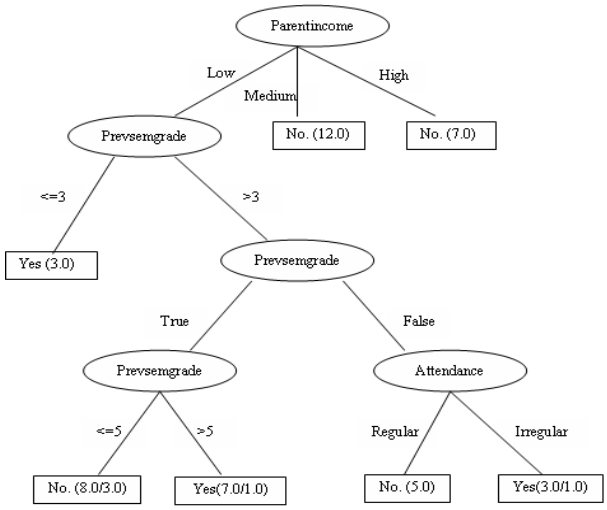
This model, make use of the software Weka The J4.8 algorithm(J4.8 implements a later and slightly improved version called C4.5) is used for predictive data mining.

Modeling student drop outs: The modeling process starts by studying the relationship between student drop outs and underlying risk factors including gender, attendance, previosus semester grade, parent education, parent income, scholarship, first child and whether the student is working or not.

A hybrid method is developed for this study-the modeling process is a combination of the decision tree techniques and logistic regression. First, the decision tree algorithm is used to identify the factors that influence dropouts. After the factors are identified, the logistic regression technique is used to quantify the dropouts and the effect of each risk factor.

The Table 1 shows the variable that influence the drop outs

Variable	Variable Type	Description
Gender	Nominal	Male, Female
Attendance	Nominal	Regular, Irregular
Prevsemgrade	Numeric	1..10
Parentedn	Nominal	Educated, Not educated
Parentincome	Nominal	Low,Medium,High
Scholarship	Nominal	Getting, Not getting
Firstchild	Nominal	True,False
Working	Nominal	Working, Not working
Dropout	Nominal	Yes,No



The Fig. 1 shows the tree diagram for analysis.

The drop out frequency varies with the most important risk factor parent income (in this study) among all the other variables. The low-income level has great influence on the drop out feature, than the medium and high-income levels.

The fact that whether the child is a first child in the family also has an influence on the drop out feature. Through the previous semester grade is above 5, the drop out feature seem to be high due to the responsibility of the student as a first child (for male). Girl students face the problem of getting married as a 'first child' of the family. Based on tree analysis, gender, parent education, scholarship, part time job are the irrelevant factors. They should not be included in the claim frequency model.

Based on tree analysis, logistic regression is used to estimate the probability of drop out feature based on the factors under consideration. Logistic regression attempts to predict the probability of drop out feature as a function of one or more independent inputs.

V Analysis

The analyzer component incorporates a number of machine learning methods for automatically analyzing the data in the log database. In addition to getting a better insight into the underlying relationship in the data, this also allows for prediction and classification of future sessions. Many machine-learning methods provide their output in an intelligible, human readable form. For instance, methods for generating decision trees from data such as C4.5 [9], allow for a tree-shaped representation of the learning results.

The aim of using the decision tree method was to characterize the students' motivation in terms of the other attributes that are automatically generated from the log data, in order to provide an abstracted view on the underlying data as well as to allow for predicting motivational aspects in other students.

VI Conclusion

This study introduced the data mining approach to modeling drop out feature and some implementation of this approach

The key to gaining a competitive advantage in the educational industry is found in recognizing that student databases, if properly managed, analyzed and exploited, are unique, valuable assets. Data mining uses predictive modeling, database segmentation, market basket analysis and combinations to more quickly answer questions with greater accuracy. New strategies can be developed and implemented enabling the educational institutions to transform a wealth of information into a wealth of predictability, stability and profits.

VII References

1) Shyamala, K and S.P. Rajagopalan, 2006. Data mining model for a better higher educational system. Information Tech. J.,5 :560-564
2) DesJardins, S.L., D.A. Ahlburg and B.P. McCall, 2002. A temporal investigation of factors related to timely degree completion. J. Higher Education, 73:555-581.
3) Murphy, K and F. Welch, 1993. Inequality and relative wages. Ameri. Economic review, 83: 104-109.

4) Murtaugh, P.A., L.D. Burns and J. Schuster, 1999 Predicting the retention of university students. Higher Education, 4: 355-357.
5) Kane, J., 1994. College entry by blacks since 1970. The role of college costs, family background and the returns to education. J. Political Econo., 102: 878-911
6) Manski, C. and D. Wise, 1983. College choice in America, Cambridge, MA: Harvard University Press.
7) Quinlan, J.R., 1983. Induction of Machine learning Machine learning, 1:81-106.
8) Han, J. and M. Kambar, 2003. Data mining: Concepts and techniques. Morgan Kaufmann Publishers, New Delhi.
9) Quinlan, R. (1993). C4.5 : Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, CA.

A Novel Optimization of Route Discovery in Dynamic Source Routing (DSR) Protocol for MANET

GJCST Computing Classification
F.2.2 ,C.2.1, C.2.5 & C.2.2

V.Ramesh [1] Dr.P.Subbaiah [2] N.Koteswar Rao [3] K.Sangeetha Supriya [4]
[1] Research scholar,Sathyabama University ,Chennai & Assoc.Prof, CMRCET, Hyderabad, AP,India.
v2ramesh634@yahoo.co.in.
[2] Principal, Veerabrahmendra Institute of Tech. & Sci., Badvel, Kadapa, AP,India
subbaiah_nani@sify.com.
[3] Associate Professor, Dept of IT, Narayana Engg. College,Gudur,AP,India, rao0007@gmail.com.
[4] Assistant Professor, Dept of IT, BVRIT, Medak, AP, India, sangeethasupriyak@gmail.com.

***Abstract*- Ad hoc networks are useful for providing communication support where no fixed infrastructure exists or the deployment of a fixed infrastructure is not economically profitable, and movement of communicating parties is allowed. Therefore, such networks are designed to operate in widely varying environments, from military networks to low-power sensor networks and other embedded systems. Frequent topology changes caused by node mobility make routing in ad hoc wireless networks a challenging problem. A novel technique for optimizing the efficiency of route discovery is proposed. The optimization aims to minimize the number of cached route replies (RREP), which is a significant source of overhead for the dynamic source routing (DSR) protocol. Performance results show that the overall route discovery overhead can be reduced by more than 30% under high node mobility.**

Keywords- DSR, Optimization, Route Discovery, Route Maintenance, MANET.

I INTRODUCTION

There are currently two variations of mobile wireless networks. The first is known as infrastructure network. The bridges for these networks are known as base stations. A mobile unit within these networks connects to and communicates with, the nearest base station that is within its communication radius. As the mobile unit travels out of range of one base station into the range of another, a "handoff" occurs from the old base station to the new, allowing the mobile to be able to continue communication seamlessly throughout the network. Typical applications of this type of network include office wireless local area networks (WLANs). The second type of mobile wireless network is the mobile ad-hoc network or MANET. Unlike infrastructure network, this type of network needs no base station. Mobile nodes communicate to each other by either directly or through intermediate nodes. Ad-hoc network becomes popular since it can be applied in many situations, such as emergency search-and-rescue operations, classroom, meetings or conference and many more. To facilitate communication within the network, a routing protocols used to discover routes between nodes. Building a MANET routing protocol is not an easy job, since efficiency and correctness becomes the main concern. Some approach had been proposed to make routing protocol becomes efficient and correct. Routing protocols in MANET, generally, can be categorized as table-driven and on-demand. In table-driven (also called proactive protocol), like in most routing protocol for wired network, each node is required to maintain routing table keep updated whether there is or not a request for routes. In on-demand (also called as reactive protocol), each node seeks for routes only when there is need to do so. [1][2]

II OVERVIEW AND RELATED WORK

The Dynamic Source Routing (DSR) protocol is a simple and efficient routing protocol designed specifically for use in multi-hop wireless ad hoc networks of mobile nodes. The DSR protocol allows source nodes to dynamically discover a route to any destination node in the ad hoc network. Each data packet sent has in its header the complete ordered list of nodes through which the packet must pass, and avoiding the need for up-to-date routing information in the intermediate nodes through which the packet is forwarded. DSR cache the routing information for future use. DSR protocol contains two major phases, route discovery and route maintenance.

A. Route Discovery

Route Discovery is the process by which a source node S needs to send a packet to a destination node D and hence obtains a route to D. Route Discovery is used only when source node S needs to send a packet to destination node D, it looks up its route cache to locate an unexpired route to the destination and if it fails, then it initiates the route discovery process through broadcasting a Route Request (RREQ) packet. Each node on receiving a RREQ packet, it rebroadcast the packet to its neighbors if it has not forwarded already. Route Request packet (RREQ) contains <Destination Address, Source Address, route Record, Request ID>. On receiving the RREQ packet the destination replies the RREQ packet the destination replies to the source with a Route Reply (RREP) packet. When an intermediate

node detects that the link to the next-hop node towards the destination is broken, it immediately remove this link from the route cache and returns a route error message to the source node. The source node again activates a new route discovery. DSR works for small to medium size MANET when nodes speed is moderate and every node has enough battery power. Its main feature is that every data packet follows the source route stored in its header. This route gives the address of each node through which the packet should be forwarded in order to reach its final destination. Each node on the path has a routing role and must transmit the packet to the next hop identified in the source route.[3]

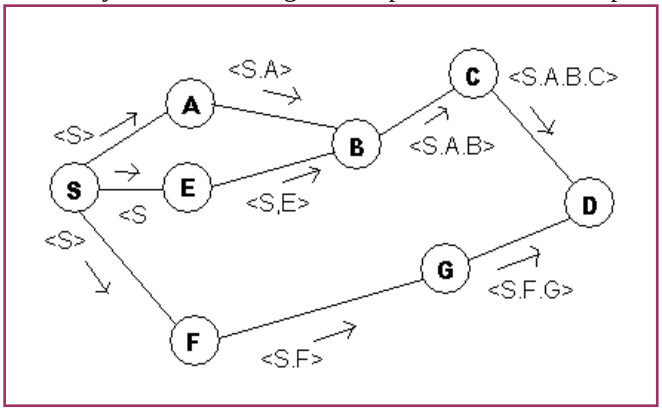


Fig: 1 Example Network

When a source node S want to send message to the destination node D, it initiates route discovery by broadcasting the RREQ packet to its neighbors (A, E, F) as shown in Fig 1. The intermediate nodes (A, E, F) on receive the RREQ packet rebroadcast the packet to its neighbors by appending its id in the route record of the RREQ packet.

Similarly other intermediate nodes also forward the RREQ packet to the destination. When the destination node D receives two or more RREQ packets from the same source through different routes, it finds the best route based on the no of hopes.

The destination node D sends Route Reply (RREP) packet using the route (<S, F, G>)

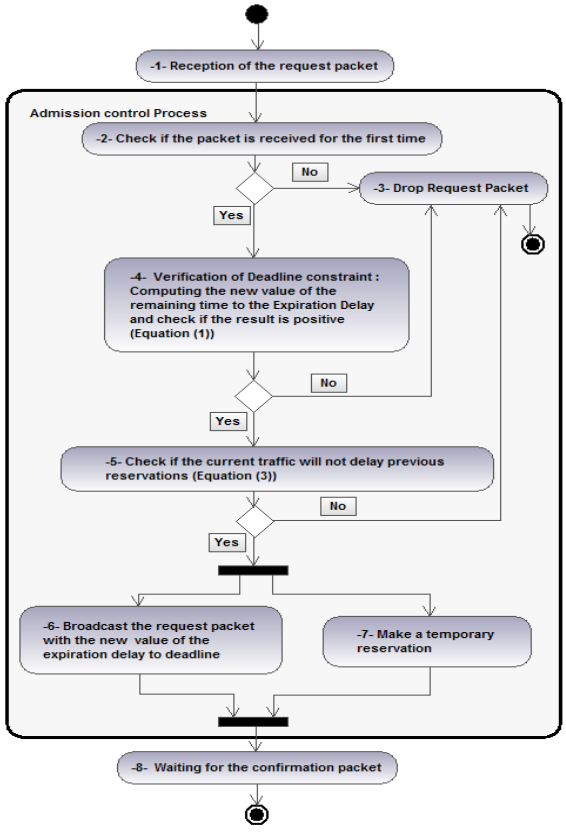


Fig 2: Activity diagram of intermediate node behavior in real time route request phase

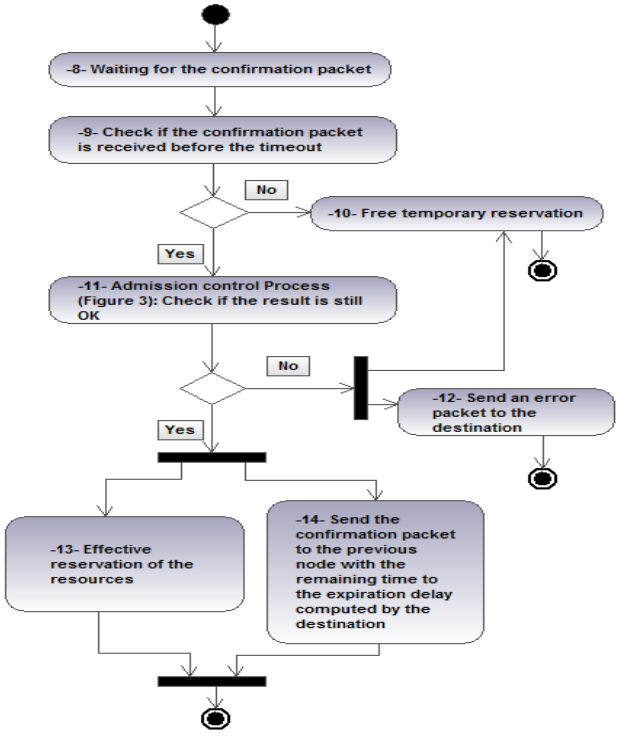


Fig 3: Activity diagram of intermediate node behavior in real time route reply phase

B. Route Maintenance

Each node maintains a Route cache in which it stores every source route it has learned. When a node needs to send a data packet, it checks first its route cache for a source route to the destination. If no route is found, it attempts to find new route using the route discovery mechanism and hereby increases the control overhead and connection setup delay. [4][1]

III Optimization

This phenomenon can be put to good use to minimize the number of cached RREP. For example, the overheard RREP contains information about i) the hop length of the returned source route, and ii) the ID of the RREQ, for which this RREP is generated. Together with the source address of the RREQ, which is found from either the returned source route, or destination address of the RREP, these three pieces of information could be used to decide if a node should reply upon receiving a RREQ, even for the first time.

Existing DSR algorithm dictates that a node will reply if it receives a RREQ for the first time, for which it has a route to the destination. It is almost certain that node D in Fig. 1 would send a cached RREP, since it knows at least a route to the destination, which is that it overhears from the RREP. However, the returned route may not be useful if it is longer than the one previously returned, since route selection at the source is typically based on the shortest path.

We thus propose that if a node overhears a RREP for a RREQ it has not seen before (known by the RREQ ID and source address), the node shall record the three pieces of information from the RREP, namely the i) hop-length of returned source route, ii) RREQ ID, and iii) RREQ source address, as mentioned before. Subsequently, if the node receives this RREQ, it will compare the hop-length of its route (to be returned) with that seen previously. It will reply if it has a shorter route, and discard otherwise.

DSR has a scheme with some similarity for preventing "Route Reply Storms". However, the scheme does not propose the use of other information received from the RREP as we mentioned above. Furthermore, the scheme listens for shorter routes only after RREQ is received. This inherently introduces a delay, which adds to the route acquisition latency.

IV Simulation Environment

We evaluate the performance of our proposed technique using an ns simulator with Monarch wireless extensions. A total of 100 nodes are simulated for 500s over a network space of 1342m x 1342m. The network traffic is modeled as 40 CBR sources with data sent in 64-byte packets at 2 packets/s. Five movement patterns are generated based on random waypoint model for each value of pause time: 0, 100, 200, 300, 400, 500s. A pause time of 0s corresponds to continuous motion (at speed of up to 20m/s) and a pause time of 500s (length of simulation) corresponds to no motion.

V Performance Results

Fig. 4 shows the number of RREPs transmitted by both DSR, and a modified version of DSR with our optimization. The DSR herein operates with all of its existing optimizations, thus providing a more challenging base for comparison than with a non-optimized, simple flooding-based DSR. The RREPs are further segregated into Cached RREPs and Target RREPs, the former being transmitted by non-destination nodes, the latter by destination nodes. Due to source routing and aggressive caching, DSR has an inherent high hit ratio for its route caches, which could explain why there are much more Cached RREPs than Target RREPs as shown in the figure 4.

The results show that the proposed optimization reduces the number of cached RREP significantly, in particular at higher node mobility (lower pause time). At pause time of 0s (highest mobility) where all nodes are in continuous motion, the number of Cached RREPs is fewer by slightly more than 50%. In addition, expectedly, this margin of improvement decreases with mobility, since lower speed lead to fewer route discoveries to be performed. At pause time of 500s (lowest mobility) where all nodes are stationary, no significant difference in the number of Cached RREPs is observed. In addition, since our optimization is aimed at Cached RREPs, the number of Target RREPs remains relatively unchanged. The same can be said for the RREQs. Fig. 5 shows the overall route discovery overhead, comprising of both RREQs and RREPs. By taking RREQ into account (not shown in the figure for clarity but should

be easily extracted), Modified DSR achieves an overall overhead reduction of 33.8% under highest mobility (zero pause time).

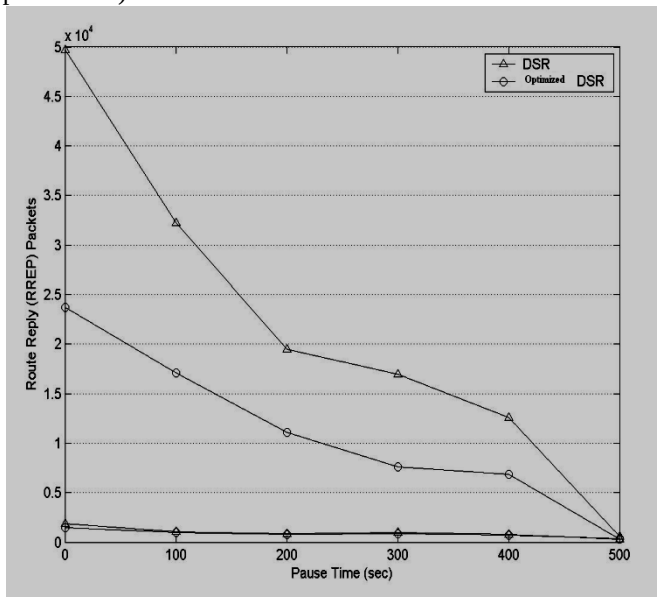


Fig: 4 RREP packets Vs Pause Time

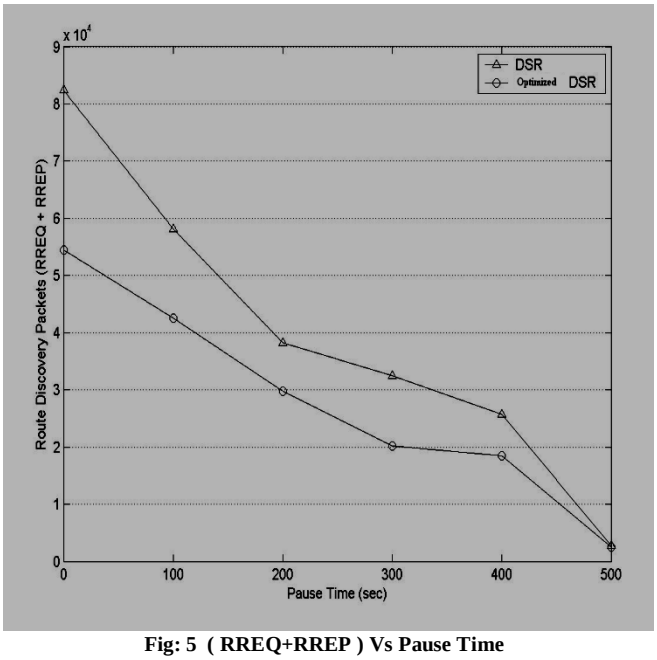


Fig: 5 (RREQ+RREP) Vs Pause Time

VI Conclusion

We have presented a new optimization for DSR to minimize the number of cached RREPs in a route discovery. The proposed optimization is simple in concept and implementation, and is shown to be effective in decreasing the route discovery overhead, in particular under high node mobility.

VII References

1) Siva Ram Murthy, B.S, Manoj, "Routing Protocols for Ad Hoc wireless Networks," in Ad Hoc wireless networks: Architectures and Protocols, Chapter 7. Pearson Publication.
2) Hongbo Zhou, "A Survery on Routing Protocols in MANETs," Technical. Note March 2003.
3) Rashida Hashim, Qassim Nasir, Saad Harous," Congestion Aware Multi-path Dynamic Source Routing Protocol (CAWMP-DSR) for Mobile Ad-Hoc Network" in Proceedings of MoMM2007, pp.199-205.
4) S.Basagni, I. Chlamtac, V. R. Syrotiuk, and B. A. Woodward, "A distance routing effect algorithm for mobility (dream)," in Proceedings of the IEEE/ACM international Conference on Mobile Computing and Networking (MOBICOM'98), 1998, pp. 76–84.
5) Staub.T., (2004). Ad-hoc and Hybrid Networks: Performance Comparison of MANET Routing Protocols in Ad-hoc and Hybrid Networks. Institute of Computer Science and Applied Mathematics, University of Berne, Switzerland, pp.1-38.
6) Tutorial for the network simulator "ns". *http://www.isi.edu/nsnam/ns/tutorial/*

Online Customer Value Identification Based On Site Usage Time through Data Mining Analysis

[1]R. Deepa, [2] R. Hamsaveni
[1]M.Phil
[1] rdeepa_25@yahoo.co.in.com,
[2] amsavenisasikumar@rediffmail.com

GJCST Computing Classification
H.2.8 & J.1

***Abstract*- CRM is required to evaluate the customer performance, discover the trends or patterns in customer behavior, and understand the factual value of their customers to their company. Data mining is playing an important role in analyzing and utilizing the abundant scale of information gathered from customers. This paper is on identifying online customer value based on the time duration of his stay in the online shop, through Data Mining Analysis. In this paper, the sequence of online site time usage can be modeled by using a Hidden Markov Model (HMM) and show how it can be used for the detection of a potential buyer and hence use appropriate promotion campaigns. The clustering technique is applied to identify the stay time category of the customer when he ends up buying a product. An HMM is initially trained with the normal web site stay of a site visitor. If a customer stay time reaches the time as calculated by the trained HMM with sufficiently high probability, he is considered to be a potential buyer. A detailed experimental result to show the effectiveness of our approach is presented.**

Keywords-HMM,K-Means,Model arameter Estimation

I INTRODUCTION

The popularity of online shopping is growing with every passing day. According to the ACNielsen study conducted in 2005, one-tenth of the world's population is shopping online. Customers visit online shops frequently with the intention of buying or making the decision process for buying. Companies usually aim at their produces, products and then give all the customers same sales promotion. This kind of sales promotions neglects the differences among customers. In most cases, these promotions cost a lot, but only get few real profits from customers. That means many promotions are waste. Data mining can solve many typical commercial problems, such as Database Marketing, Customer Segmentation and Classification, Profile Analysis, Cross-selling, Churn Analysis, Credit Scoring, Fraud Detection, and so on. Since those data mining technologies appeared, companies have changed their sales target from products to customers. How to classify customers? How to find out the common character of customers from database? How to dig up the potential customers? How to find out the most valuable customers? These kinds of questions become the most popular data mining applications in marketing.

Nevertheless, the recent customer relation analyses techniques do not consider the time duration factor of the customer's stay on the site and hence has drawbacks. The most important one is that based on those analyses company usually consider the customer visit and stay time on the site as an isolated object and having value only when he/she buys products deals with this company. The duration of the customers stay is neglected in analyzing the value from potential purchase probability. This paper reveals about the application of data mining in online customer value identification, and to use new visual angle to improve these drawbacks.

II RELATED WORK ON ONLINE CUSTOMER VALUE ANALYSIS

The web Usage analysis includes straightforward statistics, such as page access frequency, as well as more sophisticated forms of analysis, such as finding the common traversal paths through a Web site. Usage information can be used to restructure a Web site in order to better serve the needs of users of a site. Long convoluted traversal paths or low usage of a page with important site information could suggest that the site links and information are not laid out in an intuitive manner. The design of a physical data layout or caching scheme for a distributed or parallel Web server can be enhanced by knowledge of how users typically navigate through the site. Usage information can also be used to directly aide site navigation by providing a list of "popular" destinations from a particular Web page.

Web Usage Mining is the application of data mining techniques to large Web data repositories in order to produce results that can be used in the design tasks mentioned above. Some of the data mining algorithms that are commonly used in Web Usage Mining are association rule generation, sequential pattern generation, and clustering. Association Rule mining techniques [1] discover unordered correlations between items found in a database of transactions. In the context of Web Usage Mining a transaction is a group of Web page accesses, with an item being a single page access. Examples of association rules found from an IBM analysis oft he server log oft he Official 1996 Olympics Web site [7] are: – 45% oft he visitors who accessed a page about Indoor Volleyball also accessed a page on Handball.

- 59.7% oft he visitors who accessed pages about Badminton and diving accessed a page about Table Tennis.

The percentages reported in the examples above are referred to as *confidence*. Confidence is the number of transactions containing all of the items in a rule, divided by the number of transactions containing the rule antecedents (The

antecedents are Indoor Volleyball for the first example and Badminton and Diving for the second example).

The problem of discovering *sequential patterns* [17, 26] is that of finding inter transaction patterns such that the presence of a set of items is followed by another item in the time-stamp ordered transaction set. By analyzing this information, a Web Usage Mining system can determine temporal relationships among data items such as the following Olympics Web site examples:

- 9.81% oft he site visitors accessed the Atlanta home page followed by the Sneak peek main page.
- 0.42% oft he site visitors accessed the Sports main page followed by the Schedules main page.

The percentages in the second set of examples are referred to as support. Support is the percent of the transactions that contain a given pattern. Both confidence and support are commonly used as thresholds in order to limit the number of rules discovered and reported. For instance, with a 1% support threshold, the second sequential pattern example would not be reported.

A. *Clustering Analysis*

Clustering analysis allows one to group together users or data items that have similar characteristics. Clustering of user information or data from Web server logs can facilitate the development and execution of future marketing strategies, both online and off-line, such as automated return mail to visitors falling within a certain cluster, or dynamically changing a particular site for a visitor on a return visit, based on past classification of that visitor. As the examples above show, mining for knowledge from Web log data has the potential of revealing information of great value. While this certainly is an application of existing data mining algorithms, e.g. discovery of association rules or sequential patterns, the overall task is not one of simply adapting existing algorithms to new data. Ideally, the input for the Web Usage Mining process is a file, referred to as a user session file in this paper that gives an exact accounting of who accessed the Web site, what pages were requested and in what order, and how long each page was viewed. A user session is considered all of the page accesses that occur during a single visit to a Web site. The information contained in a raw Web server log does not reliably represent a user session file for a number of reasons that will be discussed in this paper. Specifically, there are a number of difficulties involved in cleaning the raw server logs to eliminate outliers and irrelevant items, reliably identifying unique users and user sessions within a server log, and identifying semantically meaningful transactions within a user session.

There are several data preparation techniques and algorithms that can be used in order to convert raw Web server logs into user session files in order to perform Web Usage Mining. The specific contributions include (i) development of models to encode both the Web site developer's and users' view of how a Web site should be used, (ii) discussion of heuristics that can be used to identify Web site users, user sessions, and page accesses that are missing from a Web server log, (iii) definition of several transaction identification approaches, and (iv) and evaluation oft he different transaction identification approaches using synthetic server log data with known association rules.

B. *Data Mining Process*

There are three main steps of data mining process.

i. Data Preparation

In the whole data mining process, data preparation is somehow a significant process. Some book says that if data mining is considered as a process then? Data preparation is at the heart of this process. However, nowadays databases are highly susceptible to noise, missing and inconsistent data. So preprocessing data improve the efficiency and ease of the data mining process, this becomes an important problem. Several consulting firms, such as IBM, have approved that data preparation costs 50% - 80% resource of the whole data mining process. From this view, there is a need to pay attention for data preparation. There are three data preprocessing techniques should be considered in data mining:

- Data cleaning-**Inconsistent Data**- Not all the data taken is "clean". For example, a list of Nationality may have the values of "China", "P.R.China", and "Mainland China". These values refer to the same country, but are not known by the computer. Therefore, this is a consistency problem. b) Missing values Data from a company's database often contains missing values. Sometimes the approaches require rows of data to be complete in order to mine them, but the database may contain several attributes with missing values. If too many values are missing in a data set, it becomes hard to gather useful information from this data. c) Noisy data Noise is a random error or variance in a measured variable.
- Data integration usually the data analysis task will involve data integration. It combines data from multiplying sources into a coherent data store. Those sources include multiple database or flat files. Several issues should be considered during data integration, such as schema integration, correlation analysis for detecting redundancy, and detection and resolution of data value conflicts. Careful integration of the data can help improve the accuracy and speed of the mining process.
- Data reduction If you select data from a data warehouse, you probably find the data set is huge. Data reduction techniques can be applied to obtain a reduced representation of the data set. Mining on reduced data set should be more efficient yet produce the same analytical results. It includes several strategies, such as data cube aggregation, dimension reduction, data compression, numerosity

reduction, and discretization and concept hierarchy generation

ii. Knowledge Discovery In Database

As a core data mining techniques, knowledge and information discovery has several main components:

- Determine the type of data mining tasks in order to confirm that the functions and tasks to be achieved by recent system belong to which kind of classification or clustering.
- Choose suitable technologies for data mining. So the appropriate data mining technologies based on the tasks have to be confirmed. Such as, classification model often use learning neural network or decision tree to realize; while clustering usually use clustering analysis algorithms to realize; association rules often use association and sequence discovery to realize.
- Select a specific algorithm based on the technologies. Furthermore, a new efficient algorithm can be designed by the specific mining tasks. For choosing the data mining algorithms determine the hidden pattern in selecting the data.
- Mining data use the selected algorithms or algorithms portfolio to do repeated and iterative searching. Extract the hidden and innovative patterns from data set.

iii. Model Explain and Estimate

Explain and estimate the patterns got from data mining, get the useful knowledge. For instance, remove some irrespective and redundant patterns, after filtration the information should be presented to customers; Use visualization technology to express the meaningful model, in order to translate it into understandable language for users. A good application of data mining can change primal data to more compact and easily understand form and this form can be defined definitely. It also includes solving the potential conflict between mining results and previous knowledge, and using statistical methods to evaluate the current model, in order to decide whether it is necessary to repeat the previous work to get the best and suitable model. The information achieved by data mining can be used later to explain current or historical phenomenon, predict the future, and help decision-makers make policy from the existed facts.

III HMM BACKGROUND

An HMM is a double embedded stochastic process with two hierarchy levels. It can be used to model much more complicated stochastic processes as compared to a traditional Markov model. An HMM has a finite set of states governed by a set of transition probabilities. In a particular state, an outcome or observation can be generated according to an associated probability distribution. It is only the outcome and not the state that is visible to an external observer [18]. HMM-based applications are common in various areas such as speech recognition, bioinformatics, and genomics. Lane [23] has used HMM to model human behavior. Once human behavior is correctly modeled, it can be used to predict the future behaviour and hence help in managing future actions to the predicted behaviours.

An HMM can be characterized by the following [18]:

i. N is the number of states in the model and the set of states $S = \{S_1, S_2, \ldots . S_N\}$ where $S_i, i = 1,2, \ldots N$ is an individual state. The state at time instant t is denoted by q_t.

ii. M is the number of distinct observation symbols per state. The observation symbols correspond to the physical output of the system being modeled. Let the set of symbols be $V = \{V_1, V_2, \ldots . V_M\}$, where $V_i, i = 1,2, \ldots M$ is an individual symbol.

iii. The state transition probability matrix $A = [a_{ij}]$, where $a_{ij} = P(q_{t+1} = S_j | q_t = S_i), 1 \leq i \leq N; 1 \leq j \leq N; t = 1, 2, \ldots$ For the general case where any state j can be reached from any other state i in a single step, then $a_{ij} > 0$ for all i, j. Also, $\sum_{j=1}^{N} a_{ij} = 1, 1 \leq i \leq N$.

iv. The observation symbol probability matrix $B = [b_j(k)]$, where $b_j(k) = P(V_k | S_j), 1 \leq j \leq N, 1 \leq k \leq M$ and $\sum_{k=1}^{M} \mathrm{b_j}(\mathrm{k}) = 1, 1 \leq \mathrm{j} \leq \mathrm{N}$.

v. The initial state probability vector $\pi = [\pi_i]$, where $\pi_i = P(q1 = S_i), 1 \leq i \leq N$, such that $\sum_{i=1}^{N} \pi_\mathrm{i} = 1$

vi. The observation sequence $O = O_1, O_{i2} \ldots O_R$ where each observation O_t is one of the symbols from V, and R is the number of observations in the sequence.

It is evident that a complete specification of an HMM requires the estimation of two model parameters, N and M, and three probability distributions A, B, and π. The notations are used to indicate the complete set of parameters of the model, where A, B implicitly include N and M.

A. Use Of HMM For Identifying Online Customer Value

Ideally, the input for the Web Usage Mining process is a file, referred to as a user session file in this paper that gives an exact accounting of who accessed the Web site, what pages were requested and in what order, and how long each page was viewed. A user session is considered to be all of the page accesses that occur during a single visit to a Web site. The information contained in a raw Web server log does not reliably represent a user session file for a number of reasons. Specifically, there are a number of difficulties involved in cleaning the raw server logs to eliminate outliers and irrelevant items, reliably identifying unique users and user sessions within a server log, and identifying semantically meaningful transactions within a user session.
The analyzing program runs at the application layer or the business layer of the web application. Each web customer

time is submitted to the application on server. The application receives the current session duration details. It tries to find the user's normal site stay time that results in the purchase of a product based on the stay time profile calculated by clustering. If the customer stay time has approached the profile time, he can be identified as a more valuable customer and hence provided with appropriate campaigns. In this section, how HMM can be used for online customer value identification is explained.

B. HMM Model For Identifying Online Customer Value

To map the customer site stay or usage time in terms of an HMM, first decide the observation symbols in the model. Then quantize the stay time values x into M stay ranges V1; V2; . . . VM, forming the observation symbols at the business logic layer. The actual stay duration for each symbol is configurable based on the stay habit of individual customers. These duration ranges can be determined dynamically by applying a clustering algorithm on the values of each customer's site stay time, as shown in Section III.C.

In this work, it is considered only three time ranges, namely, low l, medium m and high h. Our set of observation symbols is, therefore, V = {l, m, h} making M = 3. For example, let l = (0,15 mins), m = (30 mins, 45 mins) and h =(1 hour,site logout time). If customer stays on site for 30 mins, then the corresponding observation symbol is m. A customer stays on the site for different amounts of time over a period. One possibility is to consider the sequence of duration of stay and look for similarities in them. However, the sequence of types of stay time is more stable compared to the sequence of exact stay times. The reason is that, a customer stays online depending on his need for procuring different types of items over a period. This, in turn, generates a sequence of site stay times.

Each individual stay time usually depends on the corresponding type of site stay. Hence, the transition in the type of stay time is considered to state the transition in our model. The type of each stay is linked to the decision making time for products from the corresponding merchant. This information about the customer's decision-making time is not known to the online shop application. Thus, the intention of the customer is hidden from the application. The set of all possible types of decisions forms the set of hidden states of the HMM.

One possibility is to consider the sequence of duration of stay and look for similarities in them. However, the sequence of types of stay time is more stable compared to the sequence of exact stay times. The reason is that, a customer stays online depending on his need for procuring different types of items over a period of time. This, in turn, generates a sequence of site stay times. Each individual stay time usually depends on the corresponding type of site stay. Hence, the transition in the type of stay time is considered as state transition in the model. The type of each stay is linked to the decision making time for products from the corresponding merchant. This information about the customer's decision-making time is not known to the online shop application. Thus, the intention of the customer is hidden from the application. The set of all possible types of decisions forms the set of hidden states of the HMM.

After deciding the state and symbol representations, the next step is to determine the probability matrices A, B, and _ so that representation of the HMM is complete. These three model parameters are determined in a training phase using the Baum-Welch algorithm[18] . The initial choice of parameters affects the performance of this algorithm and, hence, they should be chosen carefully.

One possibility is to consider the sequence of duration of stay and look for similarities in them. However, the sequence of types of stay time is more stable compared to the sequence of exact stay times. The reason is that, a customer stays online depending on his need for procuring different types of items over a period this, in turn, generates a sequence of site stay times. Each individual stay time usually depends on the corresponding type of site stay. The type of each stay is linked to the decision making time for products from the corresponding merchant. This information about the customer's decision-making time is not known to the online shop application. Thus, the intention of the customer is hidden from the application. The set of all possible types of decisions forms the set of hidden states of the HMM.

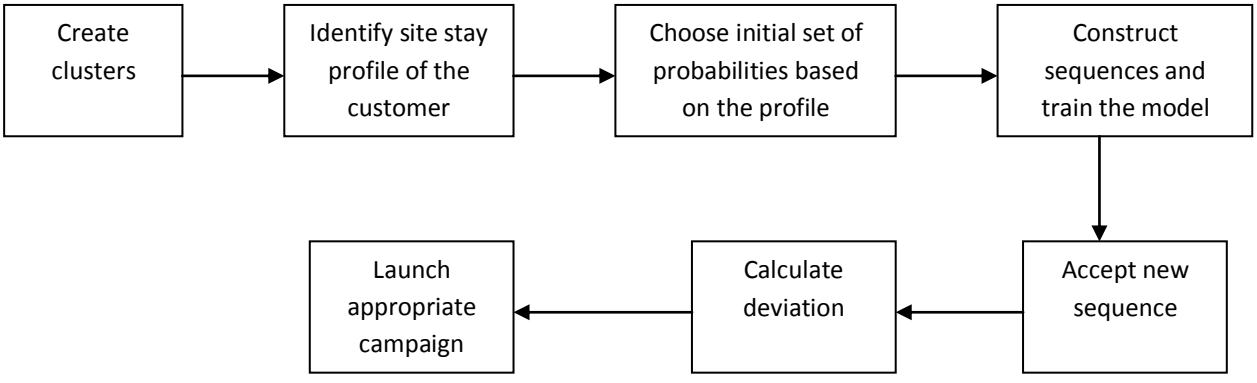


Fig 1. Flow of Process

Thus, the intention of the customer is hidden from the application. The set of all possible types of decisions forms the set of hidden states of the HMM.
The special case of fully connected HMM is considered, in which every state of the model can be reached in a single step from every other state. Site stay profiles of the individual customers are used to obtain an initial estimate for probability matrix B of (2).

C. *Dynamic Generation Of Observation Symbols*

Data Mining in general is an incremental process which includes the following phases as described by the CRISP-DM Process Model (Cross-Industry Standard Process for Data Mining):

- **Business Understanding**- defining objectives and requirements
- **Data Understanding**- data collecting, checking quality and consistency
- **Data Preparation**- preparing the data for the data mining algorithms / tools (includes selection, normalization etc.)
- **Modelling**-applying various data mining techniques to the data; stepping back to 3. may be necessary due to requirements of some algorithms
- **Evaluation**- evaluation of the model to be certain that it meets the requirements; if not, stepping back to 1. may be necessary
- **Deployment**-Clustering algorithms calculate a partitioning of a dataset into subsets (clusters) in a way that instances within a subset are more similar to each other than to instances within another subset. This is usually done using a distance measure D(r1,r2), which is specific for the data to be compared. Typical distance measures are the Euclidian and the Manhattan distance.

For each customer, we train and maintain an HMM. To find the observation symbols corresponding to individual customer's transactions dynamically. Finally a clustering algorithm is executed on his past transactions. Normally, the duration of site stay details are stored in the web session log contain many attributes. For our work, r only the time duration that the customer spent in his site visit. Although various clustering techniques could be used, we use K-means clustering algorithm [24] to determine the clusters.

A. *K-Means*

The K-Means algorithm is a simple, well known, and widely used clustering algorithm. The algorithm is based on the idea that objects (input vectors, records) are grouped into clusters according to a distance function, for example the Euclidian distance. The resulting clusters contain objects with a minimum within-cluster distance. The algorithm is performed as follows:

- A number of so called centroids are randomly spread among the input range. For each cluster one centroid will be calculated.
- Each object is assigned to the cluster represented by the closest centroid according to the distance function.
- The new position of the centroid is found by calculating the centre of all objects that are assigned to that cluster. This will cause the centroids to 'move around'.
- Point 2. und 3. are repeated until the centroids do not change any more.

K-means is an unsupervised learning algorithm for grouping a given set of data based on the similarity in their attribute (often called feature) values. Each group formed in the process is called a cluster. The number of clusters K is fixed a priori. The grouping is performed by minimizing the sum of squares of distances between each data point and the centroid of the cluster to which it belongs.
In our work, K is the same as the number of observation symbols M. Let c1; c2; . . . cM be the centroids of the generated clusters. These centroids or mean values are used to decide the observation symbols when a new entry comes in. Let x be the amount spent by the customer u in visit times T. As mentioned before, the number of symbols is 3 in our system. Considering M ¼ 3, and executing K-means algorithm on the example transactions in Table 2, the clusters are obtained, as shown in Table 3, with cl, cm, and ch as the respective centroids. It may be noted that the duration values 5, 10, and 10 have been clustered together as cl resulting in a centroid of 8.3. The percentage ðpÞ of total number of transactions in this cluster is thus 30 percent. Similarly, duration time values 15, 15, 20, 25, and 25 have been grouped in the cluster cm with centroid 20, whereas time values 40 and 80 have been grouped together in cluster ch. cm and ch, thus, contain 50 percent and 20 percent of the total number of transactions.
When the application layer receives a new entry T for this customer, it measures the distance of the stay time x with respect to the means cl, cm, and ch to decide the cluster to which T belongs and, hence, the corresponding observation symbol. As an example, if x =10mins, then in Table 2 using (9), the observation symbol is V1= l.

Table 1 Example stay times in Each visit

Visit No	1	2	3	4	5	6	7	8	9	10
Duration of stay	40	25	15	5	10	20	15	20	10	80

B. *Site Stay Profile Of Customers*

The site stay profile of a customer suggests his normal site stay behavior. Customers can be broadly categorized into three groups based on their browsing duration habits, namely, long-staying (hs) group, medium-staying (ms) group, and low-staying (ls) group. Custmers who belong to the hs group, normally take more time to make their decisions about purchasing an item. Similar definition applies to the other two categories also. Site stay profiles of customers are determined at the end of the clustering step.

Let pi be the percentage of total number of visits of the customer that belong to cluster with mean ci.
Thus, site stay profile denotes the cluster number to which most of the site visits of the customer belong. In the example in Table 2, the site stay profile of the customer is 2, that is m and, hence, the customer belongs to the ms group.

Table 2 Output of K-Means Clustering Algorithm

Cluster mean/ centroid name	cl	cm	ch
Observation symbol	V1=l	V2=m	V3=h
Mean values (Centroid)	8.3	20	60
Percentage of total transactions (p)	30	50	20

C. Model Parameter Estimation and Training

To estimate the HMM parameters for each customer Baum-Welch algorithm is used. The algorithm starts with an initial estimate of HMM parameters A, B, and _ and converges to the nearest local maximum of the likelihood function. Initial state probability distribution is considered to be uniform, that is, if there are N states, then the initial probability of each state is 1=N. Initial guess of transition and observation probability distributions can also be considered to be uniform. However, to make the initial guess of observation symbol probabilities more accurate, site stay profile of the customer, as determined in Section 3.4, is taken into account. So three sets of initial probability is made for observation symbol generation and also for three site stay groups—ls, ms, and hs. Based on the customer's site stay profile, the corresponding set of initial observation probabilities is chosen. The initial estimate of symbol generation probabilities using this method leads to accurate learning of the model. Since there is no a priori knowledge about the state transition probabilities, the initial guesses are considered to be uniform.

From now start training the HMM. The training algorithm has the following steps:

- initialization of HMM parameters,
- forward procedure, and
- backward procedure.

Details of these steps can be found in [18]. For training the HMM, the customer's site stay duration is converted into observation symbols and form sequences out of them. At the end of the training phase, an HMM corresponding to each customer is obtained. Since this step is done offline, it does not affect the application performance, which needs online response.

D. Launch The Appropriate Campaign

After the HMM parameters are learned, the symbols from a customer's training data are taken and form an initial sequence of symbols. Let O1;O2; . . .OR be one such sequence of length R. This recorded sequence is formed from the customer's transactions up to time t. We input this sequence to the HMM and compute the probability of acceptance by the HMM. If the deviation is not there, the customer is most probably going to make a purchase now and hence an appropriate campaign to close the sale can be launched.

IV Results

Predicting a potential buyer using real data set is a difficult task. Online shops do not, in general, agree to share their data with researchers. There is also no benchmark data set available for experimentation. Therefore, the large-scale simulation studies are performed to test the efficacy of the system. A simulator is used to generate a mix of visit times. The customers are classified into three categories as mentioned before—the low, medium, and hs groups. The effects of group and the percentage of visits that belong to the low, medium, and high-range clusters. First carry out a set of experiments to determine the correct combination of HMM design parameters, namely, the number of states, the sequence length, and the threshold value. Once these parameters were determined, the process was tested to appropriately predict a customers stay time and classify the customer to identify his buying decision and hence his value.

V Conclusion

In this paper, we have proposed an application of HMM in online customer value identification. The different steps in customer site stay and decision-making process to buy a product are represented as the underlying stochastic process of an HMM. The ranges of site stay times as the observation symbols are used, whereas the types of stay have been considered to be states of the HMM. We have suggested a method for finding the site stay profile of customers, as well as application of this knowledge in deciding the value of observation symbols and initial estimate of the model parameters. It has also been explained how the HMM can identify a potential buyer and predict a customer session duration. Experimental results show the performance and effectiveness of our system and demonstrate the usefulness of learning the session duration profile of the customers. Comparative studies reveal that the Accuracy of the system is close to 80 percent over a wide variation in the input data. The system is also scalable for handling large volumes of transactions.

VI Reference

1) "Global Consumer Attitude Towards On-Line Shopping," http://www2.acnielsen.com/reports/documents/2005_cc_online shopping.pdf, Mar. 2007.
2) D.J. Hand, G. Blunt, M.G. Kelly, and N.M. Adams, "Data Mining for Fun and Profit," Statistical Science, vol. 15, no. 2, pp. 111-131, 2000.

3) Charles X. Ling and Chenhui Li : Data Mining for Direct Marketing Problems and Solutions , KDD-98
4) Peter Van Der Putten, Data Mining In Direct Marketing Databases, World Scientific, October 15, 1998
5) Data Mining in the Insurance Industry,Solving Business Problems using SAS® Enterprise Miner™ Software
6) Shailendra K, 2005, "Understanding the Relationship Between loyalty and Churn", BI Reports, December 2005
7) Berry, M.J.A., Linoff, G.S.: Mastering Data Mining. Wiley, New York (2000)
8) Adomavicius, G. Tuzhilin, A., (2001), Using data mining methods to build customer profiles. IEEE computer,Volume: 34, Issue: 2, pp.74-82
9) Reichardt, Ch., One-to-One-Marketing im Internet. Wiesbaden: Gabler verlag, 2000.
10) Greenberg, P., CRM – Customer Relationship Management at the Speed of Light. 2nd ed., Berkeley: McGraw-Hill/Osborne, 2002.
11) Feelders, A.J.; Daniels, H.A.M.; Holsheimer, M., (2000), Methodological and practical aspects of datamining, Information & Management, vol.37 , pp.271-281
12) Liu,H.;Hussain, F.;Tan, C.L&Manornjan Dash, (2002), Discretization: An Enabling Technique, Data Mining and Knowledge Discovery, Volume 6, Number 4, pp.393–423
13) Coltman, T., "Why build a customer relationship management capability?", in Journal of Strategic Information Systems, vol. 16, 2007, pp. 301-320
14) He, Z., Xu, X., Huang, J. Z., Deng, S., "Mining class outliers: concepts, algorithms and applications in CRM", in Expert Systems with Applications, vol. 27, 2004, pp. 681-697.
15) Böttcher, M., Nauck, D., Borgelt, C., & Kruse, R., (2006) , A framework for discovering interesting business changes from data, BT Technology Journal , volume 24, issue 2,pp. 219 ¬228

An Experimental Study of Edge Detection Methods in Digital Image

Anu Suneja[1]
E-mail: anusuneja3@gmail.com
Gaurav Kumar[2]
E-mail: kumargaurav.in@gmail.com

GJCST Computing Classification
I.4.6 & I.5.m

***Abstract*- Image segmentation is the major step for analysis of image. All image segmentation methods depend upon either on intensity values of pixel or on discontinuity and similarity in pixel's intensity value. In first category, image is segmented based on abrupt change in intensity value. Such abrupt change in intensity value represents an edge in the image. A number of edge detection methods have been developed in area of image processing. Each of which is based on its own method to detect edge points. Edge detection methods are based on discontinuity of image intensity function. Such discontinuity can be detected either using first order derivative or using second order derivative of image intensity function. Different edge detection methods results in identification of different types of edges. Like a few edge detection methods mark edges as thick lines, whereas, some other methods result in detection of edges as thin lines. Some methods use first order derivative of intensity function to detect edges. Whereas, some other methods detect edges using second order derivative of intensity function. From such a large number of edge detection methods, which method one should use depends upon problem being solved and requirement of user.**

Keywords- Edge Detection, Edge Localization, Image Gradient, Intensity Function

I INTRODUCTION

The primary requirement for image analysis is to detect edges in an image. To identify any particular object in a scene its boundaries are find out which are edges of the image. In most of edge detection methods, the abrupt change in pixel's intensity is used to detect edge points. To find out whether a pixel lies on edge or not, it is compared with its neighbors. We can use either 4-neighbor or 8-neighbor method for detecting edges.

N1	N2	N3
N8	P	N4
N7	N6	N5

Table 1: 8-Neighbors of a pixel

	N1	
N2	P	N3
	N4	

Table 2: 4-Neighbors of a Pixel

[1] Edges are representation of discontinuities of image intensity function. To find discontinuity in a function derivative operators are used. Robert and 4-neighbor operators are based on first order derivative approach of edge detection, whereas, Laplacian method is based on second order derivative approach.

Edge detection process in itself is not a single step process. It consists of further sub steps like noise smoothing, edge localization, edge enhancement and next step is edge detection. In first step, image is smoothed to remove blurring effect. Edge localization marks pixels lying on edge. In edge enhancement process, strong and weak edges are detected. Most of edge detection methods are based on first order and second order derivative of image function. The points lying over edge have maximum intensity value of their first order derivative and zero crossing value of their second order derivative.

II EDGE DETECTION

[2] Edges can be defined as a set of connected pixels that lie on the boundary between two regions. To find such pixels first order derivative or second order derivative of image function is used. Pixels lying on edge have peak value as compare to its neighbors. Using first order derivative it's detected whether pixel lies on ramp or not. If it lies on ramp it is pixel on edge, otherwise not. [2]In Second order derivative method, sign is checked to detect if pixel lies on dark or light side of edge. The zero crossing property of second order derivative is useful for locating the centers of thick edges.

Edges are representations of discontinuity in intensity function. To detect discontinuity in intensity function a mask is run through the image in segments. In a mask of n x n order some weight is assigned to all the pixel positions in n x n region of image. The value of weight assigned depends upon property being used for discontinuity detection. This mask is rotated all over the image, pixel wise and detects if pixel lies over the edge or not. [3]At each pixel position (x, y), the response of mask of order 3 x 3 is calculated using the equation:

$$R=w(-1,-1)f(x-1,y-1)+w(-1,0)f(x-1,y)+\text{--------}w(0,0)f(x,y)+\text{---------}\quad w(1,0)f(x+1,y)+w(1,1)f(x+1,y+1) \qquad \text{.....(1)}$$

In linear filtering of intensity function f of order M x N with mask of size m x n is given by:

$$g(x,y)=\sum\sum w(r,s)\,(f(x+r,y+s) \qquad \text{.....(2)}$$

Using equation 2 edges are detected in different edge detection methods.

A. *Prewitt Method*

It is based on first order derivative of intensity function. [4] In this method, similar weights are assigned to all the neighbors of candidate pixels whose edge strength is being calculated. Mask used by Prewitt operator are

1	1	1
0	0	0
-1	-1	-1

0	1	1
-1	0	1
-1	-1	0

-1	0	1
-1	0	1
-1	0	1

This method can detect vertical edges better. Also the effect of noise can be reduced by Prewitt method.

B. *Sobel Method*

[4, 5] Like Prewitt method it is also based on first order derivative of intensity function. In this method, higher weights are assigned to pixels close to candidate pixels. Mask used by Sobel operator are

1	2	1
0	0	0
-1	2	-1

0	1	2
-1	0	1
-2	-1	0

-1	0	1
-2	0	2
-1	0	1

This method can detect diagonal edges better as compare to Prewitt method. The effect of noise can also be reduced by Sobel method by inherent averaging of neighbor pixels.

C. *Laplacian Method*

This method is used for approximating second order derivative of image function. For a given image Laplacian of image is second order partial derivative along x and y axis which is given by following equations:

Along x-axis derivative is g(r,c)-2r(r,c-1)+g(r,c-2)
And along y-axis derivative is g(r,c)-2g(r-1,c)+g(r-2,c)

Laplacian mask used for 4-neighbors and 8 neighbors are as:

0	1	0
1	-4	1
0	1	0

1	1	1
1	-8	1
1	1	1

III EXPERIMENTAL RESULTS

Image of Figure-1 is used as original image. This image and Laplacian mask is passed in function findedge( ), which returned Figure-2 image after detecting edges. Similarly masks for Sobel and Prewitt methods are passed in findedge() function to return Figure-3 and Figure-4 image.
Following is the Source Code for the implementation of Algorithm

```
findedge(filter[3][3],IB14[][][,IB17[][],m,n)
begin
z=0;
sum=0.0;
for(x=0 to m)
begin
for(y=0 to n)
begin
sum=0.0;
if((x==0)OR(x==m-1))
sum=0.0;
else if((y==0)OR(y==n-1))
sum=0.0;
else
begin
for(i=-1 to 1)
begin
for(j=-1 to 1)
begin
t=IB14[x+i][y+j];
sum=sum+t*filter[i+1][j+1];
end;
```

IV CONCLUSION

From the above results, it has concluded that in Laplacian method thin lines are used to mark edges whereas, in Sobel and Prewitt method edges are marked with thick lines. Thus edges detected in Laplacian method are sharp and clear. At the same time problem with Laplacian method is that some non edge pixels are also get marked in it. Prewitt method detects vertical edges better as compare to Sobel method. Whereas, Sobel method detects diagonal edges better as compare to Prewitt method.

V REFERENCES

1) B.Chanda, D.Dutta Majumder, "Digital Image Processing and Analysis", PHI, pp 239.
2) C.Gonzalez Rafel, "Digital Image Processing", Pearson education, pp 594-598.

3) C.Gonzalez Rafel, "Digital Image Processing", Pearson education, pp 138.
4) "Image Processing and edge detection", Computer Vision, GZ Yang and DF Gillies, http://www.doc,ic.ac.uk/
5) B.Chanda, D.Dutta Majumder, "Digital Image Processing and Analysis", PHI, pp 244.
6) Pratt, W.K. [2001], "Digital Image Processing", 3rd edition, John Wiley & sons.
7) Jain, Anil. K. [1989], "Fundamentals of digital Image Processing", PHI.
8) http://www.owlnet.rice.edu/~elec539/Projects97/segment/edge.html date accessed 12th Jan, 2010.
9) http://www.owlnet.rice.edu/~elec539/Projects97/morphjrks/moredge.html Date Accessed 12th Jan, 2010.
10) http://homepages.inf.ed.ac.uk/rbf/HIPR2/log.htm Date Accessed 16th Jan, 2010.

Statistical Image Watermarking In DWT with Capacity Improvement

M. Devapriya
Email : devapriya_gac@rediffmail.com
Dr. K. Ramar.
Email : ramar_nec@rediffmail.com

GJCST Computing Classification
I.4.9, G.3 & F.2.1

***Abstract*- Abundant techniques has been widely used to design robust image watermarking schemes but in most cases due significance is not given on capacity and data imperceptibility aspects. Robustness of an image-watermarking scheme is the ability to detect the watermark after intentional attacks and normal audio/visual processes. This paper proposes a well-organized blind watermark detection scheme using DWT coefficients. Discrete Wavelet Transform (DWT) is widely applied to image watermarking applications because it decomposes a cover image into spatial domain as well as frequency domain simultaneously. The proposed method improves the capacity of image watermarking. The proposed paper concentrates on some of the main attributes necessary for image watermarking. They are embedding scheme, maximum likelihood detection, decision threshold, and the Laplacian model for image watermarking. The embedding method is multiplicative and done at second level of DWT decomposition by most favorable choice of the embedding strength. The watermark detection is based on the maximum likelihood ratio. Neyman-Pearson criterion is used to reduce the missed detection probability subject to a fixed false alarm probability. The DWT coefficients are assumed to be modeled using the Laplacian distribution. The proposed method is tested for imperceptibility, robustness, and capacity and proved to have better robustness and better imperceptibility and better capacity than other conventional watermarking techniques that were proposed earlier in literature.**

Keywords- Decision Threshold, DWT, Laplacian Distribution, Maximum Likelihood Detection, Neyman-Pearson Criterion, Watermarking.

I INTRODUCTION

Multimedia can be defined to be the combination and integration of more than one media format (e.g., text, graphics, images, animation, audio and video) in a given application. Nowadays, multimedia data is stored in the digital form which makes the processing and storage easy. But this leads to unauthorized duplication of the digital data. Digital watermarking is used to solve the above problem. The following requirements are generally considered for to evaluate a watermark system. They are readability, security, imperceptibility, and robustness [14]. It deals with techniques to embed the copyright information into a digital media by making small changes in the media content.

A digital watermark is a prototype of bits inserted into a digital image, an audio or video file. The name comes from the hardly visible text or graphics embossed on stationery that identifies the manufacturer of the stationery. There are more than a few proposed or actual watermarking applications [12]: broadcast monitoring, owner identification, proof of ownership, transaction tracking, content authentication, copy control, and device control. Specifically, watermarking appears to be useful in plugging the analog hole in consumer electronics devices [13].

Watermarking can be done in either spatial domain or transform domain. Spatial domain approaches like LSB technique are not content based and are simple to implement. Transform domain approaches are more robust and can be implemented adaptively. Among the transform domain, techniques DCT and DWT are commonly used. In [17] [18] [20], for example, the most significant DWT coefficients are selected and modified to carry the watermark. In DWT-based watermarking, the DWT coefficients are modified to embed the watermark data. Because of the conflict between robustness and transparency, the modification at a given level is usually made in HL, LH, and HH sub bands. Additionally, discrete wavelet transform (DWT) based watermarking techniques are gaining more recognition because DWT has a number of advantages over other transform such as progressive and low bit-rate transmission, quality scalability and region-of-interest (ROI) coding stipulate more competent and adaptable image.

The embedding of watermark in the cover image can be done either by additive or multiplicative rule. Usually, for additive embedding, correlation detection is used to detect the watermark. Additive methods are simple and used widely. Non-additive methods are very efficient because of their ability to achieve image dependent embedding and flexibilities in using HVS models. The security of a watermark can be defined to be the ability to prevent hostile attacks such as unauthorized removal, unauthorized embedding, and unauthorized detection. The comparative importance of these properties depends on the requirements of a given application.

For non-additive schemes in DWT domain [1] and [2] suggest Maximum likelihood detection using Bayes Decision theory and Neyman-Pearson criterion for detection. [4] [5] discuss statistical detections in DFT and DCT domain respectively. In paper [1] and [3], third level decomposition is employed and sub bands LH3, HL3 and HH3 are embedded with watermarks. In this paper, we use the level 2 decomposition and embed only in HH2. This improves the payload and imperceptibility. To achieve maximum protection, the watermark should be: 1) undeletable; 2) perceptually invisible; 3) statistically undetectable; 4) resistant to lossy data compression; 5)

resistant to common image processing operations; and 6) unambiguous [15].

The remainder of the paper is structured as follows. Section 2 explains our proposed methodology for image watermarking. Section 3 illustrates the experimental results and discussions and Section 4 concludes the paper with fewer discussions.

II Methodology

Watermarking is done by altering the wavelets coefficients of carefully selected DWT sub-bands. Figure 1 represents the Multi-resolution DWT sub bands of an image.

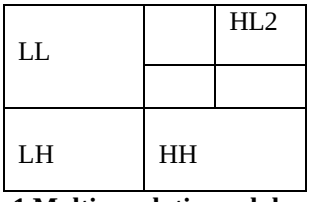


Fig.1 Multi-resolution sub bands

A. *Embedding Scheme*

The proposed scheme embeds watermarks by modifying log-scaled singular value of selected coefficients of all sub-bands [16]. In a DWT-based scheme, the DWT coefficients are modified with the data that represents the watermark. Let $X = \{x_1, x_2 \dots x_N\}$ and $Y = \{y_1, y_2 \dots y_N\}$ be the vectors representing DWT coefficients of cover image and watermarked image in the HH2 region. For embedding, a bit stream is transformed into a sequence. This sequence is used as the watermark. In our case, the watermark $W = \{w_1, w_2 \dots w_N\}$ which is chosen from a set M, is embedded into X giving Y. W is inserted into the X by using multiplicative rule,

$$y_i = x_i(1 + \alpha_i w_i) \quad i = 1, 2, \dots N$$

where α_I is the embedding strength and x_i, w_i and y_i are the values of the random variable X_i, W_i and Y_i whose Probability Distribution Functions (PDFs) are $f_{X_i}(s_i)$, $f_{W_i}(w_i)$ and $f_{Y_i}(y_i)$ respectively for i = 1, 2, … N. The elements of the watermarks from the set M are independent and uniformly distributed in the interval [-1, 1].

B. *Maximum Likelihood Detection [1]*

The watermark detection is based on the maximum likelihood ratio. If $W^* = \{w_1^*, w_2^* \dots w_N^*\}$ is the embedded watermark, we can write $M = M_o \cup M_1$, where $M_o = \{W: W \neq W^*\}$ and $M_1 = \{W^*\}$. The null watermark $W = \{0, 0 \dots 0\}$, which indicates that no watermark is embedded, is already included in M_o.

Two hypotheses can be established as follows:

H_o = Y has W*

H_1 = Y does not have W*

The statistical decision test or watermark presence detection test is interpreted as deciding if the input of the detector in the outcome of the random process with the pdf conditioned to H1 and H0. It compares the ratio between the pdf conditioned to H1 and the pdf conditioned to H0 against a threshold as given below.

If the likelihood ratio,

$$l(y) = \frac{f_Y(y/M_1)}{f_Y(y/M_0)} > \lambda \qquad (1)$$

where $f_Y(y/M_j)$, j = 0, 1 are the conditional pdfs and λ is the decision threshold.

Since $\alpha < 1$, from [4]

$$f_Y(y/M_o) \approx f_Y(y/0) \qquad (2)$$

Assuming that the transform coefficients are statistically independent, (1) can be expressed as

$$l(y) = \frac{\prod_{i=1}^{N} f_{Y_i}(y_i / w_i^*)}{\prod_{i=1}^{N} f_{Y_i}(y_i / 0)} \qquad (3)$$

$$= \frac{\prod_{i=1}^{N} \frac{1}{1+\alpha_i w_i^*} f_{X_i}\left(\frac{y_i}{1+\alpha_i w_i^*}\right)}{\prod_{i=1}^{N} f_{X_i}(y_i)} \qquad (4)$$

Since log x is an increasing function of x, log l(y) will reach its maximum value when l(y) reaches its maximum. Hence, taking natural log on both sides

$$z(y) = \sum_{i=1}^{N}\left[\ln f_{X_i}\left(\frac{y_i}{1+\alpha_i w_i^*}\right) - \ln f_{X_i}(y_i)\right] \qquad (5)$$

where $\lambda' = \ln \lambda' + \sum_{i=1}^{N} \ln(1+\alpha_i w_i^*)$ is the modified decision threshold.

C. *Decision Threshold [1]*

The Neyman-Pearson criterion is stated in terms of certain probabilities associated with a particular hypothesis test. Neyman-Pearson criterion is used to reduce the missed detection probability subject to a fixed false alarm probability. The Neyman-Pearson criterion is used to find λ' to minimize the missed detection probability for a fixed false alarm probability, PFA. PFA is fixed as 10^{-9}.

$$P_{FA} = P(z(Y) = \lambda'/M_o) = P(z(X) > \lambda')$$

$$= \int_{\lambda'}^{\infty} f_{z(x)} \ (z(X))\ d\ z(x) \qquad (6)$$

As the number of Z(x) is more than 30, central limit theorem can be applied and PDF of Z(x) can be assumed to be Gaussian.

Thus,

$$P_{FA} = \int_{\lambda'}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2_{z(x)}}} e^{\left(\frac{[z(x)\mu_{z(x)}]}{2\sigma^2_{z(x)}} \right)} d\ z(x) \qquad (9)$$

which gives $\lambda' = \text{erfc}^{-1}(2P_{FA})\ \sqrt{2\sigma^2} z(x)\ + \mu_{Z(x)}$

D. *Laplacian Model [1]*

The DWT coefficients are assumed to be modeled using the Laplacian distribution.

Each of the DWT coefficients is modeled by the Laplacian PDF given below

$$f_{X_i}(x_i) = 0.5\ b_i \exp(-b_i |x_i - \mu_i|) \quad -\infty < x_i < \infty \qquad (10)$$

with $b_i = \sqrt{2}/\sigma_i$ where σ_i^2 is the variance of X_i and μ_i is the mean of X_i. Substituting (10) in (4),

$$z(y) = \sum_{i=1}^{N} b_i \left[| y_i - \mu_i | - | 1 + \alpha_i w_i^* |^{-1} | y_i - \mu_i - \mu_i \alpha_i w_i^* | \right] > \lambda' \qquad (11)$$

Mean and variance are derived to be

$$\mu_{Z(x)} = \sum_{i=1}^{N} [1 - | 1 + \alpha_i w_i^* |^{-1} \{ b_i | \mu_i \alpha_i w_i^* | + \exp(-b_i | \mu_i \alpha_i w_i^* |) \}] \qquad (12)$$

and

$$\sigma^2_{z(x)} = \sum_{i=1}^{N} [1 + | 1 + \alpha_i w_i^* |^{-2} \{ 2 - \exp(-2b_i | \mu_i \alpha_i w_i^* |)] - 2 | 1 + \alpha_i w_i^* |^{-1} \exp(-b_i | \mu_i \alpha_i w_i^* |) - 2b_i | \mu_i \alpha_i w_i^* | \exp(-b_i | \mu_i \alpha_i w_i^* |) \{ | 1 + \alpha_i w_i^* |^{-1} + | 1 + \alpha_i w_i^* |^{-2} \}] \qquad (13)$$

Substituting (12) and (13) in (9) the decision threshold λ' is obtained.

III EXPERIMENTAL RESULTS AND DISCUSSIONS

Experiments are performed to prove that the proposed scheme has the robustness against to a wide range of attacks such as JPEG compression, Gaussian noise addition, median filtering, blurring, shift, and rotation. Images of Lena, and Crowd at the size of 512 x 512 are used, as shown in Fig. 1. Lena contains little detail; and crowd contains a large amount of detail [6].

Lena **Crowd**

Fig.1 Test Images

Digital image watermarks can be detected in the transform domain using maximum-likelihood detection, whereby the decision threshold is obtained using the Neyman-Pearson criterion. Each image is transformed by DWT. Generally, in a two-dimensional DWT, each level of decomposition produces four bands of data denoted by LL, HL, LH, and HH. The LL sub-band can further be decomposed to obtain another level of decomposition. This process is continued until the desired number of levels determined by the application is reached [19]. In our experiments a Daubechies filter is used to obtain a third and a second level decomposition. In the third level decomposition, embedding is done in the high frequency sub bands LH3, HL3 and HH3. Total number of coefficients after combining the three bands is N = 12,288. If a coefficient belongs to the particular band, mean μI and variance σi2 are estimated from the equations,

$$\mu_i' = \frac{1}{N_B} \sum_{i=1}^{N} y \qquad (14)$$

$$\sigma_i^2 = \frac{1}{(N_B - 1)} \sum_{i=1}^{N} (y - \mu_i')^2 \qquad (15)$$

where $N_B = 4096$ (Number of coefficients in one band).

Y is the DWT coefficient in band B of the watermarked image. $\alpha = 0.3$.

The results of detection are listed in Table I.

Image	PSNR for α=0.3	Number of successful detections for level III embedding for 10 trials					
		Gaussian noise	Mean filter	Blur	Rotation	JPEG Compression	Crop
Lena	37.24	10	10	5	0	10	10
Crowd	30.89	10	10	9	0	10	10

scheme has the robustness against to a wide range of attacks such as JPEG compression, Gaussian noise addition, median filtering, blurring, shift, and rotation. Images of Lena, and Crowd at the size of 512 x 512 are used, as shown in Fig. 1.

Table I contains the number of successful detections for some standard attacks on the watermarked images. Gaussian noise has zero mean and variance 0.5. Blurring is caused by circular filter of the size 31 x 11.

Rotation is upto 10° in the counter clockwise directions. JPEG compression is done to offer 50% quality. Cropping is done to obtain an image whose size is 300 x 300. Mean filter filters the image by using adaptive wiener filter, using neighborhoods of size 4 x 4.

The embedded watermark is chosen from a set of 100 randomly generated watermarks of length N. Number of trials is 10.

In the second level of decomposition, embedding is done only at HH2. Total number of coefficients in HH2 is 16,384. μ and $\sigma 2$ are calculated only for this band using 14 and 15. α = 0.5 and PFA = 10^{-9}. The results of detection are listed in Table II.

Table II

Image	PSNR for α=0.5	Number of successful detections for level II embedding for 10 trials					
		Gaussian noise	Mean filter	Blur	Rotation	JPEG Compression	Crop
Lena	44.97	9	10	10	10	10	10
Crowd	40.63	10	10	10	10	10	10

Comparing Table I and II, we observe that Level II HH2 embedding better imperceptibility and better robustness. Also its capacity is better. Table III lists the PSNR value of level III embedding for different images.

TABLE III

Image	PSNR for α = 0.5	PSNR for α = 0.3
Lena	32.00	37.24
Crowd	26.49	30.89

IV Conclusion

Watermarks and watermarking techniques can be divided into various categories and in various ways. The indispensable and most frequently used partitioning of image watermarking is the spatial domain, transform domain, and parametric domain watermarking. The embedding of watermark in the cover image can be done either by additive or multiplicative rule. This paper proposed an efficient blind watermark detection scheme using DWT coefficients. A maximum likelihood detection scheme based on Laplacian modeling of coefficients of DWT transformation is implemented. The results obtained at level II, HH2 sub-band embedding are better than the results obtained using the existing method of embedding at level III. The proposed method is tested for imperceptibility, robustness and capacity and proved to have better robustness and better imperceptibility and better capacity than other conventional watermarking techniques that were proposed earlier in literature. In future this can be extended by implementing other statistical modeling.

V References

1) M.Ng, H.K. Garg, "Maximum-Likelihood Detection in DWT domain Image Watermarking using Laplacian Modeling", *IEEE Signal Processing Letters*, Vol. 12, No. 4, pp. 285-288, April 2005.
2) Seong-Geunkwon, Suk-HwanLee, Kee-Kpp Kwon, Ki-Rying Kwon and Kuhn-II Lee, "Watermark Detection Algorithm using Statistical Decision Theory", in *Proceedings of IEEE International Conference on Multimedia Expo,* Vol. 1, pp. 561-564, 2002.
3) Ng Tek Ming, Hari Krishna Garg, "A Maximum a Posteriori Identification Criterion for Wavelet Domain Watermarking", *IEEE International Conference on Distributed Computing Systems Workshops ICDCSW '04*, pp. 178-183, 2004.
4) Barni, Bartolini, De Rosa and Piva, "A new decoder for the optimum recovery of Non addtivie watermarks", in *IEEE transactions on image processing*, Vol. 10, No. 5, pp. 755 – 766, 2001.
5) Cheng, Huang, "A DCT-Domain blind watermarking system using optimum detection on Laplacian model" in *IEEE International Conference on Image Processing,* Vol. 1, Vancouver, BC, Canada, pp. 454-457, 2000.
6) Gonzalez and Woods, "*Digital Image Processing"*, Second Edition, Pearson Education, 2004.
7) K. Stefan and F. A. P. Petitcolas, Eds., Information Hiding Techniques for Steganography and Digital Watermarking, Artech House, Boston, 2000.
8) M. L. Miller, G. J. Doerr, and I. J. Cox, "Dirty-paper trellis codes for watermarking," *in Proceedings of IEEE International Conference on*

Image Processing, Rochester, New York, USA, vol. 2, pp. 129–132, 2002.

9) M. L. Miller, G. J. Do¨err, and I. J. Cox, "Applying informed coding and embedding to design a robust high-capacity watermark," *IEEE Transactions on Image Processing*, vol. 13, pp. 792–807, June 2004.
10) Q. Cheng and T.S. Huang, "Blind digital watermarking for images and videos and performance analysis," *IEEE International Conference on Multimedia and Expo*, Vol. 1, pp. 389-392, 2000.
11) G.C. Langelaar, I. Setyawan and R.L. Lagendijk, "Watermarking digital image and video data: a state-of-art overview," *IEEE Signal Processing Magazine*, pp. 20-46, Sept. 2000.
12) J. Cox, M. L. Miller, and J. A. Bloom, "Digital Watermarking," Morgan Kaufmann Publishers, 2002.
13) Content Protection Status Report III, November 7, 2002, available at http://judiciary.senate.gov/special/mpaa110702.pdf .
14) Chirawat Temi, Somsak Choomchuay, and Attasit Lasakul, "A Robust Image Watermarking Using Multi resolution Analysis of Wavelet," *Proceedings of ISCIT* 2005.
15) Chiou-Ting Hsu, and Ja-Ling Wu, "Multi resolution Watermarking for Digital Images," *IEEE Transactions on circuits and systems II, Analog and Digital Signal Processing*, vol. 45, no. 8, pp. 1097-1101, 1998.
16) Kyung-Su Kim, Min-Jeong Lee, and Heung-Kyu Lee, "Blind Image Watermarking Scheme in DWT-SVD Domain," *Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIHMSP 2007*, vol. 2, pp. 477-480, 2007.
17) H. Inoue, A. Miyazaki, A. Yamamoto, and T. Katsura, "A digital watermark based on the wavelet transform and its robustness on image compression and transformation," *IEICE Transactions on Fund. Electronics and Communication, Computer Science*, vol. E82-A, pp. 2–10, Jan. 1999.
18) H.-J. M. Wang, P.-C. Su, and C.-C. J. Kuo, "Wavelet-based digital image watermarking," Opt. Express, vol. 3, no. 12, pp. 491–496, Dec. 7, 1998.
19) Emir Ganic, and Ahmet M. Eskicioglu, "Robust Embedding Of Visual Watermarks Using DWT-SVD," *Journal of Electronic Imaging*, 2005.
20) Ding Wei, Yan Weiqi, and Qi Dongxu, "Digital image watermarking based on discrete wavelet transform," *Springer, Journal of Computer Science and Technology*, vol. 17, no. 2, pp. 129-139, 2002.

An Algorithm to Reconstruct the Missing Values for Diagnosing the Breast Cancer

F. Paulin, A.Santhakumaran

GJCST Computing Classification
I.2.6, F.1.1 & H.2.m

***Abstract*- The treatment of incomplete data is an important step in pre-processing data prior to later analysis. The main objective of this paper is to show how various methods can be used in such a way that they are able to process dataset with missing values. Computer–aided classification of Breast cancer using Back propagation neural network is discussed in this paper. The classification results have indicated that the network gave the good diagnostic performance of 99.06%.**

Keywords- Artificial Neural Networks, Back propagation, Breast cancer, Successive Iteration.

I INTRODUCTION

Cancer begins in cells, the building blocks that make up tissues. Tissues make up the organs of the body. Normally, cells grow and divide to form new cells as the body needs them. When cells grow old, they die and new cells take their place. Sometimes, this orderly process goes wrong. New cells form when the body does not need them, and old cells do not die when they should. These extra cells can form a mass of tissue called growth or tumor. Tumors can be either cancerous (malignant) or non-cancerous (benign). Malignant tumors penetrate and destroy healthy body tissues. Cancer that forms in tissues of the breast, usually the ducts (tubes that carry milk to the nipple) and lobules (glands that make milk) are called Breast cancer. Breast cancer is one of the leading cancers for women worldwide. It is the second most common cause of cancer death in white, black, Asian/Pacific Islander and American Indian/Alaskan native women [1], [2]. Early detection and improved therapy planning are crucial for increasing the survival rates of cancer patients. To aid clinicians in the diagnosis of breast cancer, recent research has looked into the development of computer aided diagnostic tools.

Neural networks have been widely used for breast cancer diagnosis [3], [5]. The effectiveness of breast cancer classification by training neural networks using a linear programming technique is demonstrated in [7].

Manuscript received "26/02/2010"

F. Paulin is with the MCA department as Senior Lecturer, CMS College of Science and commerce, Coimbatore, Tamilnadu, India. She is doing her ph.d in computer science from Mother Teresa university, Kodaikannel, India. Her research interest is artificial neural network.
(Telephone: 98422 67441 email: paulinrex@rediffmail.com)
Santhakumaran is currently a Reader in the department of statistics, Salem Sowdeswari College, affiliated to Periyar University, Salem, and TamilNadu. He received his PhD in Mathematics-Statistics from the Ramanujam Institute for advanced study in Mathematics, University of Madras. His research interests statistical Quality Control and Stochastic Processes and their applications.
(Telephone: 9443995082 m)

Artificial Neural Network (ANN) has made a significant mark in the domain of health-care applications. The brain learns from experience, in ANN, learning is typically achieved through progressive adjustment of the weighted interconnections of neurons and other network parameters, guided by learning algorithm. Feed forward neural networks have been trained with standard Back propagation algorithm [8]. They are supervised networks so they require a desired response to be trained. They have been shown to approximate the performance of optimal statistical classifiers in difficult problems.

There is much research on medical diagnosis of breast cancer with Wisconsin Breast Cancer Data (WBCD) in neural network literature [9]-[13]. In real world applications, missing values often abound. Therefore, there is a need for algorithms that can cope with missing values. Missing values in a datum mean that the values for some of the attributes of that datum are unknown. In [4], the 16 missing value instances have been left out while using WBCD for Breast Cancer diagnosis. The constructed feed forward neural network has been evaluated for breast cancer detection without replacing missing values [16]. Eliminating some instances will affect the diagnosis accuracy. The seventh attribute called Bare Nuclei of WBCD has 16 missing values. This paper presents a result of direct classification of data after replacing missing values using various methods for the WBCD dataset with a given number of classes.

A. Data Set

This breast cancer database is downloaded from the UCI machine-learning repository [14], which was collected by Dr. William H. Wolberg from the University of Wisconsin Hospitals, Madison [6]. The dataset is comprised of elements that consist of various scalar observations. The total number of the original samples is 699 with 16 samples contain missing values. The dataset contains two classes referring to benign and malignant samples. There are 458 samples in the dataset that are assigned to benign and the other 241 samples are malignant. The original dataset contains 11 attributes including both sample id number and class label, which are removed in the actual dataset that are used in our experiments. The remaining 9 attributes represent 9 cytological characteristics of breast fine-needle aspirates (FNAs), as shown in Table 1. The cytological characteristics of breast FNAs were valued on a scale of one to ten, with one being the closest to benign and ten the most malignant.

Table 1
Attribute Information

No.	Attribute	Domain
1	Clump thickness	1-10
2	Uniformity of cell size	1-10
3	Uniformity of cell shape	1-10
4	Marginal Adhesion	1-10
5	Single Epithelial cell size	1-10
6	Bare Nuclei	1-10
7	Bland Chromatin	1-10
8	Normal Nucleoli	1-10
9	Mitoses	1-10
10	Class	2 for benign and 4 for malignant

Number of instances : 699
Missing Attributes : 16
Benign : 458

II METHODOLOGY

In ANN, the assigned weights for each connector of node resemble the long term memory. They contain information of the input's importance and ANN learns by repeated adjustments of these weights. The weight adjustments are carried out according to the mathematical functions known as learning or activation function, which will be compared to the threshold value of the network. A feed forward back propagation artificial can learn a function of mapping inputs to outputs by being trained with cases of input-output pairs. Back propagation neural network (BPNN) is actually a descending slope method to minimize the total square of the output, calculated by the network [15]. There are three phases in the training process: first is to send the signal pattern forward, second is to calculate the propagated error and the last is to update all weights in the network. In addition BPNN also has the advantages of faster learning in multilayer Neural Network, especially sigmoidal activation function is represented by hyperbolic tangent. The neurons in feed forward networks can be any transfer function of the designer wishes to use. The usually used transfer function is the sigmoid function with threshold defined as in equation (1).

$$f(\sum_{i=1}^{n} w_i x_i - \theta) = 1/(1+\exp(-(\sum_{i=1}^{n} w_i x_i - \theta))) \qquad (1)$$

where x_i is the input to the node and w_i is the corresponding input weight, θ is a value which is usually called the threshold, n is the number of inputs to the node.
The network performance and convergence depends on many parameters like initial weights, learning rate and momentum used during the training process. Fig1. illustrates the flowchart of the overall processes used in this research.

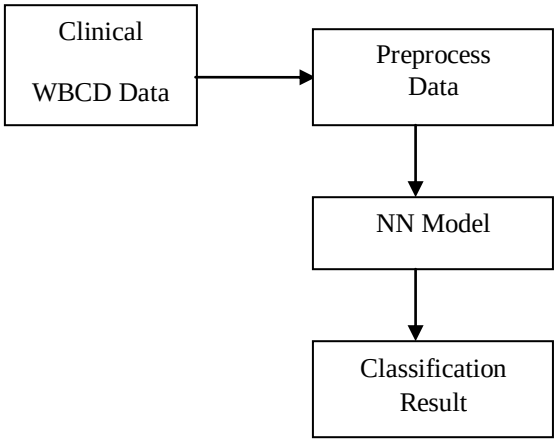


Fig.1. Overall Processes Flow Chart.

The proposed algorithm used in this research is follows.

i. Load Data set.
ii. Replace missing values by any one of the missing values replacement method (mean or median or mode or successive iteration).
iii. Normalize each variable of the data set, so that the values range from 0 to 1. We call this data set as normalized data set.
iv. Create a separate training set and testing set by randomly drawing out the data for training and for testing.
v. Create an initial ANN architecture consisting of three layers, an input, an output and a hidden layer. The number of nodes in the input layer is the same as the number of inputs of the problem. Randomly initialize the nodes of the hidden layer. The output layer contains 1 node. Randomly initialize all connection weights within a certain range.
vi. Train the network on the training set by using Back propagation algorithm until the error is almost constant for a certain number of training epochs, this is specified by the user.
vii. Present the test data to the trained network and evaluate the performance.

Replacement of missing values using Mean Method:

i. Find mean (average) for the Bare Nuclei (This attribute contains missing values).
ii. All the missing value of this attribute replaced by this mean value.

Replacement of missing values using Median Method:

i. Find median (middle value) for the Bare Nuclei (This attribute contains missing values).
ii. All the missing value of this attribute replaced by this median value.

Replacement of missing values using Mode Method:

i. Find the highest value for the Bare Nuclei (This attribute contains missing values).
ii. All the missing value of this attribute replaced by this highest value.

Replacement of missing values using Successive Iteration Method:

i. Find mean for the Bare Nuclei (This attribute contains missing values).
ii. Replace this mean for the first missing value.
iii. Again, find the mean for the entire attribute.
iv. If the new mean and old mean are same then replace this mean value for all missing values and stop the iteration
v. Else, perform step 1.

III AN APPLICATION

Preprocessing the input data set for a knowledge discovery goal using the neural network approach usually consumes the biggest portion of the effort devoted in the entire work. A simple analysis shows that the WBCD data set has missing information in the field of Bare Nuclei for 16 records. In this research, these missing values have been replaced by the calculated value using various replacement methods.
The following pre-classification rule have adopted in this work. In which three fields are included: Clump thickness, Bare Nuclei, and Mitoses as given below.
If (Clump thickness < 7 and Uniformity of cell size<8 and Uniformity of cell shape<3 and Normal Nucleoli<9) then

Benign

Else

Malignant

A. Normalize the Data

One of the most common tools used by designers of automated recognition systems to obtain better results is to utilize data normalization. Data normalization can also speed up training time by starting the training process for each feature within the same scale.
Input data has been normalized by the formulae as in (2), in the range between 0 and 1:

$$\bar{X} = (x - x_{\min}) / (x_{\max} - x_{\min}) \qquad (2)$$

Where, $\bar{X}$ is standard value of input,
x is Observed value,
$x_{\max}, x_{\min}$ are minimum and maximum actual observed values.

The above mentioned approaches for the proposed algorithm have been implemented and tested with the breast cancer dataset from the University of Wisconsin Hospitals, Madison, collected by Dr. W. H. Wolberg. To analyze the data neural network toolbox, which is available in MATLAB, software is used.
The feed forward neural network (FFNN) consists of an input layer, an output layer and one hidden layer. With 9 features in each input vector and 2 values in each output vector, we select 5 nodes for the hidden layer. The training algorithm is a standard back-propagation with a set of initialized parameters. The non-linearity in the hidden layer is the sigmoid function. The Levenberg-Marquardt (trainlm) algorithm was applied to increase the training speed. This trainlm algorithm appears to be the fastest method for training moderate-sized feed forward neural networks (up to several hundred weights). In this application, 80% of the data were selected randomly and used to train and construct the network. The remaining 20% of the data were then used to test capability of the resulting network. The proposed algorithm was executed 10 times each with a different set of missing value replacement methods. In Table 3 the percentage of correct classification indicates the percentage of the patterns that were correctly classified by the constructed networks. Table 2 shows the accuracies of different replacement methods.

Table 2
Performance of the Replacement methods

S. No.	Missing Value Replacement Methods	Percentage of Correct Classification
1	Mean	98.92%
2	Median	99.06%
3	Mode	98.56%
4	Successive Iteration	98.63%

IV CONCLUSION

In this research, a feed forward neural network is constructed and the Back propagation algorithm is used to train the network. The proposed algorithm is tested on a real life problem, the Wisconsin Breast Cancer Diagnosis problem. In a paper four missing value replacement methods are used, among these four methods, Median method gave the good result of 99.06%. Preprocessing using min-max normalization is used in this diagnosis. Further work is needed to increase the accuracy of classification of breast cancer diagnosis.

V REFERENCES

1) American Cancer Society, Cancer Facts and Figures 2007, 2007th ed. American Cancer Society, 2007.

2) U. S. Cancer Statistics Working Group, "United states cancer statistics 2003 incidence and mortality (preliminary data)," National Vital Statistics, vol. 53, no. 5, 2004.
3) Tuba Kiyan And Tulay Yildirim, "Breast Cancer Diagnosis Using Statistical Neural Networks" Istanbul University, *Journal Of Electrical And* Electronics Engineering, Year 2004, vol. 4, Number 2, pp.1149-1153
4) Anupam Shukla, Ritu Tiwari and Prabhdeep Kaur, "Knowledge Based Approach for Diagnosis of Breast Cancer" IEEE International Advance Computing Conference, Patiala, India, March 2009, pg 6-12
5) Sudhir D. Swarkar, Ashok Ghatol, Amol P. Pande, "Neural Network Aided Breast Cancer Detection and Diagnosis Using Support Vector Machine" Proceedings of the International conference on Neural Networks, Cavtat, Croatia, June 12-14, 2006, pp. 158-163.
6) W. H. Wolberg and O.L. Mangasarian, "Multisurface method of pattern separation for medical diagnosis applied to breast cytology," in Proceedings of the National Academy of Sciences, vol. 87, pp. 9193-9196, U.S.A., December 1990.
7) K. Bennett and O. L. Mangasarian, "Neural Network Training via Linear Programming," Advances in Optimization and Parallel Computing. Elsevier Science Publishers, 1992.
8) Renato De Leone, Rosario Capparuccia and Emanuela Marelli, "A Successive Overrelaxation Backpropagation Algorithm for Neural-Network Training" IEEE Transactions on Neural Networks, vol. 9, No. 3, May 1998, pg 381-388
9) Jun Zhang MS, Haobo Ma Md MS, "An Implementation of Guildford Cytological Grading System to diagnose Breast Cancer Using Naïve Bayesian Classifier", MEDINFO 2004, M.Fieschi et al. (Eds),Amsterdam:IOS Press
10) Punitha, C.P.Sumathi and T. Santhanam, "A Combination of Genetic Algorithm and ART Neural Network for Breast Cancer Diagnosis" Asian Journal of Information Technology 6 (1):112-117, 2007, Medwell Journals, 2007.
11) S.M. Kamruzzaman and Md. Monirul Islam, "Extraction of Symbolic Rules from Artificial Neural Networks" Proceedings of world Academy of science, Engineering and Technology, vol. 10, Dec. 2005, ISSN 1307-6884
12) Rudy Setiono and Huan Liu, "Neural-Network Feature Selector" IEEE Transactions On Neural Networks, vol. 8, No. 3, May 1997, pg 664-662
13) Wlodzislaw Duch and Rafal Adamczak and Krzysztof Grabczewski, "A New methodology of Extraction, Optimization and Application of Crisp and Fuzzy Logic Rules" IEEE Transactions On Neural Networks, vol. 12, No. 2, March 2001, pg 227-306

14) UCI Machine Learning Repository. [http://archive.ics.uci.edu/ml/]. Irvine, CA: University of California, Center for Machine Learning and Intelligent Systems.
15) D.W. Ruck,, S.K. Rogers., M.Kabrisky., P. S. Meibeck., and M. E. Oxley., "Comparatives Analysis of Backpropagation & the extended Kalman Filter for Training Multilayer perceptrons", IEEE Transactions on Pattern Analysis and Machine Intelligence, June 1992, Vol 14, No 6, pg 686-691
16) F.Paulin and A.Santhakumaran, "Extracting Rules from Feed Forward Neural Networks for Diagnosing Breast Cancer" CiiT International Journal of Artificial Intelligent Systems and Machine Learning, vol. 1, No. 4, July 2009, pg 143-146

Locomotion and Devices Control in Quadriplegic People

Mauricio Plaza Torres Universidad Libre
mplaza200@yahoo.com

GJCST Computing Classification
J.3, C.5.m, D.1.2, I.6.3 & I.2.9

Abstract- **The persons who have suffered permanent paralysis of his members develop a dependency that limits his possibilities and his quality of life. Independence in his daily activities improves notably his autoesteem and liberates his family of an additional responsibility.**

The basic needs are the locomotion, activation of devices as light, television, doors remote control, or communications. The present project intends to identify the basic needs of quadriplegic people and the possible solutions using technology of infrared tracking.

I INTRODUCTION

The persons with limitations in his members develop dependency from other people, for his primary needs of movement and common and daily activities (Fig 1). This dependency deteriorates his life quality and of the person dedicated to his care. For this limitation, the autoesteem of the person with limitations is low, creating psychological and physical problems that can be solved partially by means of some type of technology.

Fig 1. Traditional Electric chair

The devices have to take special characteristics like low cost, control with movement of the head, in some cases with finger activation, good accuracy and easily to maneuver. The basic idea of the project consists of develop an infrared tracking device that allows the devices activation by means of the movement of the head to quadriplegic people [1].
In some cases the affected person has the possibility of basic finger movement that allows activating the devices like a switch or the activation by means of a trackball. It will be studied and will leave the use possibility of the two devices depending on the injury type and the mobility possibilities [3]. The development of this device improves the quality of life of the affected people and opens new alternatives for handling different activities in this kind of persons [2]. The involved costs are low, the operation capacity is simple and his development is possible. The prototype include a infrared tracking system, a computer that processes the signal from the system and a display with the chosen options, a power system for the relay control and a digital system that is the interface with the power unit [4].
The main objective is to determine the needs of the people with limitations of mobility and to propose solutions with devices bases on infrared tracking (Fig 2).

Fig 2. System proposed for the research

The specific objectives are:

i. To determine the primary needs of persons with limitations of movement
ii. To study the possible solutions to obtain a certain percentage of independence
iii. To use technology of infrared tracking to supply part of the needs of mobility and devices control.
iv. To escalate the project to different grades of incapability (small movements of hand, total immobility of members...)
v. To realize a prototype that allows a person with a certain grade of incapability, to be able to have high level of independence with the control of some devices and mobility in a specific area.

II METHODOLOGICAL APPROACH

i. Bibliographical review. Allows knowing the current advances in the devices control for quadriplegic systems. It allows the comparison of alternative systems as control with EEG. It allows

the review of current technologies in infrared systems.

ii. Surveys in the health sector. It allows understanding the specific needs in persons with these limitations. It allows validating the proposal and to determine if there are improves in the model.
iii. Prototype preparation and simulation.
iv. Prototype test with quadriplegic people and improvement of the model.
v. Final prototype development
vi. Technical test and prototype improvement
vii. Field test
viii. Model improvement
ix. Final prototype development

III Expected results

The final prototype improves the quality of life of quadriplegic people. The prototype must have the possibility of allowing the control of electrical wheel chairs, of some devices connected to a wireless control system, only by means of the movement of the head.

We hope that the solution should be practical, easy use and should not generate very high costs, as well as safety and immunity to the noise. These characteristics allow that the system should be robust characteristics and easy to reproduce. The persons who have this system must have the possibility of easy movement in a specific area by means of the movement of some part of the body that still has control.

The fact that is an infrared system allows being immune to electrical noise, as generated by the electric chairs and allows low costs and easy design of engineering. These characteristics are desirable to reduce costs and to make the project possible.

IV Reference

1) Axel Mulder. "Human movement tracking technology". Technical report, School of Kinesiology, Simon Fraser University, 2000
2) Kenneth Meyer , Hugh L. Applewhite , Frank A. Biocca, A survey of position trackers, Presence: Teleoperators and Virtual Environments, v.1 n.2, p.173-200, Spring 1999
3) Olivier Faugeras, Three-dimensional computer vision: a geometric viewpoint, MIT Press, Cambridge, MA, 2001
4) R. T. Tsai "An Efficient and Accurate Camera Calibration Technique for 3-D Machine Vision", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2001

Architectural Crises in Vehicular Ad-Hoc Networks

[1]P. Sai Kiran, [2]B. Thirumala Rao,
[3]Dr. L.S.S. Reddy
[1,2] Associate Professor
[1]psaikiran@hotmail.com,
[2]email2thiru@gmail.com, [3]director@lbrce.ac.in

GJCST Computing Classification
C.2.1 & C.2.m & J.m

***Abstract*- Vehicular Ad-Hoc Networks (VANETs) deals with cars equipped with short-ranged radios communicating with each other exchanging the information for increasing passenger safety and comfort. VANET will enable both vehicle-to-vehicle and vehicle-to roadside communications. Vehicular networking protocols will require nodes, that is, vehicles or road-side infrastructure units, to communicate directly when in range, or in general across multiple wireless links (hops). In this paper we would like to exploit the various Architectural Crisis faced by the VANETS and consider various issues to be addressed by the protocols designed for VANET.**

I INTRODUCTION

According to the definition of IEEE 802.11 Mobile Ad Hoc Network(MANET) is A network composed solely of stations within mutual communication range of each other via the wireless medium (WM). An emerging application area for Mobile Ad hoc Networks (MANET) is the automotive sector. Vehicular Ad-Hoc Networks (VANETs) deals with cars equipped with shortranged radios communicating with each other exchanging the information for increasing passenger safety and comfort.

VANETs have similar characteristics as mobile ad hoc networks, often in the form of multi-hop networks. Due to the high mobility of nodes network topology changes occur frequently. All nodes share the same channel leading to congestion in very dense networks. The decentralized nature of VANETs leads to the need for new system concepts and information dissemination protocols. In addition, new approaches for data and communication security have to be designed to fit the specific network needs and to guarantee reliable and trustworthy services.

VANET will enable both vehicle-to-vehicle and vehicle-to roadside communications. Vehicular networking protocols will require nodes, that is, vehicles or road-side infrastructure units, to communicate directly when in range, or in general across multiple wireless links (hops). Nodes will act both as end points and routers, since vehicle-to-vehicle communication can often be the only way to realize safety and driving assistance applications, while the deployment of an omnipresent infrastructure can be impractical and too costly. In fact, vehicular networks are emerging as the first commercial instantiation of the mobile ad hoc networking (MANET) technology. VANET therefore is not an architectural network and not an ad hoc network but a combination of both; this unique characteristic combined with high speed nodes complicates the design of the network. Because vehicles in motion have short connection durations with roadside access points, efficient use of this duration is important.

State-of-the-art vehicle safety systems are based on various types of sensors, e.g. radars, lidars, and vision sensors. However, sensor based systems give rise to the following drawbacks: i) the limited range and Field-of-view (FOV) limit sensing to nearest vehicles that are immediately around the vehicle of interest, and ii) the cost associated with these possibly sophisticated sensors limits their applicability only to luxury vehicles. Therefore, there is strong interest in the automotive community to investigate the key role communication-based safety systems could play in either complimenting or replacing some of the sensing-based systems due to their versatility (ability to support a wide variety of applications) and competitive cost. This paper is organized as follows. Section II gives details about the services provided by VANET. W hile Section III exploits the Architectural Crisis in VANETs and concludes in Section IV.

II SERVICES PROVIDED BY VANET

A. Inter-Vehicle Services

Vehicle-to-vehicle communication can be used to disseminate messages of multiple services generating their content using sensors within the vehicle. These services can include accident warning, information on traffic jams or warning of an approaching rescue vehicle. In addition, information on road or weather conditions can be exchanged. More elaborate inter-vehicle services are direct collision warning or intersection assistance with information on cross traffic.

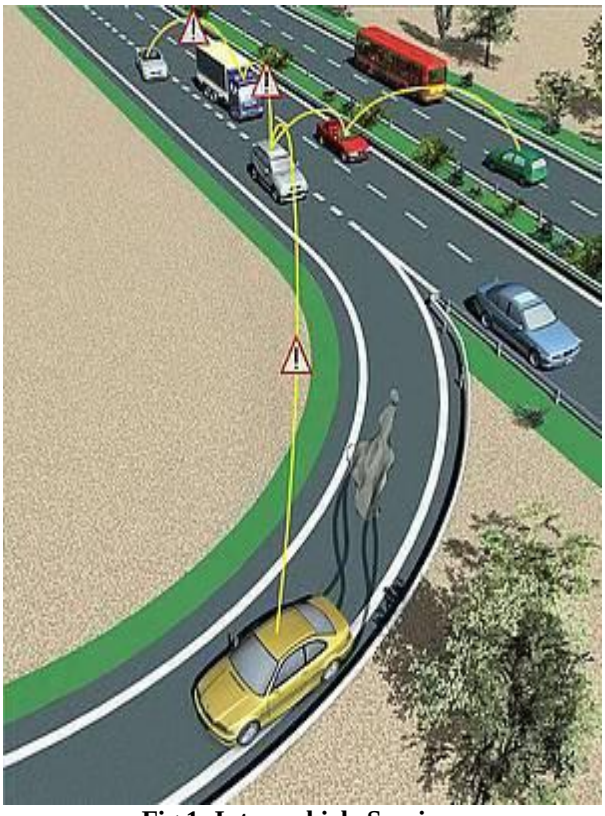


Fig 1: Inter-vehicle Services

B. Services Of Road Side Units

Communication between vehicles and RSUs can also increase safety. Traffic lights or road signs could be equipped with a communication device to actively inform vehicles in the vicinity. Hence, drivers can receive information on traffic flow, road conditions or construction sites directly from the respective RSU. In addition, static hazard areas, e.g. construction sites, could be equipped with a RSU to warn surrounding vehicles. RSU-based services will play an important role during the introduction phase, since they are almost unaffected by the penetration rate.

C. Portal-Based Services

Besides the safety related services, many other services related to the vehicle or providing entertainment to the passengers can be brought to future vehicles. The on board unit (OBU) inside the vehicle collects all incoming messages and sensor information. In addition, it relies on a server-based infrastructure providing many additional services. These can include information on parking or hotels as well as sightseeing information. A standardized solution opens the market to multiple service providers and reduces the time to market for service applications.

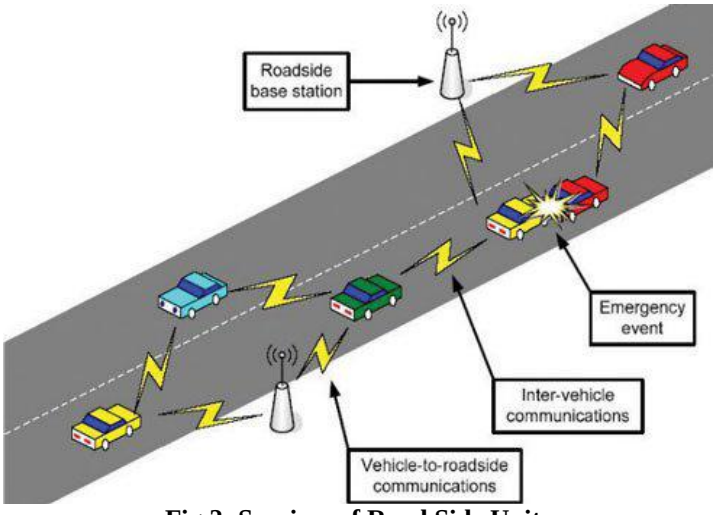


Fig 2: Services of Road Side Units

D. Integration Of Vehicles Into Backend Business Processes

Vehicles will certainly play a major role in everyday business processes that are currently
handled by enterprise IT systems. Two different ways of integrating cars into business processes
are considered valuable: First, data such as geographical position, covered distance or average speed may be transmitted to a company's backend system to allow for mobile asset management services. Logistics providers, for example, who nowadays run complex IT systems to manage their fleet, could feed real-time information into their applications to improve flexibility and adaptivity of their business processes. If such a system was enabled to receive the current,
geographical position of all vehicles, the firm could react to customer demands more agilely due to better capacity

forecasting mechanisms. Insurance companies and their customers might also be interested in connecting vehicles to backend IT services. Initiatives such as "Pay-as-youdrive" currently investigate the market potential of such applications. Drivers who only cover short distances and drive carefully would have to pay less than someone driving long distances. Besides the transmission of data from the car to backend IT application landscapes, the provisioning of car drivers with access to external data is a promising possibility of applying vehicular communications as well. Business people, which are always "on the move", such as sales persons or consultants, may be highly interested in leveraging their cars' onboard systems as a conventional workplace. Via speech input, drivers could trigger their cars to remotely access a company portal and to download crucial information for their next customer visit, for example.

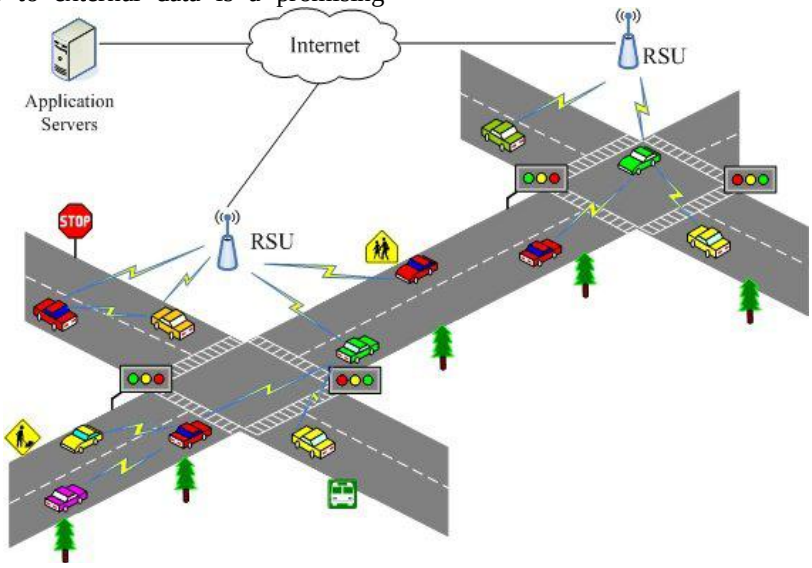


Fig 3: Integration of vehicles into backend business processes

III ARCHITECTURAL CRISIS

A. *Relationship To Sensor Networks*

While the notion of sensor networks usually stands for (non)-mobile wireless networks with lowpower and low-capability devices distributed gather sensor information, there are some important similarities between sensor networks and VANETs that might influence architectural considerations.

First, a vehicle can be seen as a high capability sensor device with sensors for environmental information such as road grip or temperature, and for information about the vehicle itself such as movement.

Second, the sensor information coming from different vehicles en route can be combined in order to eliminate redundancy, minimize the number of transmissions, and improve the quality of the sensor information. This 'data centric routing', as opposed to 'address-centric routing', is well known from sensor networks. In addition, the whole communication system might react to sensor information in the sense that sensor events are an integral part of network protocols. However, the main difference between VANETs and classical sensor networks is most likely that for VANETs, the main goal of these protocols is not the preservation of energy but a 'low channel utilization' to keep the system accessible for urgent safety messages.

B. *Packets vs. information*

Along the lines of the first observation, one has to differentiate between 'packets' and 'information': In classical networks, the data payload of a packet is meant to be delivered unchanged to the addressed application instance(s). However, VANET applications will most likely evaluate the information contained in a packet, merge it with their own state and then decide how to communicate this updated information. This operation is known as 'in-network' processing.

C. *End-To-End Notion Revisited*

In a traditional network, peer application and protocol entities are well-defined on all 'communication endpoints'—either by an ID or by a multicast group. However, the VANET communication entities might not only address specific peer entities, but also geographical or topological areas whose members are likely to change over time. Furthermore, a communication between two peers might only be possible in one direction but not vice versa.

D. *Network Protocol Requirements*

Among other things, the last observation directly leads to different requirements for multi-hop packet-forwarding protocols. On the one hand, traditional unicast and multicast protocols using ID-based addressing might still be needed

for infotainment applications or the extension of hotspot access. On the other hand, the challenge for VANET network protocols lies in efficient geocasting and flooding. Additionally, there might be potentially severe requirements concerning reliability and/or timeliness due to the safety purpose of some applications.

E. Granularity Of Control

In classical network management, the control parameters are set as 'mid-term' or 'long-term' parameters. E.g., the setting of an IP address tends to be long-term and even in UDP communication sessions packet options are usually changed for every session (mid-term) but not for single packets within one session. In a VANET, however, it seems that various control parameters will have to be set on a per-packet basis, as when sending successive packets different MAC algorithms could be used, different transmission powers could be set, or the packets could be sent to different physical channels.

F. Information Sharing

In a VANET, the communication system generates information that is of high value to many protocol entities. Beacon packets could be used to generate a list of neighboring nodes, that could be used both for driver assistance and packet forwarding decisions. Thus, we observe the need to share information in an efficient and clean manner without creating complex control interactions. In addition, the integration of these events into protocol state machines demands a standardized means to access them, if implementation portability is desired.

G. Application Requirements Vs. Medium Conditions

The safety-focused nature of VANETs requires the communication system to be dependably able to deliver important packets. To achieve this, the packets have to contend with (a) the sending demands of other nodes and (b) the allocation of the radio channel by other nodes. In addition to that, the channel itself is highly probabilistic. Thus, in order to meet application requirements, not only will all nodes have to cooperate among themselves but also all applications and protocols on a single node.

H. Challenges In VANET Protocols

VANET protocols that are able to make it to the product will stage the need to work under very different conditions. For the first couple of years, a car equipped with a VANET system will find hardly any other cars with which to productively exchange messages. Thus, the first task of the protocols will be to operate under these conditions. They will—in the beginning—not care very much about channel usage to maximize utility. I.e., in the beginning, the probabilistic channel will be used frequently to increase the utility range of VANET messages.

However, as system penetration increases, the scarcity of the radio channel as a resource will increase. Paying this will imply the absolute necessity of minimizing its usage acknowledging the increasing likelihood of packet collisions. Consequently, a significant challenge lies in building protocols that work in both cases, and a great danger lies in building protocols that are hastily tailored to cope with the low-density situation. The high-density situation, however, creates the greater challenge of seeing the multi-hop effect of single-hop broadcasts. This means that whenever information triggers broadcasts, the subsequent message exchange is, in fact, part of a multihop protocol, which has to be evaluated on a non-local scope by people with knowledge in ad-hoc networking.
It is—in our view—simply not enough to provide singlehop broadcast to application developers and then let them worry about the rest.

IV Conclusion

The architecture of a future VANET system is still not clear, at least for the projects we were involved in. While many people consider this an academic discussion, it has some impact on how protocol development can be separated and cooperation can be stimulated. Obviously, the Internet's end-to-end paradigm has to be reconsidered, since there is technically no backbone vs. end system structure, but every node is both end system and router. For the cooperation part, the know-how lies in the hands of different groups: On the one hand, there are network researchers holding knowledge about multi-hop protocols, retransmission timing, broadcast redundancy etc., and on the other, there are people with an understanding of traffic flow, time-criticalness of information distribution and so on. In our opinion, since both groups directly influence any resulting protocol's "radio profile", a stronger interaction is necessary to avoid a system with protocols that either will not really work in the beginning or choke the channel later. Especially the extreme high density situation in a congested highway under the assumption that every car is running a VANET system will create the ultimate protocol challenge. Any available protocol design trick will have to be used to tackle these problems: from using infrastructure to control the channel or to coordinate information gathering, over the usage of classical algorithmic methods to increase scalability like hierarchization, up to methods involving the electrical properties of the signals like power control or different physical coding to stabilize transmissions. Moreover, the problems will have to be tackled quickly since the car manufacturing industry is eager to roll out a car-to-car communication system, and the consequences in rolling out a closed-box system have a long reach.

V References

1) W. Franz, H. Hartenstein, and M. Mauve, Eds., "Inter-Vehicle- Communications Based on Ad Hoc Networking Principles— The FleetNet Project". Karlsruhe, Germany:

Universitätsverlag Karlsruhe, November 2005.
2) Doetzer, F., Kosch, T., Strassberger, M.: "Classification for traffic related intervehicle messaging". In: Proceedings of the 5th IEEE International Conference on ITS Telecommunications, Brest, France (2005)
3) Franz, W., Eberhardt, R., Luckenbach, T.: "Fleetnet - internet on the road". In: Proceedings of the 8th World Congress on Intelligent Transportation Systems. (2001)
4) Kosch, T.: "Local danger warning based on vehicle ad-hoc networks: Prototype and simulation". In: Proceedings of 1st International Workshop on Intelligent Transportation (WIT 2004). (2004)
5) Vehicle Safety Communicatiotus Consortium. htp: www-nrd.nhtsa.dot. gov/pdf/nrdl2/CAMP3/pages/VSCC.htm.
6) L. Wischhof, A. Ebner and H. Rohling, "Information Dissemination in Self-Organizing Intervehicle Networks", IEEE Transaction on Intelligent Transportation Systems, Vol. 6, No. 1, March 2005.
7) J.Lou, J-P. Huaux, " A Survey of Inter-Vehicle Communication", Technical Report, School of Computer and Communication Science, EPFL, Switzerland, 2004.
8) MI. M. Artimy. W. Robertson, anld W. Phillips. "Connectivity in Inter-Vehicle Ad Hoc Networks", CCGEI 2004, Niagara Falls, Maylmai, 2004.
9) S. Schnaufer, H. Filler, M. Transier, and W. Effelsberg, "Vehicular Ad-Hoc Networks: Single-Hop Broadcast is not enough," in Proceedings of 3rd International Workshop on Intelligent Transportation, Hamburg, Germany, March 2006, pp. 49–54.
10) J.J. Blum, A. Eskaindarian and L.J. Hoffman, J Challenges of Inter-vehicle Ad Hoc Network-s", IEEE Transaction on Intelligent Transportation Systems, Vol. 5, No. 4, December 2004.

Compiling Mechanical Nanocomputer Components

GJCST Computing Classification
D.2.2, C.5.3 & B.1.4

Thomas Way, Tao Tao
and Bryan Wagner
Department of Computing Sciences Villanova University
Villanova, PA, USA thomas.way@villanova.edu

Abstract- **Computer component fabrication is approaching physical limits of traditional photolithographic fabrication techniques. Alternative computer architectures are being enabled by the rapidly maturing field of nanotechnology, and range from nanoelectronics and bioelectronics to nano-mechanical computational machines and other nanoscale components. In this study, the design of a nanocompiler, which targets a simulated hydrocarbon assembler, is presented. The compiler framework demonstrates the feasibility of a hardware compiler to produce building block components, a necessary first step in full molecular assembly of nano-mechanical computers. As a proof of concept, the resulting nano-mechanical machine components are simulated using a Colored Petri Net model of a 32-bit adder and an atomic-level gate simulator. Performance and size bound estimates and key nano-mechanical component design issues are given.**

Keywords- Nanocompilers, Mechanical logic gates, Compiler-directed mechanosynthesis, modeling.

I Introduction

Despite recent significant advancements in feature size to 45nm, announced in 2007 by Intel, there is growing consensus that the familiar density-doubling prediction of Moore's Law as it concerns 2D fabrication techniques is reaching limits [1,2]. The trend clearly is away from techniques that fit increasing numbers of transistors onto a chip, as manufacturers pursue technology that enables increasing numbers of processing cores on a single chip [3]. The movement toward increased coarse-grained parallelism, the seeming approach of inherent limits of photolithographic techniques, and the continued maturation of the field of nanotechnology could hint at a serendipitous convergence of needs and capabilities.

The way forward in chip design and fabrication may well include applied computational nanotechnology as originally foreseen by Eric Drexler [4], and furthered by many others [5-8]. Current chip design techniques, and in fact virtually all software and hardware design of any significance, make use of a variety of automated compiler tools to generate complex designs, layouts or executable code from an original human-readable specification or source program [2]. Molecular manufacturing, and other applications of nanotechnology, are likely to require a similar approach in order to manage the scope and complexity of translating a high-level processor specification into nanoscale components. Computer science techniques, such as compiler design, optimization and software engineering, are likely to play an important role in the molecular design and fabrication process. [4]

\

Modern compilers for high performance computer architectures apply a sequence of sophisticated analyses and optimizations to translate a source language program into efficient binary machine code. Machine specific optimizations, customized to the particular target architecture, are required to achieve significant speedup on modern, high-performance architectures [2,9]. In spite of public announcements made in early 2007 of advances in feature-size reduction by Intel and AMD, heat dissipation and barriers of physics remain as problems [1,5]. Nanotechnology, manufacturing performed through manipulation of atoms and molecules, or through other nanoscale manufacturing techniques, is capable of overcoming these barriers [4]. The continuing trend toward flexible computer architectures with higher degrees of parallelism suggests that the field of reconfigurable computing, perhaps enabled through the use of nanotechnology, is the next evolutionary step in processor design [5,10].

Nanocomputing is taken to mean the class of highly reconfigurable, nanotechnology-based, computer architectures, and a nanocompiler is the software-hardware system that targets such a nanocomputer. In this paper, we present the design of a nanocompiler framework that targets one form of nanoscale computer architecture, nano-mechanical computing devices. This compiler framework translates a source code program into both an optimized executable program and a customized nanocomputer on which the executable program will be ideally suited to run. Much as traditional compilers customize the program to suit the machine [9], this proposed compiler customizes the machine to fit the program. Since no such nanocomputer architecture yet exists, our study demonstrates the approach using a molecular design language, a simulated hydrocarbon assembler, and a mathematical modeling tool to demonstrate and estimate the performance of this approach.

II Background

Nanocomputer architectures, which form a subset of reconfigurable architectures that include FPGA, FCCM, cellular array, synthetic neural systems and many others [6,10], are produced using some form of a molecular manufacturing and provide a natural successor to current general-purpose microprocessor architectures [6]. To be accepted, nanocomputers functionally must be at least as capable as their predecessors, fast, inexpensive, robust and capable of operating at room temperature and of executing legacy code [5]. In order to produce such nanoscale architectures, molecular manufacturing and

mechanosynthesis techniques must be understood, including the use of hydrocarbon assemblers. As this technology is not yet realized, the use of a variety of research tools to model and simulate the proposed architecture is required.

The capabilities of various nanotechnologies as applied to science and manufacturing generally are classified into one of a range of generations or stages. Roco describes four generations [11] of nanotechnology development: passive nanostructures, active nanostructures, systems of nanosystems and molecular nanosystems. Hall defines a set of five stages [12] of increasing precision, complexity and difficulty: bulk process chemistry, molecular self-assembly, cellular-scale machinery, special purpose macro-scale molecular assembly, and general molecular manufacturing. Current advancements in nanotechnology are in Roco's second generation or Hall's second stage. The research reported in this paper assumes advancements will continue forward to at least Roco's third generation and Hall's fourth stage.

A. *Nanoscale Computer Architectures*

Among the proposed nanoscale computer architectures are silicon-based resonant-tunneling devices (RTDs) consisting of tunneling diodes paired with field-effect transistors (FETs) [5], carbon nanotube semiconductors [13], diamondoid carbon transistors synthesized using Chemical Vapor Deposition (CVD) techniques [14], a variety of quantum, DNA-based and single-electron transistor (SET) nanoscale electronic devices [7], and novel nano-mechanical computing devices constructed of moving, nanoscale components [4]. Such nano-mechanical computers are reminiscent of Charles Babbage's Analytical Engine [2], albeit nine or ten orders of magnitude smaller. Although these potential technologies hold promise, all but nano-mechanical devices appear to be constrained to two dimensions, limiting their potential for improvement over time in the same way that current photolithographic techniques are limited. The recent press release announcement regarding the maturation of IBM's chip-stacking technology may hint at a future for 3D electronic devices, although issues of resistance, heat dissipation, interconnectivity and quantum tunneling remain as significant hurdles [4,7,12].

Nano-mechanical devices such as logic gates and registers will be constructed from a series of molecular logic rods called *interlocks* and driven by kinetic forces [4]. Although this atomic-scale computational machinery will operate more slowly than traditional electronic devices, the difference will be only approximately one order of magnitude due to its much smaller size. Nano-mechanical devices can have much higher densities since interlock logic gates can be stacked in three dimensions while transistors must be placed on a 2-dimensional substrate. The density of nano-mechanical devices is estimated to be 10^{11} greater than that of silicon transistor devices, enabling the very real potential to produce massively parallel networks of nanoscale processors [4]. Other benefits to nano-mechanical computing architectures include precision, tolerance to physical wear, and improved fault tolerance [15].

The interlocks used for logic gates as described in Drexler's architecture consist of sliding rods that have knob protrusions which slide between one another. Depending on their position, one rod may block another or allow a rod to continue sliding along its vector of movement, with input and output provided by additional interlock rods, all enclosed in a stiff housing. Rods can be combined easily to form a logic gate that computes the output of a NAND operation, a logical component from which all other logical operations can be constructed. Estimates for clock speeds of nano-mechanical logic gates are 1000 MIPS, or approximately 1 GHz. The ability to fabricate in three dimensions means that massively parallel processors could be formed in very dense volumes. A complete nano-mechanical CPU system of 106 logic gate "transistors" forming a cube 400 nm on a side, with each logic gate being smaller than a single rhinovirus, would fit in the same surface area as just 80 transistors at the 45 nm scale. [4]

B. *Mechanosynthesis And Hydrocarbon Assemblers*

Building molecular mechanical computers with atomic precision will require direct control of the chemical reactions that occur between atoms and molecules. The process of manufacturing machinery through such methods is known as Molecular Manufacturing [14]. So far, chemistry has relied on methods of controlled, probabilistic reactions resulting from collisions between masses of atoms in order to synthesize useful compositions, but without guarantee of atomic precision. Mechanosynthesis is performed using atomically precise tools that rely on chemical bonding to produce positional control of mechanical forces [16], which would be used to synthesize useful molecules from their constituent atoms, as well as higher-order structures [14].

Ideally, mechanosynthesis would be performed using a Universal Constructor, a nano-mechanical computer controlled machine that could follow sequences of instructions to assemble raw atomic material into arbitrary molecular structures, including exact copies of itself [15]. Drexler's proposal for a general-purpose molecular assembler is an example of such a constructor [4,15], an approach that is reminiscent of Alan Turing's notion of a Universal Machine which eventually led to the modern, stored-program computer [2]. Research has focused on the hydrocarbon assembler, which is a simplified universal constructor that builds diamondoid structures [4], including copies of itself. The small molecular computers in each assembler are controlled by a broadcast mechanism emanating from a macroscale computer. Instructions from the macroscale controller in such a broadcast architecture are passed through acoustic waves that vibrate through a liquid environment surrounding the assemblers. The instructions are received by mechanical pressure-actuated devices in the assemblers. These pressure activated devices initiate instructions to the mechanical logic units in the

assembler computer. As a result, the assembler can be reprogrammed to make different molecular structures [8].

C. Modeling Of Nano-Mechanical Structures

Molecular machinery can require billions or trillions of atoms in the final design. Thus, there is a need for a set of "molecular compilers," which take a high level of abstraction input for molecular components and transform them into atomically positioned devices [10]. To enable high-level specification of nano-mechanical structures, an abstract scripting language, MolML, was developed that can be compiled into hydrocarbon assembler instructions [17]. MolML is an Extensible Markup Language (XML) based language that is designed to factor out the redundancy in molecular structures that have large amounts of symmetry and repetition. The language is written to provide communication between a macroscale computer and a molecular assembler, facilitating the correspondence of input instructions to the assembler. Using MolML, it is possible to define 3-dimensional molecular structures of arbitrary complexity [17].

In order to visualize and simulate the molecular structures defined with MolML, a software tool was developed to parse, display and animate a MolML design. The MolSim tool [17] implements basic Newtonian Laws by using vector geometries, accounting for the position, velocity, and acceleration of particles in the simulation environment. The simulation engine was tested against a MolML document that defined the structure of Drexler's nano-mechanical NAND gates, using the tetrahedral lattice molecular structure of rigid diamond molecules to form the logic rods. For efficiency concerns, the housing was omitted. The molecular design was inspired from an examination of the repetitive symmetries inherent in the lattice [6].

Simulation of atomic interactions and molecular structures in the current version of MolSim is infeasible for larger structures, so these larger structures such as adders can be constructed as formal, mathematical models using Colored Petri Nets (CPNs) [18]. Petri Nets provide a modeling language that is well suited for larger systems, drawing on the power of generalization provided by mathematical modeling techniques. In practice, a CPN model is created using a graphical tool, enabling visual representation and analysis of a CPN model. With the addition of timing parameters, CPN can realistically model the function of a nano-mechanical NAND gate and higher-order molecular structures constructed hierarchically by reusing the NAND model in varying configurations.

D. Nanocompiler Design

The two major goals of the work reported in this paper are refinement of the design of a nanocompiler [19] and a simulation of a common nano-mechanical component that can be generated by the nanocompiler. Our nanocompiler design generates as its output both as an executable version of the original source program and the description of a machine on which the executable will run. Traditional compilers take the source code and translate it into a binary form suitable for a specific processor, optimized to run as well as possible on that target machine. Knowledge of the target machine is needed to perform machine-dependent optimizations. Our approach is a generalization of compilation for reconfigurable computing in that the configuration of the target machine is unknown when compilation begins. The machine configuration is extracted from the source program, based on analyses of program characteristics. In this way, the resulting machine is an excellent fit to the program.

Figure 1 illustrates the organization of the proposed nanocompiler from a high level. Source code is processed by the Front end of the compiler, including machine-independent optimizations. The resulting intermediate form is passed to a Machine requirements analysis phase, which performs static analysis, providing metrics to the Machine description generation phase. The resulting machine description is used by the Processor generator phase to generate or reconfigure the target machine, and by the compiler Back end to perform machine-dependent optimizations and generate the executable code. The Processor generator and even runtime profiling information can feed information back into the Machine requirements analysis phase to enable iterative refinement of the machine description, and thus of the processor itself.

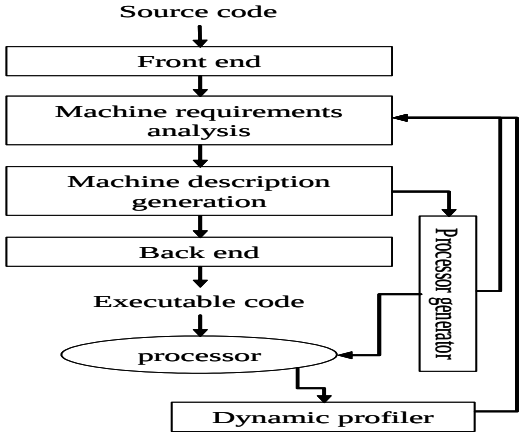


Fig.1. Organization of a nanocompiler

Inside the Processor generator (Figure 2), a machine is reconfigured or generated using a nanotechnology approach. The machine description is analyzed through a sequence of phases that translate the description into a layout of circuits (e.g., VHDL) or other structures (e.g., MolML) that implement the machine, which in turn is implemented using logic gates, which are either reconfigured as with FPGAs or assembled using molecular manufacturing techniques, to produce the target processor.

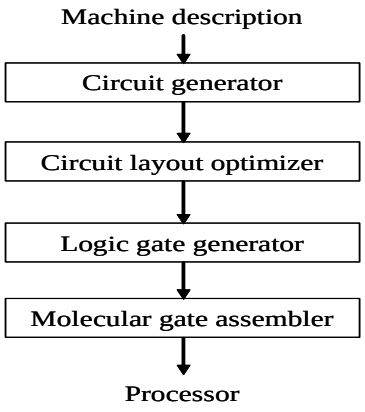


Fig.2. Organization of a nanotechnology-based processor generator.

In the current implementation, we focus on the generation of a nano-mechanical machine description for an addition instruction. Thus, we restrict the scope of the research results reported in this paper to the machine description and processor generation phases, and more specifically on how a machine description for low-level components, nano-mechanical NAND gates, can be combined into a higher-level component, a 32-bit nano-mechanical adder. Designing nano-mechanical computer components at this scale raises a number of important issues that may need to be addressed in future research, including

A. Compilation Time

Determine the time needed to analyze source code, produce a machine specification, and fabricate components using hydrocarbon assemblers, which can enable near-exponential assembly capability through an initial self-replication process.

B. Design Feasibility

Determine whether the approach is feasible, and identify further what technology must be developed to construct a complete, working nanocompiler. Extend 2-D fabrication to an understanding of 3-D fabrication made possible with mechanical components.

C. Performance Efficiency

Determine the efficiency of inherently slower nano-mechanical processing, evaluate for potential speedup through increased parallelism, and predict throughput and related latencies. Design realistic models that reflect predicted behavior.

D. Requirements Analysis

Determine what essential information is needed in the compiler, including analyses performed for parallelization, resource requirements, and efficiently scalable reconfiguration and fabrication.

E. Usability Of Massive Parallelism

Identify approaches that can successfully utilize the extreme parallelism that may be available, including integration of ILP, thread- and task-level parallelism.

III SIMULATION OF NANOSCALE COMPONENTS

Simulation of a 32-bit nano-mechanical adder was accomplished through a sequence of four developmental stages: nano-mechanical NAND gate, CPN model of a NAND gate, CPN model of a 1-bit adder, and a CPN model of a 32-bit adder. In the first stage of development, an atomic-level nano-mechanical description language (MolML) and a simulation tool (MolSim) were designed, enabling a realistic, visual modeling of a single NAND gate (Figure 3). By extending earlier work on MolSim [17], we created an atomic and molecular description of a NAND gate that could be generated given a few positional parameters, which is then translated by MolSim into the full atomic and molecular description that provides details of the placement and configuration for each atom and molecule.

Based on proposed functional parameters of nano-mechanical logic gates [4], a variety of estimated characteristics are provided in Table I. The rod component of the NAND gate can be constructed using relatively few Carbon atoms, while the housing which encloses the gate (not shown) requires significantly more. Atomically precise modeling tools are needed to quantify these material requirements. The proposed speed for a nano-mechanical processor is approximately 1 ns per operation, operating at 1000 MIPS, or 1 GHz, with logic and arithmetic operations requiring about 1 ns [4]. We assume 0.1 ns per NAND calculation, with 0.2 ns for reset latency. The volume of a single NAND gate is approximately 16 nm3 [4], so 6.25e+16 such components can fit in a single cubic millimeter. Clearly this is a much higher density than the current best practical feature size of 45 nm.

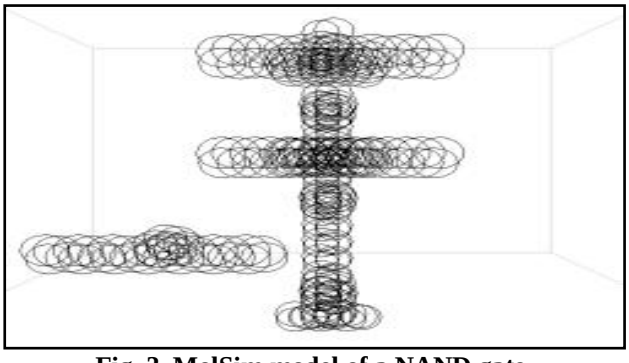


Fig. 3. MolSim model of a NAND gate.

Table I. Estimated characteristics of mechanical NAND gate model [4].

Characteristic	Estimated value
Time to perform operation (Transistors require 0.01 ns)	0.1 ns
Time to reset (2x op time)	0.2 ns

Surface area of gate (Transistors each require 10^6 nm^2)	4 nm^2
Volume of gate (gates are stackable in 3D)	16 nm^3
Improvement in volumetric packing density compared to transistors	$>10^{11}$

In the second stage, CPN was used to model the same NAND gate at a higher level of abstraction (Figure 4), but with identical behavior characteristics. Time units are measured by the addition of a timestamp notation to the NAND gate, such that the function of each NAND gate accounts for one unit of time. Time units accumulate during the simulation, enabling straightforward scaling to other time units as needed. Because of the nature of the binary calculations being performed, the NAND gate was designed to operate on Boolean, rather than decimal, values.

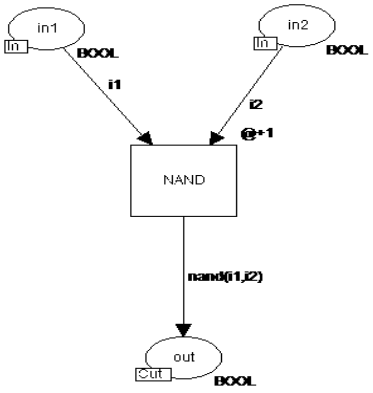


Fig. 4. CPN model of a NAND GATE.

In the third stage, a CPN model was constructed of a 1-bit adder by connecting nine NAND gate models in a standard adder configuration (Figure 5). Time measurement of the NAND gate model produced output of a carry value in 5 units of time, and a sum value in 6 time units, which equates to a total time of 0.6 ns to perform a 1-bit addition.

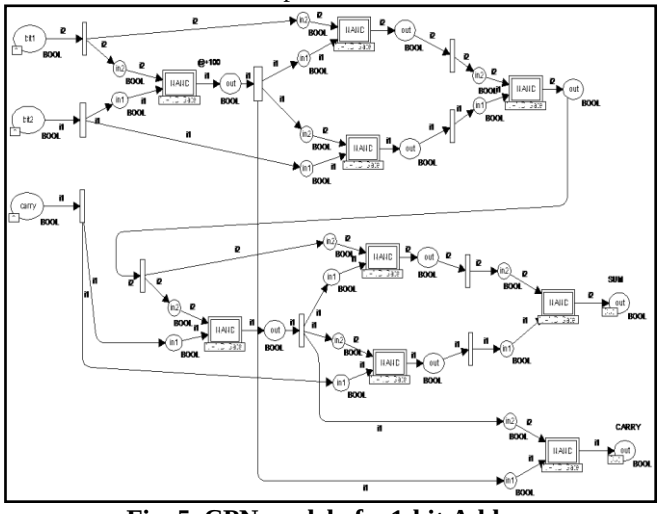


Fig. 5. CPN model of a 1-bit Adder.

In the fourth and final stage, a CPN model of a 32-bit adder was constructed using the hierarchical features of CPN Tool, combining four 1-bit adders into a 4-bit adder, then four 4-bit adders in a 16-bit adder, and finally two 16-bit adders into a 32-bit adder (Figure 6). Timing for addition of two 32 bit values was a total of 68 time units, which is generalized to the equation: 2(n-1)+6, where n is the number of bits. Although the carry value requires 5 units to calculate in each individual adder, because calculation of the sum does not require the previous carry for the first 3 time units, significant overlapping (i.e., parallel) computation occurs.

Table II summarizes selected characteristics, providing initial estimates. Based on our models, and other proposed characteristics [4], a 32-bit nano-mechanical adder requires 288 NAND gates connected using roughly 96 connector rods, possibly constructed of nanotubes in a housing. The models predict 6.8 ns per addition, with latency between the start of subsequent additions of 7.0 ns, including the 0.2 ns reset time. Drexler's proposed CPU requires approximately 1 ns per operation [4], suggesting that some amount of parallelism may be inherent in a nano-mechanical adder that is not captured in the current model. At this estimated size, 2e+14 such adders would fit in a volume of one cubic millimeter. Extrapolating these values, the presented 32-bit adder model is within specification of a proposed nano-mechanical CPU that would be contained in a cube 400 nm on a side [4].

Table II.
Estimated characteristics of a 32-bit nano-mechanical adder model.[4]

Characteristic	Est. value
Total NAND gates @ 16 nm^3	288
Total connectors @ 1 nm^3	96
Estimated volume (approx 16 nm cube) (288 x 16 nm^3) + (96 x 1 nm^3)	4,704 nm^3
Estimated time for 1-bit addition	0.6 ns
Time to perform 32-bit addition based on simulation	6.8 ns
Predicted time of 32-bit addition [6]	1 ns
Estimated throughput of model	142 MIPS
Predicted throughput [6]	1000 MIPS

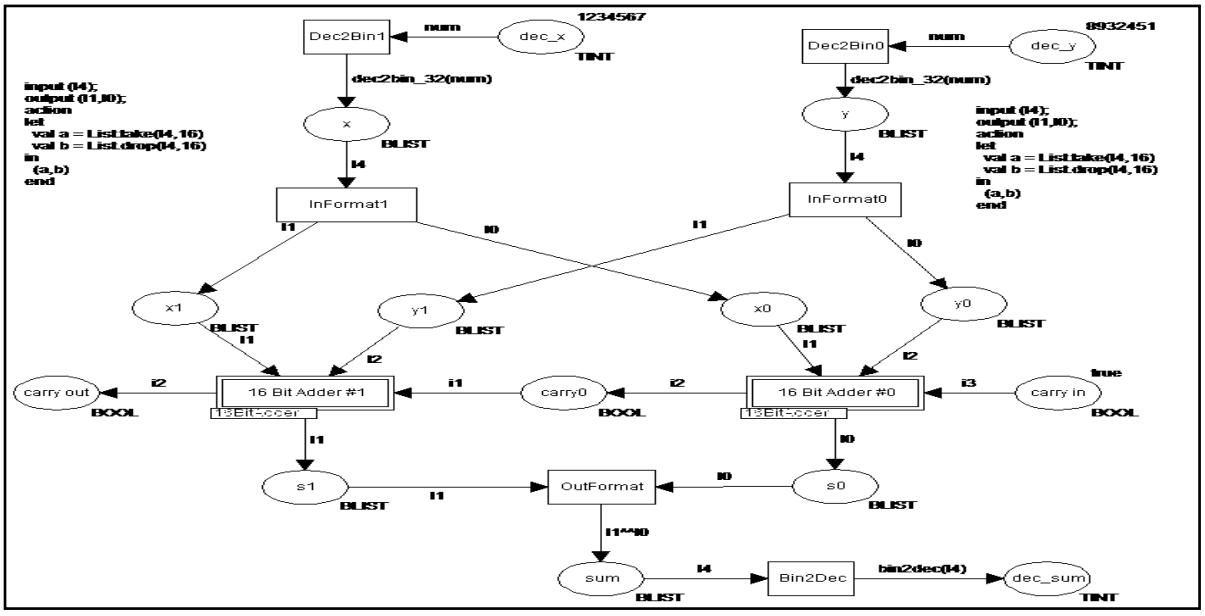


Fig. 6. CPN model of a 32-bit Adder

IV Related Work

Reconfigurable architectures, including nanocomputing, are likely to provide a better fit and allow continued performance improvements for general-purpose computation. [5,6,10,20]. Design space exploration as applied to compilation for FPGA-based and other reconfigurable architectures demonstrates the performance improvements possible with customized architectures [21-23]. The concept of "program in, chip out" (PICO) relies on compiler analysis, particularly targeted automatic parallelism, to identify program fragments that will benefit most from customized hardware [21]. While PICO targets primarily embedded processors and uses design space exploration, our approach envisions a desktop or embedded computer that reconfigures its own hardware to any arbitrary configuration, using machine requirements analysis and nanotechnology. This process may resemble FPGA reconfigurability, or physical molecular reassembly, performed at compile-time or run-time.

An innovative approach to large-scale, homogeneous, undifferentiated, reconfigurable architecture improves upon FPGAs using a less expensive nanoscale cell matrix approach [6]. This work describes how networks of atomic-scale switches can be configured in parallel and used to fabricate scalable processors that are customized to specific tasks. Our research can be targeted to a nanocomputer cell matrix architecture, focusing on use of the compiler to automatically generate reconfiguration instructions.

Although recent advancements in nanotechnology center on medicine, pharmaceuticals, chemistry, physics and computer engineering [12], most are early-generation nanotechnology. Our research pursues basic and applied nanocompiler research, focusing on applied molecular manufacturing techniques. Continued research in this area in the near future will require more powerful modeling and simulation tools.

Although CPN provides a flexible framework for modeling, true molecular models that more accurately depict the underlying physics and enable large-scale simulation are needed, though they will be computationally expensive. A variety of tools are available for molecular modeling and simulation (i.e., nanohub.org), with a promising recent tool being Nanorex's NanoEngineer (i.e., nanorex.com).

V Conclusion and Future Work

Reconfigurable computing is a rapidly advancing area, and the early promise of nanotechnology is being recognized. In this paper, we have proposed the design of a nanocompiler framework and demonstrated how molecular computing components can be generated from a higher-level representation, such as source code. Although this research is limited to simulation of a nano-mechanical adder component, the same approach can be used to generate the full range of components needed to construct an arbitrarily complex nano-mechanical multi-processor. With the significant flexibility and capability of nanocomputers, it is likely that the responsibility for guiding the configuration will fall to the compiler, and this research demonstrates the feasibility of that approach. Rather than the compiler customizing the program to suit the machine as in traditional compilation, the compiler may customize the machine to suit the program, extending code generation to include the ability to reconfigure the processor or guide its design and fabrication.

We are conducting extensive research and experimentation in the area of nanocompiler design and molecular nano-mechanical machine generation, including program characterization analysis, automatic parallelization, and molecular modeling. The long-term goal of this research is compiler control of physical, molecular assembly hardware

such as STMs and their descendants. Early work with an 8-bit adder model that uses a carry-lookahead approach [24] indicates that practical performance improvements are possible. Models for other components, including logic units, multipliers, and memory, are being developed using Colored Petri Nets for to conceptualize design and behavior and NanoEngineer to visualize the structure and physics of these nanoscale components.

VI References

1) Gibbs, W. (2004) A Split at the Core, Scientific American, 291, 96-101.
2) Hennessy, J. L. and Patterson, D. A. (2006) Computer architecture: a quantitative approach, fourth edition, San Francisco: Morgan Kaufmann Publishers.
3) Held, J., Bautista, J. and Koehl, S. (2006) From a few cores to many: a tera-scale computing research overview, Intel white paper.
4) Drexler, K. E. (1992) Nanosystems: molecular machinery, manufacturing and computation, New York: John Wiley & Sons, Inc.
5) Beckett, P. and Jennings, A. (2002) Towards nanocomputer architecture, Proceedings of the Seventh Asia-Pacific Conference on Computer Systems Architecture, Melbourne, Australia, Sept. 2002, 19, Darlinghurst, Australia: Australian Computer Society, 141-150.
6) Durbeck, L. J. K. and Macias, N. J. (2001) The Cell Matrix: an architecture for nanocomputing, Nanotechnology, 12, 217-230.
7) Goser, K., Glosekotter, P. and Dienstuhl, J. (2004) Nanoelectronics and nanosystems: from Transistors to molecular and quantum devices, Berlin: Springer-Verlag.
8) Merkle, R. C. (1996) Design considerations for an assembler, Nanotechnology, 7, 210-215.
9) Aho, A., Lam, M., Sethi, R. and Ullman, J. D. (2007) Compilers: Principles, Techniques, and Tools, New York: Addison-Wesley.
10) Compton, C. and Hauck, S. (2002) Reconfigurable computing: a survey of systems and software, ACM Computing Surveys, 34, 171-210.
11) Roco, M. C. (2007) National nanotechnology initiative: past, present and future Handbook on Nanoscience, Engineering and Technology, second edition, London: Taylor and Francis.
12) Hall, J. S. (2005) Nanofuture: what's next for nanotechnology, Amherst, New York: Prometheus Books.
13) Cohen, M. L. (2001) Nanotubes, nanoscience and nanotechnology, Materials Science and Engineering, 15, 1-11.
14) Merkle, R. C. (1993) Molecular manufacturing: adding positional control to chemical synthesis, Chemical Design Automation News, 8:9-10, 55-61.
15) Merkle, R. C. (1991) Computational nanotechnology, Nanotechnology, 2, 134-141.
16) Merkle, R. C. and Freitas, R. A. (2003) Theoretical analysis of a carbon-carbon dimer placement tool for diamond mechanosynthesis, Journal of Nanoscience and Nanotechnology, 3, 319-324.
17) Wagner, B. W. and Way, T. P. (2006) MolML: an abstract scripting language for assembly of mechanical nanocomputer architectures, International Conference on Computing in Nanotechnology, Las Vegas, June 2006, 258-264.
18) Jensen, K. (1997) Coloured petri nets: basic concepts, analysis methods and practical use, EATCS Monographs on Theoretical Computer Science, Berlin: Springer-Verlag.
19) Way, T. P. (2006) Compilation for future nanocomputer architectures nanocomputer architectures, International Conference on Computing in Nanotechnology, Las Vegas, June 2006, 251-257.
20) Carrillo, J. E. and Chow, P. (2001) The effect of reconfigurable units in superscalar processors, Proceedings of the 2001 ACM/SIGDA Ninth international Symposium on Field Programmable Gate Arrays, Monterey, California, Feb. 2001, New York: ACM Press, 141-150.
21) Kathail, V., Aditya, S., Schreiber, R., Rau, B. R., Cronquist, D. C. and Sivaraman, M. (2002) PICO: automatically designing custom computers, IEEE Computer, 35:9, 39-47.
22) Sekar, K., Lahiri, K. and Dey, S. (2003) Dynamic platform management for configurable platform-based system-on-chips, Proceedings of the International Conference on Computer Aided Design (ICCAD 2003), 641-648.
23) So, B., Hall, M. and Diaz, P. (2002) A Compiler approach to fast hardware design space exploration in FPGA-based systems, ACM SIGPLAN Notices, 37:5, 165-176.
24) Cheng, F. C., Unger, S. H. and Theobald, M. (2000) Self-timed carry-lookahead adders, IEEE Transactions on Computers, 49, 659-672.

Performance Evaluation of New-Fangled Digital Watermarking Techniques

Ajay Goel[1] , O.P.Sahu[2] ,
Sheifali Gupta [3] , Rupesh Gupta[4]
goelajay1@gmail.com, ops_nitk@yahoo.co.in,
sheifali@yahoo.com, rup_esh100@yahoo.co.in

GJCST Computing Classification
I.4.m, B.8.2 & D.4.6

***Abstract*- The use of digital formatted data offers several advantages over analog media such as high quality, easy editing, or high fidelity copying. The development of multimedia services and environments and thus the ease by which digital media can be duplicated and distributed, requires new concepts to support the protection of the media during the production and the distribution. In this work, I intend to disseminate the general concept of digital watermarking. I will illustrate the properties of a digital watermark and a description of methods used to insert watermarks in media and then move on to a discussion of what requirements a watermarking system must meet, as well as new-fangled methods are discussed for evaluating the strengths of various algorithms. This paper will focus almost exclusively on the watermarking of digital images; however most of these same ideas could easily be applied to the watermarking of digital video and audio [4] [5].**

Keywords- Watermarking Techniques, Least Significant Bit Modification, Correlation-Based Techniques, Stenography.

I INTRODUCTION

Unlike analog media that are becoming obsolete by now, digital media can be stored, duplicated, and distributed easily and with no lost of fidelity. It is clear that documents in digital form present a lot of advantages, but they also create problems, for parties who wish to prevent unauthorized reproduction and distribution of valuable digital medium (copyrighted, commercial, sensitive, secret documents…). Encryptions technologies can be used to prevent unauthorized access to the digital document - protect the content during the transmission of the file, but once it is received and decrypted, the document is no longer protected and is in clear. As a complement to encryption and/or copy protection, digital watermarking has been proposed as "last line of defense" against document misuse. Digital watermarking describes the process of embedding additional information into a digital media, without compromising the media's value. Hiding this piece of information to anybody besides a special designed detector is achieved by using a special watermarking technique called steganography. The watermark is then hidden in such a way that it may be imperceptible to a human observer, but easily detected by a computer. The development of watermarking methods involves several design tradeoffs due to the properties that a digital watermark should fulfill, and depending on the application field [1]. The framework, requirements, properties, evaluation and the design constraints will be discussed more throughout this paper.

A. *Digital Watermarking*

The Framework That watermark can be detected or extracted later to make an assertion about the media. For example, a very simple yet widely used digital watermarking technique[2] would be for images to add a visible seal on top of an existing image as shown as in the following figure 1:-

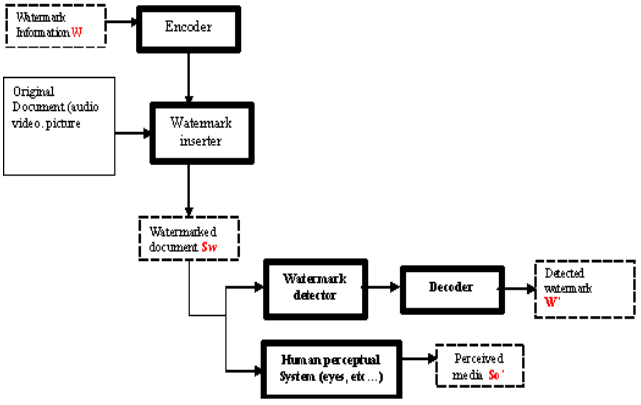


Figure-1: Generic Digital Watermarking Insertion & Recovery.

As illustrate in the above figure 1 the first step is to encode the watermark into a form that will be easily inserted with the media. The watermark inserter then combines the encoded representation of the watermark with the document. If the watermarking insertion process is design correctly, the result is media that appears identical to the original when perceived by a human, but which yields the encoded watermark information when processes by a watermark detector.

B. *Choice Of a Watermark-Object*

The most straight-forward approach would be to embed text strings into an image, allowing an image to directly carry information such as author, title, date…and so forth. The drawback however to this approach is that ASCII text in a way can be considered a form of LZW compression, which each letter being represented with a certain pattern of bits.

By compressing the watermark-object before insertion, robustness suffers [11].

Due to the nature of ASCII codes, a single bit error due to an attack can entirely change the meaning of that character, and thus the message. It would be quite easy for even a simple task such as JPEG compression to reduce a copyright string to a random collection of characters. Rather then characters, why not embed the information in an already highly redundant form, such as a raster image? Not only do images lend themselves to image watermarking applications, but the properties of the HVS can easily be exploited in recognition of a degraded watermark[3] [7].

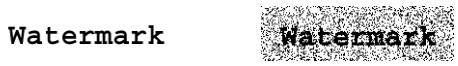


Figure 2 - Ideal Watermark-Object vs. Object with 25% Additive Gaussian Noise

II THE WATERMARKING TECHNIQUES

A. *Least Significant Bit Modification*

The most straight-forward method of watermark embedding would be to embed the watermark into the least-significant-bits of the cover object [6]. Given the extraordinarily high channel capacity of using the entire cover for transmission in this method, a smaller object may be embedded multiple times. Even if most of these are lost due to attacks, a single surviving watermark would be considered a success.

LSB substitution [12] however despite its simplicity brings a host of drawbacks. Although it may survive transformations such as cropping, any addition of noise or loosy compression is likely to defeat the watermark. LSB embedding is one of algorithm that uses spatial domain. When LSB is applied in the spatial or temporal domains, these approaches modify the Least Significant Bits (LSB) of the host data. The invisibility of the watermark is achieved on the assumption that the LSB data are visually insignificant [13]. Furthermore, once the algorithm is discovered, the embedded watermark could be easily modified by an intermediate party. LSB modification proves to be a simple and fairly powerful tool for stenography, however lacks the basic robustness that watermarking applications require.

B. *Correlation-Based Techniques*

Another technique for watermark embedding is to exploit the correlation properties of additive pseudo-random noise patterns as applied to an image [9]. A pseudo-random noise (PN) pattern W(x,y) is added to the cover image I(x,y), according to the equation shown as below in equation 1.

$$I_w(x,y) = I(x,y) + k * W(x,y) \quad \ldots\ldots (1)$$

In equation 1, k denotes a gain factor, and IW the resulting watermarked image. Increasing k increases the robustness of the watermark at the expense of the quality of the watermarked image. To retrieve the watermark, the same pseudo-random noise generator algorithm is seeded with the same key, and the correlation between the noise pattern and possibly watermarked image computed. If the correlation exceeds a certain threshold T, the watermark is detected, and a single bit is set. This method can easily be extended to a multiple-bit watermark by dividing the image up into blocks, and performing the above procedure independently on each block [10].

This basic algorithm can be improved in a number of ways. First, the notion of a threshold being used for determining a logical "1" or "0" can be eliminated by using two separate pseudo-random noise patterns. One pattern is designated a logical "1" and the other a "0". The above procedure is then performed once for each pattern, and the pattern with the higher resulting correlation is used. This increases the probability of a correct detection, even after the image has been subject to attack [9].

We can further improve the method by pre-filtering the image before applying the watermark. If we can reduce the correlation between the cover image and the PN sequence, we can increase the immunity of the watermark to additional noise. By applying the edge enhancement filter shown below in figure 3, the robustness of the watermark can be improved with no loss of capacity and very little reduction of image quality [9].

$$F_{edge} = \frac{1}{2}\begin{vmatrix} -1 & -1 & -1 \\ -1 & 10 & -1 \\ -1 & -1 & -1 \end{vmatrix}$$

Figure 3 - FIR Edge Enhancement Pre-Filter [9]

Rather than determining the values of the watermark from "blocks" in the spatial domain, we can employ CDMA spread-spectrum techniques to scatter each of the bits randomly throughout the cover image, increasing capacity and improving resistance to cropping. The watermark is first formatted as a long string rather than a 2D image. For each value of the watermark, a PN sequence is generated using an independent seed. These seeds could either be stored, or themselves generated through PN methods. The summation of all of these PN sequences represents the watermark, which is then scaled and added to the cover image [9].

To detect the watermark, each seed is used to generate its PN sequence, which is then correlated with the entire image [13]. If the correlation is high, that bit in the watermark is set to "1", otherwise a "0". The process is then repeated for all the values of the watermark. CDMA improves on the robustness of the watermark significantly, but requires several orders more of calculation [14]

III RESULTS

In general, algorithms were implemented in the most straightforward way, not the most computationally optimal. Three different watermarks were used, based on the theoretical and experimental information capacity of the watermarking algorithm, as shown in figures 4 & 5

Figure 4 - Small Watermark (12 x 9 pixels)

Figure 5 - Normal Watermark (50 x 20 pixels)

For our reference image, the ever-popular r (Lena) image is used, as shown below in figure 6.

Figure 6 - Lena Reference Image (512 x 512 Pixels)

Results from LSB substitution were as expected. The watermarked image shows little not noticeable degradation, while the large watermark was recovered perfectly.

Least Significant Bit Substitution

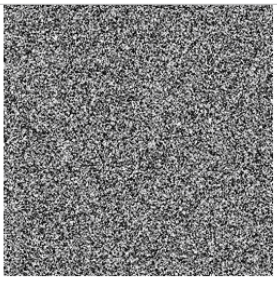


Figure 7(a) – Recovered Watermark after addition of 1% Gaussian Noise

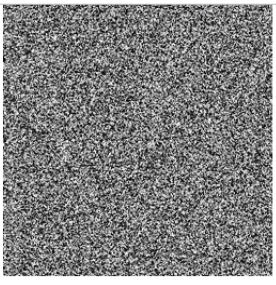


Figure 7(b) - Recovered Watermark after JPEG Compression with Quality 95

Least Significant Bit Substitution

Figure 8(a) - Watermarked Image PSNR= 102 dB

Figure 8(b) - Recovered Watermark

Least Significant Bit Substitution

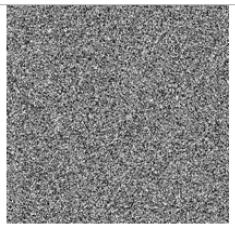


Figure 9(a) – Recovered Watermark after addition of 1% Gaussian Noise

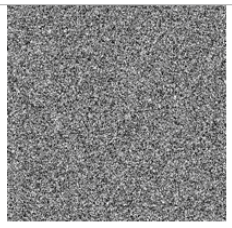


Figure 9(b) - Recovered Watermark after JPEG Compression with Quality 95

Although the watermark was recovered perfectly in the ideal case, the addition of any amount of noise, or compression of the image using JPEG fully destroys the embedded watermark, leaving nothing but noise. Even worse, the watermark can be removed with no perceivable change to the watermarked image. The message capacity of LSB embedding however is quite good, a 1:1 correlation with the size of the image.

The results of threshold-based correlation showed a vast improvement over LSB substitution in terms of robustness. Several parameters however must be discussed before we move on to results of this technique. A gain factor of $k = 5$ was chosen experimentally, however larger factors might be used for increased robustness at the expense of visual quality [8].Another issue with threshold-based techniques is the choosing of a suitable threshold for detection. Use of a smaller watermark will allow larger blocks to be used, increasing the strength of correlation and thus system robustness. Using the normal sized watermark, the largest possible block size {8,16,32…} is determined by:

$1000 \le \frac{512 * 512}{16^2}$, for a maximum block size of 16.

Threshold-Based Correlation, K = 5 BlockSize=16

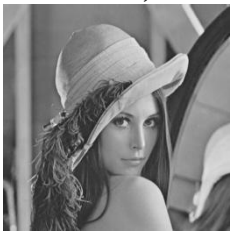


Figure 10a - Watermarked Image

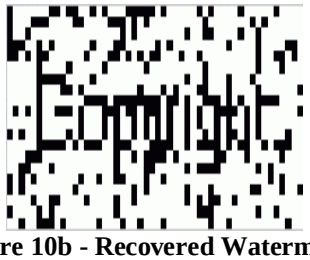


Figure 10b - Recovered Watermark

The outcome PSNR =20 As shown, the addition of the watermark is context-sensitive, to make it harder to remove. CDMA improves on the robustness of the watermark significantly, but requires several orders more of calculation.

IV Conclusion

A digital watermarking scheme is presented with the properties of digital watermark and requirement of a watermarking system. We also illustrate to insert different watermark into different image. We analysis and apply for the three different watermark to embedded directly into the LSB of channels of the pixel based on intensity of pixel, without using additional transformations and in correlation based technique with additive pseudo-random noise patterns. LSB substitution is not a very good candidate for digital watermarking due to its lack of even a minimal level of robustness. LSB embedded watermarks can easily be removed using techniques that do not visually degrade the image to the point of being noticeable. Furthermore if one of the more trivial embedding algorithms is used, the encoded message can be easily recovered and even altered by a third party. It would appear that LSB will remain in the domain of steganography due to its tremendous information capacity and Correlation-Based Techniques improves on the robustness of the watermark appreciably with more mathematical operation and additional computational time.

V References

1) S.Craver, N. Memon B.-L. Yeo, and M. Yeung, "Resolving rightful ownerships with invisible watermarking techniques: Limitations, attacks and implications," IEEE Trans. On Selected Areas of Communications, vol. 16, no. 4, pp. 573–586, 1998.
2) Stephen Wolthusen "On the limitations of digital watermarks". Fraunhofer-Institute for Computer Graphics 64283 Darmstadt, Germany.
3) Gustavus, J. Simmons "The prisoner's problem and the subliminal channel". 51-67
4) R. Walker, "Audio watermarking". BBC Research & Development, White Paper WHP 057
5) Raymond B Wolfgang, Christine l. Podilchuk, Edward J. Delp, "Perceptual Watermarks for Digital Images and Video".
6) J.J.K.Ó Ruanaidh, W.J. Dowling G.M. Boland, "Watermarking Digital Images for Copyright Protection".
7) Jonathan K. Su, Frank Hartung, Bernd Girod "Digital Watermarking of Text Image and Video Document." Telecommunications Laboratory; Universität of Erlangen-Nuremberg, Germany
8) R. G.Van Schyndel, A. Z. Tirkel, N. Mee, C. F. Osborne "A Digital Watermark, Processing of the IEEE International Conference on Image Processing." November 1994 Austin, Texas, vol. 2, pp. 86-90.
9) Langelaar, I. Setyawan, R.L. Lagendijk, "Watermarking Digital Image and Video Data", in IEEE Signal Processing Magazine, Vol 17, pp 20-43, September 2000
10) Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "A secure, robust watermark for multimedia," in Information Hiding: First Int. Workshop Proc. (R. Anderson, ed.), vol. 1174 of Lecture Notes in Computer Science, pp. 185–206, Springer-Verlag, 1996.
11) Koch, J. Rindfrey, and J. Zhao, "Copyright protection for multimedia data," in Proc. of the Int. Conf. on Digital Media and Electronic Publishing, 1994.
12) W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," IBM Systems Journal, vol. 35, no. 3/4, pp. 313–336, 1996.
13) Fabien A. P. Petitcolas, Ross J. Anderson and Markus G. Kuhn "Information hiding-a survey. Proceedings of the IEEE", July 1999 Special Issue "Identification and protection of multimedia information".
14) F.A.P. Petitcolas, "Watermarking Schemes Evaluation" ", in IEEE Signal Processing Magazine, Vol 17, pp 58-64, September 2000

An Efficient Hierarchical Certificate Based Binding Update Protocol for Route Optimization in Mobile IPv6

D.Kavitha *, K.E.Sreenivasa Murthy,
B.Sathyanarayana , V.Raghunatha Reddy,
S.Zahoor ul HuqS.K.University , Anantapur.
dwaramkavithareddy@gmail.com

GJCST Computing Classification
C.2.2, C.2.1 & D.4.6

***Abstract*- Most of the proposed security protocols for routing optimization in Mobile IPv6, including the one in the standard, depend on the special relationship between the home network and the mobile node. In this paper, we present a new protocol that does not depend on the security relationship between the home network and the mobile. The security of the protocol is analyzed and its performance evaluation is given. The results of performance evaluation show that our protocol achieves strong security and at the same time requires minimal computational overhead compared to return routability procedure.**

Keywords- Route Optimization , Mobile IPv6, Binding Update , Certificate based Binding Update , authentication , Security

I INTRODUCTION

Mobile IPv6 (MIPv6) has been proposed as an IP layer mobility protocol for the IPv6 [1]. The Mobile IPv6 standardization process started in 1995. The recent developments in public wireless networks convey that IPv6 nodes on a local link cannot necessarily trust each other anymore. They now become mutually suspicious, even when the nodes have completed an authentication exchange with the network. This had added a number of new security threats as well [3]. In addition, the lack of a global authentication infrastructure made it very hard to solve the problems with straightforward application of standard Internet security protocols, such as IPSec and IKE. For many of such reasons, the standardization of Mobile IPv6 was delayed.

Many schemes are introduced to provide security to Mobile IPv6.The method of route optimization is introduced to avoid triangle routing and which also removes some of the plain IPv6 vulnerabilities. For instance, the use of Secure Neighbor Discovery on the network where one of the end-points resides, removes some of the existing threats [4]. Yet, a security association alone is not sufficient to enhance security mechanism. General security associations typically do not show that a node owns a specific IP address, a property that is desired in the case of route optimization to authenticate home addresses. Certificate technology, for instance, usually does not track the correct IP address assignments of a large group of users. Also, the validity of care-of addresses cannot be ensured by a security Association alone. Either the security association must be accompanied by a trust relationship, or care-of addresses must be checked otherwise. This shows that enhancements to the security of route optimization are likely to employ Mobile IPv6 specific technology rather than general-purpose security tools.

In addition to above, many research problems that exist that can be divided as follows.

i. Most of the proposed security protocols for Mobile IPv6; including the one in the standard, depend on the special relationship between the home network and the mobile. It is a completely open question, what kind of security mechanisms would be needed if the home agent did not trust the information provided by the mobile.
ii. Finding care-of address verification mechanisms that employ lower layer assistance or SEcure Neighbour Discovery[4].
iii. Finding route optimization security mechanisms that do not require a reconfiguration of the shared secret between mobile node and its correspondent node.
iv. Developing adaptive out-of-band security mechanisms that are not specific to the deployment environment of a network operator.
v. Extending the developed mechanisms to full multiaddressing, i.e., including also multi-homing.

The rest of the paper is organized as follows. In the section(2) we focus the state-of-art of the existing routing optimization protocols and Section(3) emphasizes on the security of these protocols and their limitations are specified. Then in Section(4) we are presenting a protocol that is able to give a solution to a research problem that is stated in (a). Finally, we conclude our remarks in Section (5).

II SECURITY ANALYSIS OF RO PROTOCOLS

A. *Return Routabilty*

The secure RO[5] in the MIPv6 is composed of six messages and is shown in Figure 1. The first four messages are dedicated to checking the RR of the |RO protocol, and the last two messages are used to authenticate the message.

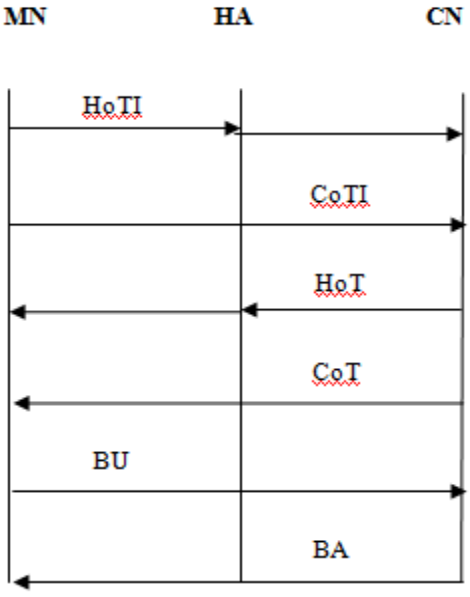


Figure 1: Return routability protocol

Figure.1 shows an illustration of secure routing optimization in MIPv6. The MN-HA path is securely protected by the IPSec tunnel.

The MN sends the Home Test Init () and the Care-of Test Init (CoTI) messages to initiate the binding update. These two messages are sent almost simultaneously but along different paths; the CoTI is sent directly to the CN , and the HoTI is sent indirectly via the HA. HoTI and CoTI messages are sent to trigger the test packets.

The MN sends the Home Test (HoT) and the Care-of Test (CoT) messages as responses to the previous messages. The HoT message is sent to the source address of HoTI while CoT message is sent to the source address of the CoTI. The HoT message consists of a home token which is calculated by taking the hash function over the concatenation of a secret key Kcn known only by the CN, the source address of the HoTI packet, and a nonce ni.The index i is also included in the HoT packet,to allow the CN to find the appropriate nonce. Similarly the CoT message consists of a care-of token which is calculated by taking the hash function over the concatenation of a secret key Kcn known only by the CN, CoA, and a nonce nj.

When MN receives both home token and care-of token it calculates a key Kbm by taking a hash function over the concatenation of home token and care-of token. By using secret key kbm it sends the actual binding update (BU) message to CN.After receiving BU message CN sends a binding acknowledgement (BA) message to MN.

B. Certificate Based Binding Update (Cbu) Protocol

In CBU[6]When MN wants to start RO operation with CN, it sends a RO request REQ = {HoA, CN, n0} to CN via reserved tunneling, where n0 is a nonce value used to match the reply message REP. Message REQ is sent to MN's HA via the IPSec protected secure tunnel.

Upon arriving at HA, REQ is intercepted by HA and it will not forward REQ to CN, instead, it creates a cookie C0 and sends COOKIE0 = {HoA, CN, C0} to CN.

In reply, CN creates a nonce n1 and a cookie C1, and sends COOKIE1 = {CN, HoA, C0, C1, n1} to MN.

After receiving COOKIE1, HA checks on the validity of C0 and replies CN with EXCH0 = {HoA, CN, C0, C1, n1, n2, g^x, TS, SIG_{HA}, $Cert_{HA}$}, where n2 is a freshly generated nonce, x a Diffie– Hellman (DH) secret value, SIG_{HA} = S_{HA}(HoA|CN|g^x|n1|n2|TS) , $Cert_{HA}$ = {HLSP, P_{HA}, Valid_Interval, SIG_{CA}} is the public key certificate of HA.

When CN receives EXCH0, it validates the cookies, the HA's public key certificate $Cert_{HA}$, the signature and importantly, checks for equality of the HA's subnet prefix strings embedded in both $Cert_{HA}$ and HoA. If all the validations and checking are positive, CN can be confident that the home address HoA of MN is authorized by its HA and the DH public value g^x is freshly generated by MN's HA. CN next generates its own secret value y and its public value g^y, and then computes the DH key K_{DH} = $(g^x)^y$, a session key K_{BU} = prf(K_{DH}, n1|n2) and a MAC MAC1 = prf(K_{BU}, g^y|EXCH0), and sends EXCH1 = {CN, HoA, C0, C1, g^y, MAC1} to MN.

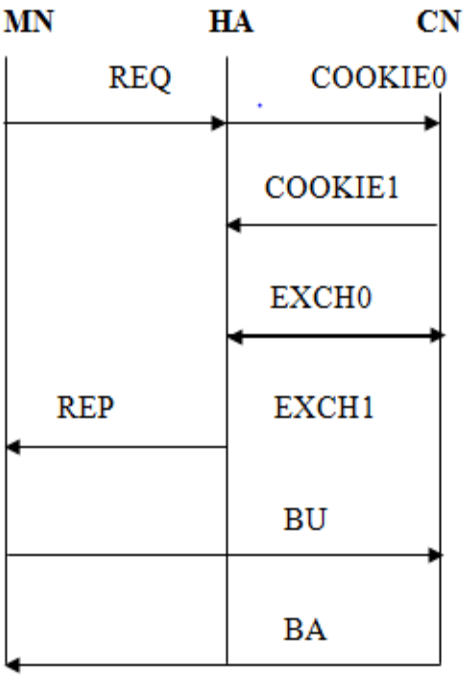


Fig. 2 CBU Protocol

Again, this message is intercepted by HA, which first validates the cookies, calculates the K_{DH} = $(g^y)^x$ and K_{BU} = prf(K_{DH}, n1|n2).

HA then computes MAC2 = prf(K_{BU}, EXCH1), and sends an optional CONFIRM = {HoA, CN, MAC2} to CN. The validity of MAC2 is checked by CN and if valid, CN creates a cache entry for HoA and the session key K_{BU}, which will be used for authenticating binding update messages from MN. Upon positive verification of MAC1, HA also sends REP = {CN, HoA, n0, K_{BU}} to MN through the secure IPSec ESP protected tunnel. After receiving REP, MN

checks that n0 is the same as the one it sent out in REQ. If so, MN proceeds to send CN binding update messages protected using K_{BU} as in the RR protocol. Once CN receives BU message it sends an acknowledgement to MN.

C. *Hierarchical Certificate Based Binding Update Protocol(Hcbu)*

In HCBU[7], When MN realizes an imminent handover , it first initializes a Binding Update request in Message 1.

Message 1. Binding Update Request (BUReq): {BU, Nm, HoA, CN}

where Nm is a fresh random nonce.

Message 2 passes the fresh nonce Nm, MN's HoA, CN's address and a DH public value g^x to CN.

Message 2. Pre-Information Exchange0 (EXCH0):

{Nm, HoA, CN, g^x}.

In reply, CN attaches its own fresh nonce N_C and DH public value g^y to the received Message 2 and thus forms Message 3.

Message 3. Pre-Information Exchange1 (EXCH1):

{Nm, Nc, HoA, CN, g^x, g^y , $Cookie_{CN}$}

where $Cookie_{CN}$ =prf(K_{CN}, Nm|Nc|HoA|CN|g^x|g^y)

CN next creates a cookie $Cookie_{CN}$ for HA using its own secret key K_{CN}.

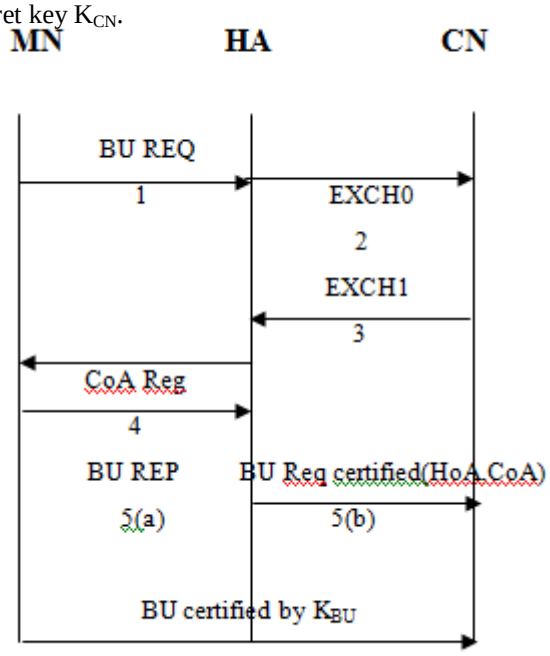


Fig. 3 HCBU Protocol

III ANALYSIS OF THE RO PROTOCOLS

The goal in designing IPv6 is to make MIPv6 at least as secure as static IPv6. But MIPv6 introduced some security vulnerabilities. Among which weak authentication and authorization of BUs is considered as the biggest vulnerability. These malicious Binding Updates open the door for many types of attacks like False Binding Update attack , Man – in-the-Middle Attack , Denial-of-Service Attack.

In the RR protocol , liveness test for the MN is done, but it is prone to false BU attack and session hijacking attack and Distributed DoS attacks as anytained to anyone who obtains Home Token and Care of token can create the session key Kbm. Moreover the path between CN and HA is not secured.To overcome these drawbacks , a variation of the RR protocol is recently proposed in

In the CBU protocol, The task of authenticating MN and its HoA is done by issuing individual certificates. But the protocol does not address certificate management issues for HA's. the prcess of certificate issuing is done directly to every individual home link subnet prefix by one CA which is not practical. This flat structure of trust management is not flexible and scalable. Another problem with the CBU protocol is that there is no way for CN to assure liveness of MN on its claimed CoA in the BU message.

The drawbacks of CBU protocol are overcome by HCBU protocol presented in the previous section. It uses a flexible and scalable 3-layer trust management framework for certificate management. Based on such a framework it is assured that both the mobile nodes Home address and Care of address are authenticated to CN. This is considered as a secured tunnel between a MN and HA. Hence in HCBU, computational costs on the protocol participants are reduced .and also latency of this protocol is fairly low. Communication efficiency is also achieved by using early binding update. Inspite of all these advantages there is an assumption that a secure tunnel is established between the MN and HA. To remove such an assumption we present a new protocol in the next section that can be considered as an enhancement to HCBU.

IV EFFICIENT HIERARCHICAL CERTIFICATE BASED BINDING UPDATE PROTOCOL

In HCBU, existence of a secure tunnel between a MN and HA is an assumption.. This assumption has two drawbacks. 1) In practice such a secure tunnel can be established by exchanging secret keys between HA and all other nodes in the home network. If the intruder performs eavesdropping on the packets that are used to exchange the secret keys, then the intruder also gets the secret key. 2) If at all such a secure tunnel is established between MN and HA, it should also exist between HA and CN. It is because even a CN can be a MN.

To overcome the above specified drawbacks, we proposed a protocol "Efficient Hierarchical Certificate based Binding Update Protocol" as an improvement to HCBU. In this protocol we remove the assumption of a secure tunnel.

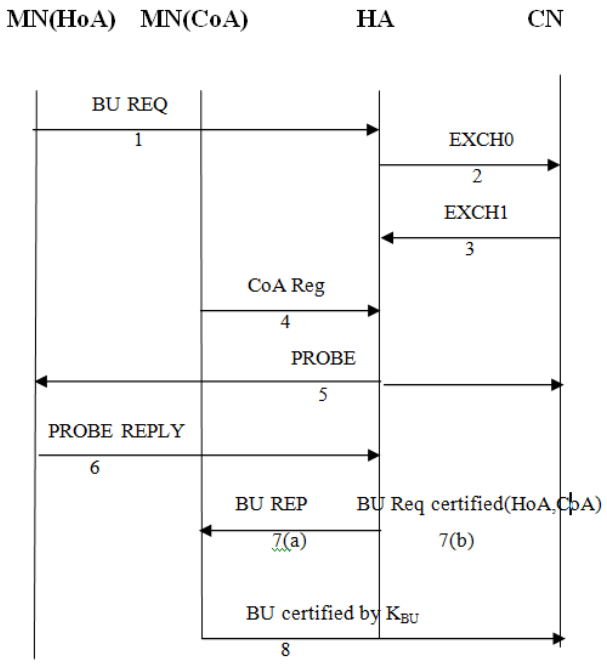


Fig. 4 . Efficient Hierarchical Certificate based Binding Update Protocol

The Proposed Scheme

The shortages in RR procedure and CBU protocols are widely discussed in current years that lead to the increased proportion of mobile users. Many papers have proposed improved mechanisms in terms of security enhancement, but very few of them are proposed with out taking secure tunnel into consideration.

Our proposed protocol EHCBU consists of 8 steps. First 3 steps are carried out when the MN discovers an imminent handover and the remaining after the handover.

Step 1

When the MN realizes that the handover is forthcoming, it sends a BU Request to HA.

BUReq : {BU , Nm , HoA , CN}

In this message MN sends its Home address , Correspondent node's address and a fresh random nonce Nm.

Step 2

HA exchanges pre information needed for Diffsie Hellman Key exchange in the message 2. HA constructs the packet

{Nm , HoA , CN , g^x}

Step 3

In reply CN also exchanges similar pre information with the HA.So it creates the packet as

{Nm , Nc , HoA , CN , g^x , g^y , $Cookie_{CN}$},

where $Cookie_{CN}$ = prf(K_{CN} , Nm | Nc | HoA | CN | g^x | g^y}

Step 4

Immediately after the hand over MN sends a CoA registration Request to HA.

CoAReq : {CoA , HoA , Valid_Interval , CN , $SIG_{HA'}$, $Cert\_Chain_{HA'}$}

HA checks the validity of the certificate chain and verifies the signature contained in the message. Negative result of either of them leads to the rejection of message

Step 5

HA sends a PROBE to MN's home address to determine if the MN is still in the HoA or not.

PROBE : {HoA , CN , Np}

Step 6

If it is not there, no reply packet is sent to the HA. If the MN is there it sends a PROBE CONFIRM to the HA.

PROBE CONFIRM : {HoA , MN , $PREV\_SIG_1$, $PREV\_SIG_2$, $PREV\_SIG_3$}

When the HA does not get a PROBE CONFIRM packet it may be because the MN is not there or the PROBE REQUEST packet may be lost. To handle the latter case it sends a PROBE REQUEST packet after a certain time out. This confirms the availability of MN in its Home address.

Step 7

Binding Update Request with Certified(HoA, CoA):

Where SIG_{HA}= S_{HA}(HoA|CoA|Valid_Interval|CN|g^{xy}|Nm|Nc)

At the same time, MN obtains the Binding Update key K_{BU} in Message 7(b) from HA and therefore, could send out the final Binding Update message.

Message 7(b). Binding Update Reply (BURep):

{HoA, CoA, CN, K_{BU}}.

Step 8:

MN sends the Binding Update message certified by K_{BU} to CN .

V SECURITY ANALYSIS OF EHCBU

We compared the computational expenses for the three protocols described in section 2 and the proposed protocol. We compared all the protocols by taking the uniform computational delay in each message.

In the figure 5, it is clearly shown that EHCBU has a little more computational delay compared to HCBU. However, the removal of an assumption of a secure tunnel is overcome at the cost of a small increase in computation delay.

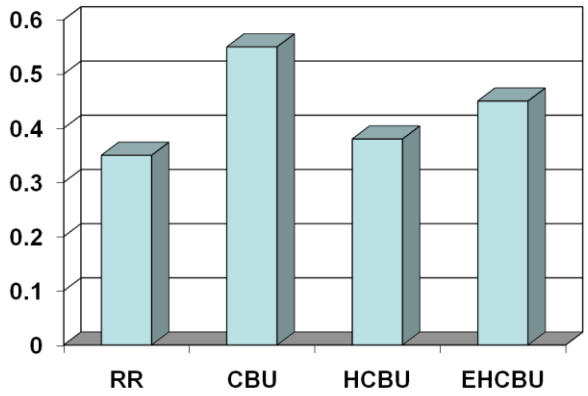


Fig. 5. Computational delay of the protocols

In addition to the superiority in performance compared to RR procedure, security protection is also assured by the proposed scheme.The table 1 shows that only EHCBU is not susceptible to false BU attack.

	RR	CBU	HCBU	EHCBU
False BU attack	No	No	Yes	Yes
False BU attack when intruder can grab the secret key	No	No	No	Yes

Table 1 : Security Analysis

VI References

1) Perkins C. Mobility support in IPv6, RFC 3775, IETF, June 2004.
2) Aura Tuomas. Designing the Mobile IPv6 security protocol. Annals of Telecommunications (Special issue on network and information systems security) March–April 2006;61(3–4)
3) Arkko Jari, Aura Tuomas, Kempf James. Securing IPv6 neighbor discovery and router discovery. In: Proc. 2002 ACM Workshop on Wireless Security (WiSe). Atlanta, GA, USA: ACM Press; September 2002. p. 77–86.
4) Arkko J, Voght C. A taxonomy and analysis of enhancements to mobile, Internet draft, October 2004.
5) P. Nikander, T. Aura, J. Arkko, G. Montenegro, Mobile IP version 6 Route Optimization Security Design Background,Expired IETF Internet Draft, 2003.
6) R. Deng, J. Zhou, F. Bao, Defending against redirect attacks in mobile IP, in: Proceedings of the of 9th ACM Conference on Computer and Communications Security (CCS), Washington, 2002, pp. 59–67.
7) K. Ren, W. Lou, K. Zeng, F. Bao, J. Zhou, and R. H. Deng, "Routing optimization security in mobile IPv6,"
8) Computer Networks, vol. 50, no. 13, pp. 2401–2419, 2006.
9) "A Secure and Lightweight Approach for Routing Optimization in Mobile IPv6 " Sehwa Song, Hyoung-Kee Choi, and Jung-Yoon Kim,EURASIP Journal on Wireless Communications and Networking
Volume 2009 (2009), Article ID 957690
10) "Security in Mobile IPv6: A survey" Khaled Elgoarany*, Mohamed Eltoweissy Published online 6 March 2007 Elsevier

Routing In Mobile Ad-Hoc Networks

V.V.S.Chandra Mouli[1],
Y.Nagi Reddy[2],S.Meena Kumari[3]
,N.Geetanjali[4]

GJCST Computing Classification
C.2.2, C.1.3 & C.2.6

***Abstract*- Y As mobile networking continues to experience increasing popularity, the need to connect large numbers of wireless devices will become more prevalent A mobile ad hoc network (MANET), is a self-configuring network of mobile devices connected by wireless links. Each device in a MANET is free to move independently in any direction, and will therefore change its links to other devices frequently. Each must forward traffic unrelated to its own use, and therefore be a router. The primary challenge in building a MANET is equipping each device to continuously maintain the information required to properly route traffic. Such networks may operate by themselves or may be connected to the larger Internet. MANETs are a kind of wireless ad hoc networks that usually has a routable networking environment on top of a Link Layer ad hoc network.**

Keywords- Manets, routing, wireless,protocol.

I INTRODUCTION

The growth of laptops and 802.11/Wi-Fi wireless networking have made MANETs a popular research topic since the mid- to late 1990s. Many academic papers evaluate protocols and abilities assuming varying degrees of mobility within a bounded space, usually with all nodes within a few hops of each other and usually with nodes sending data at a constant rate. Different protocols are then evaluated based on the packet drop rate, the overhead introduced by the routing protocol, and other measures. In ad hoc routing protocol is a convention, or standard, that controls how nodes decide which way to route packets between computing devices in a mobile ad-hoc network . In ad hoc networks, nodes do not start out familiar with the topology of their networks; instead, they have to discover it. The basic idea is that a new node may announce its presence and should listen for announcements broadcast by its neighbours. Each node learns about nodes nearby and how to reach them, and may announce that it, too, can reach them. Note that in a wider sense, ad-hoc protocol can also be used literally, that is, to mean an improvised and often impromptu protocol established for a specific purpose In our paper we concentrate on different routing protocols which are extensively used in MANETS.

1st V.V.S Chandra Mouli , 2nd Y.Nagi Reddy,3rd S.Meena Kumari, Assistant Professor, G.Pulla Reddy Engineering College,Kurnool,Andhra Pradesh,India.
(email:yerramnagireddy@gmail.com)
4th N.Geethanjali,Associate Professor,Department Of CST,SK University,Ananthapur,AP,India.

Advantages

DSDV was one of the early algorithms available. It is quite suitable for creating ad hoc networks with small number of nodes. Since no formal specification of this algorithm is

We can list the routing protocols as follows:

i. Pro-active (table-driven) routing
ii. Reactive (on-demand) routing
iii. Flow-oriented routing
iv. Adaptive (situation-aware) routing
v. Hybrid (both pro-active and reactive) routing
vi. Hierarchical routing protocols
vii. Geographical routing protocols
viii. Power-aware routing protocols

II PROACTIVE ROUTING

This type of protocols maintains fresh lists of destinations and their routes by periodically distributing routing tables throughout the network. The main disadvantages of such algorithms are:

Respective amount of data for maintenance.
Slow reaction on restructuring and failures.

Examples of pro-active routing are

Destination-Sequenced Distance-Vector Routing (DSDV) is a table-driven routing scheme for ad hoc mobile networks based on the Bellman-Ford algorithm. It was developed by C. Perkins and P.Bhagwat in 1994. The main contribution of the algorithm was to solve the Routing Loop problem. Each entry in the routing table contains a sequence number, the sequence numbers are generally even if a link is present; else, an odd number is used. The number is generated by the destination, and the emitter needs to send out the next update with this number. Routing information is distributed between nodes by sending full dumps infrequently and smaller incremental updates more frequently.

For example the routing table of Node A in this network is

Dest	Next Hop	No. of Hops	Seq Number	Install Time
A	A	0	A 46	001000
B	B	1	B 36	001200
C	B	2	C 28	001500

Selection of Route

If a router receives new information, then it uses the latest sequence number. If the sequence number is the same as the one already in the table, the route with the better metric is used. Stale entries are those entries that have not been updated for a while. Such entries as well as the routes using those nodes as next hops are deleted.

present there is no commercial implementation of this algorithm. Many improved forms of this algorithm have been suggested.

Disadvantages
DSDV requires a regular update of its routing tables, which uses up battery power and a small amount of bandwidth even when the network is idle.
Influence
While DSDV itself does not appear to be much used today[citation needed], other protocols have used similar techniques. The best-known sequenced distance vector protocol is AODV, which, by virtue of being a reactive protocol, can use simpler sequencing heuristics. Babel is an attempt at making DSDV more robust, more efficient and more widely applicable while staying within the framework of proactive protocols.

III REACTIVE ROUTING

This type of protocols finds a route on demand by flooding the network with Route Request packets. The main Disadvantages
High latency time in route finding. Excessive flooding can lead to network clogging.
Examples are AODV,DSR

IV FLOW BASED ROUTING

This type of protocols finds a route on demand by following present flows. One option is to unicast consecutively when forwarding data while promoting a new link.
Disadvantages
Takes long time when exploring new routes without a prior knowledge. May refer to entitative existing traffic to compensate for missing knowledge on routes.
Examples are LBR- LINK LIFE BASED ROUTING

V ADAPTIVE ROUTING

This type of protocols combines the advantages of proactive and of reactive routing. The routing is initially established with some proactively prospected routes and then serves the demand from additionally activated nodes through reactive flooding. Some metrics must support the choice of reaction.
Disadvantages
It depends on amount of nodes activated
Reaction to traffic demand depends on gradient of traffic volume.
Examples are TORA (Temporally-Ordered Routing Algorithm) (Temporally-Ordered Routing Algorithm), LRR(Link Reversal Routing)

VI HYBRID ROUTING

This type of protocols combines the advantages of proactive and of reactive routing. The routing is initially established with some proactively prospected routes and then serves the demand from additionally activated nodes through reactive flooding. The choice for one or the other method requires predetermination for typical cases.
Disadvantage
Reaction to traffic demand depends on gradient of traffic volume.
Examples are ARPAM, HRPLS, HSLS, OORP, ZRP.

VII HIERARCHICAL ROUTING

This type of protocols the choice of proactive and of reactive routing depends on the hierarchic level where a node resides. The routing is initially established with some proactively prospected routes and then serves the demand from additionally activated nodes through reactive flooding on the lower levels. The choice for one or the other method requires proper attributation for respective levels.
Disadvantages
Advantage depends on depth of nesting and addressing scheme.
Reaction to traffic demand depends on meshing parameters.
Examples are CBRP, CEDAR, DART, DDR, HSR.

VIII GEOGRAPHICAL ROUTING

This type of protocols acknowledges the influence of physical distances and distribution of nodes to areas as significant to network performance.
Disadvantages
Efficiency depends on balancing the geographic distribution versus occurrence of traffic.
Any dependence of performance with traffic load thwarting the negligence of distance may occur in overload.
Examples are ALARM, BGR, DREAM, LAR.

IX POWER-AWARE ROUTING

Energy required to transmit a signal is approximately proportional to d^{α}, where d is the distance and is the attenuation factor or path loss exponent, which depends on the transmission medium. When $\alpha = 2$ (which is the optimal case), transmitting a signal half the distance requires one fourth of the energy and if there is a node in the middle willing spend another fourth of its energy for the second half, data would be transmitted for half of the energy than through a direct transmission - a fact that follows directly from the inverse square law of physics.
Disadvantages
This method induces a delay for each transmission.
No relevance for energy network powered transmission operated via sufficient repeater infrastructure.
Examples are ISAIAH, PARO, EADSR, DSRPA.

X CONCLUSION

This is a list of existing definitions or even implementations of Ad hoc network routing protocols.In the above paper we tried to compare and contrast different routing protocols in MANETS. In future, we would like to study the performance of ADV and the on-demand protocols for real-

time traffic. With ADV providing lower latencies it should be a more suitable protocol for real-time traffic scenarios. It will be also interesting to investigate the effect of ADV and the on-demand protocols on TCP performance. As routes are frequently refreshed using updates in ADV, it helps maintain route connectivity all the time as required in TCP.

XI References

1) Edoardo Biagioni and Kent Bridges. The application of remote sensor technology to assist the recovery of rare and endangered species. International Journal of High Performance Computing Applications, 16(3), (2002).
2) Edoardo Biagioni and Galen Sasaki. Wireless sensor placement for reliable and efficient data collection. In Hawaii International Conference on Systems Sciences, Waikoloa, Hawaii, Jan (2003).
3) Josh Broch, David A. Maltz, David B. Johnson, and Yih-Chun Hu and Jorjeta Jetcheva. A performance comparison of multi-hop wireless ad hoc network routing protocols. In Mobile Computing and Networking, pages 85–97, (1998). IEEE Computer Society LAN/MAN Standards Committee
4) Wireless LAN medium access control (MAC) and physical layer (PHY) specification." IEEE Std. 802.11-1997. IEEE, New York, NY (1997).
5) S. R. Das, C. E. Perkins, and E.M. Royer, "Performance comparison of two on-demand routing protocols for ad hoc networks," in IEEE Infocom (2000),Mar (2000).
6) Johnson et al., The dynamic source routing protocol for mobile adhoc Networks IETF Internet Draft. http://www.ietf.org/internet-drafts/draftietf-manet-dsr-02.txt, (1999).
7) S.-J. Lee, E. M. Belding-Royer, and C. E. Perkins, Scalability Study of the Ad Hoc On-Demand Distance-Vector Routing Protocol, International Journal of Network Management, vol. 13, no. 2, March (2003).
8) S.Lee, Mario Gerla, and C.K.Toh, A simulation study of table-driven and on-demand routing protocols for mobile ad hoc networks in IEEE Network Magazine, Aug(1999).
9) E.Royer, and C.K.Toh, "A review of current routing protocols for ad hoc mobile wireless networks," in IEEE Personal Communications Magazine, pp.46-55, Apr(1999).
10) Cheng, R. Riley, and S. P. R. Kumar, A loop-free extended Bellman-Ford routing protocol without bouncing effect in ACM SIGCOMM '89, pp. 224–236, (1989).

Risk Identification and Preemptive Scheduling In Software Development Life Cycle

GJCST Computing Classification
D.2.6, D.2.2 & K.6.1

Basit Shahzad, Abdullah S. Al-Mudimigh, Zahid Ullah
College of Computer & Information Science
King Saud University, Riyadh, Saudi Arabia
Basit.shahzad@gmail.com , mudimigh@ksu.edu.sa, zahid@ksu.edu.sa

***Abstract*- Software development has emerged as a disciplined discipline and the use of process models to develop the software has increased over time. Although the software industry is blessed with quite a few tool driven approaches, and the usage of technology is increasing yet the amount of risks faced by the software development life cycle have also increased to an extent. This paper focuses on the avoidance and mitigation strategies for the already identified and prioritized risk factors.**

Keywords- Preemptive risk identification, software risk handling, risk mitigation

I INTRODUCTION

Software risk management has been a very hot area of research since last three decades. Recently, the research community looks seriously interested to identify not only the risk factors but also the causes of the appearance of the risk factors in software development life cycle and how these risks can either be handled or avoided. A recent survey of 600 firms indicated that almost 35% of them had at least one 'runaway' software project [1]. In another study, conducted on almost 13,000 projects, it was investigated that almost 25% of the projects were either delayed or faced a failure. It has been observed that most problems in the software industry are faced just because of the poor software risk handling mechanisms or due to the absence of any such mechanism at all. In this regard it is important to note that currently strong emphasis is being given on this domain to identify more and more risk factors. Pressman [3] has made an effort to identify the software risks, and has provided the ten broader risk factors. Bohme, in his work has also provided a list of top ten risk categories[4]. Basit Shahzad, [5] has also worked in this domain to identify a relatively more detailed list of software risk factors and also identifying the relative impact of each risk factors. In a recent paper on risk management, the risk factors have been prioritized according to their frequency of occurrence and the impact that they possess [6], and thus a list of eighteen risk factors with respect to their total impact has been prepared. The list is presented in the table 1 and table 2. Table 1 presents the list of all 18 risk factors, while table 2 presents the ordered list of software risk factors w.r.t. the overall impact of each risk factor.

Risk #	Impact	Probability	Overall Impact	Risk #	Impact	Probability	Overall Impact
2	519.5	50%	259.75	1	55.2	50%	27.6
15	191.4	50%	95.7	6	208.5	10%	20.1
11	185	40%	74	16	195.6	10%	19.6
13	348.5	20%	69.7	10	61.3	30%	18.39
18	208.5	30%	62.55	9	145.6	10%	14.6
17	83.5	50%	41.75	8	56	10%	5.6
5	136	30%	40.8	12	15	20%	3
4	114.8	30%	34.44	7	6.6	30%	1.98
14	292.5	10%	29.3	3	3.6	30%	1.08

Table 1: The risk factors w.r.t their identifier

The risk factor identified in this list is expected to cover a border range of the risks that may come into the software development process. Still the author feel himself restricted, not to claim that this list covers all possible risk factors. It is strongly believed that the risk identification, particularly, is an ongoing process, and apparently there is no full stop as the risk factors keep on increasing with the arrival of new technologies, people, environment, management and the circumstances. So a claim about the identification of all risk factors available in the entire software process, may not be realistic. Table 2, presents the ordered list of available risk factors, by calculating the overall impact and frequency of each available risk factor [6].

In table 2, the term "Impact" means the impact of that specific risk factor, e.g. the risk factor number 2 has the impact of 519.5 and risk factor number 7 has the impact of 6.6. The term "Probability" means the possible occurrence of the risk factor. The term "Overall Impact" describes the impact of a risk factor with respect to the probability of each factor. The risk factors have been ordered with respect to the overall impact they possess, in ascending order, showing the maximum overall impact of 259.75 for risk factor number 2 and minimum overall impact of 1.08 for risk factor number 3.

Risk #	Impact	Probability	Overall Impact	Risk #	Impact	Probability	Overall Impact
2	519.5	50%	259.75	1	55.2	50%	27.6
15	191.4	50%	95.7	6	208.5	10%	20.1
11	185	40%	74	16	195.6	10%	19.6
13	348.5	20%	69.7	10	61.3	30%	18.39
18	208.5	30%	62.55	9	145.6	10%	14.6
17	83.5	50%	41.75	8	56	10%	5.6
5	136	30%	40.8	12	15	20%	3
4	114.8	30%	34.44	7	6.6	30%	1.98
14	292.5	10%	29.3	3	3.6	30%	1.08

Table 2: List of prioritized risk factors

II STATE OF THE ART

Software risk identification and mitigation has been a prime area of research since last two decades, and this area of research has received a highly overwhelming response and contribution from the researcher both: in industry and academia, world-wide. In order to identify the recent trend and practices in the domain of software risk identification a comprehensive literature survey was conducted that has helped in the more effective management of risk factors.

Danny Lieberman [31] has worked to reduce operational risks by improving the software quality. Dannay focuses on the classification and quantitative evaluation of removing the software risks by effective software management, thus contributing to the classified risk mitigation. In a study that was conducted in 2005 [32], a sample of 167 customer's data breaches were analyzed to view the distribution of risks and threats and it were identified that 3% of the total risks are caused by accidental disclosure bye-mails, 7.8% of risks are oriented due to the human weaknesses, 40.1% risks are caused by unprotected computer/backup media and 49.1% of risks are caused due to the malicious exploitation of software risks. Thus, suggesting way mitigates the risk factors more appropriately.

The SEI reports that 90% of all software risks are due to already known defects [33], while all of the SANA top 20 internet security problems are result of poor coding, testing and sloppy software engineering.[34]

Jhon Stiuby (2009) and his team have worked on the management of risks in distributed software projects, which proposes a framework for handling the software projects that are not developed at geographically same location, and have advised a framework to e followed in this regard [35].

B.J. Alge, C. Wiethoff, and H.J (2003) . Kelin have emphasized on the effective handling of risks and problems in the software development lifecycle and in team structure by the usage of knowledge building process and effective communication[36]. E. Bradner, G. Mark, and T.D. Hertal

(2005) have worked to identify the correct team sizes for the different project sizes and have focused the problems that are experienced by over, low and poor staffing [37].

R. N. Burn (2001) and his team have discussed the risks that are oriented due to the in-appropriate application selection methodology, specially in the database projects [38]. R. N. Charatte (1989), has proposed the analysis and management of the risk factors in software development process [39].

The surveyed literature has been identified greatly in the favors of categorical identification of the risk factors as the existence of risk factors can be extremely harmful, if not attended at the proper time by giving due consideration.

III HANDLING AND AVOIDANCE MECHANISM

Table 2, summarizes the Impact, Probability and Overall Impact of each risk factor. The aim to establish the prioritized list is to help the interested community to better handle the software risks, thus, the risk factor with the highest overall impact is proposed to be addressed first and with the highest attention, perhaps even leaving all other activities at hold. While the risk factors like factors number 7 or 3 require least attention, and can be given importance only when the ample staff is free to invest time on the management of these risks factors. After having established the prioritized list of risk factors based upon the overall impact it is necessary that the risks are either to be handled or avoided, it is necessary that a strategy is proposed for each risk factor. Sub-sections 2.1-2.14, discuss the handling and avoidance strategies against each risk factors, presented in Table 1 and Table 2.

A. Requirements Are Not Properly Stated

i. Multiple requirement acquisition approaches must be used; this includes the questionnaires, interviews and direct communication. The team deployed on the requirement acquisition should be capable enough to extract the accurate/valuable information from the information lot coming from different sources. The capabilities of the analysts in terms of requirements acquisition can be determined by their performance in the previous projects. An analyst having a very good track record of determining the requirements may be more trustable for deployment in the requirement acquisition process.
ii. Facilitated Application Specification Techniques (FAST) [7] should be used to ensure the elaborated understanding of the requirements at both ends, i.e. the customer and developer. This informal way of requirement collection helps the development team to understand the requirements in the actual context. [12]
iii. The customer must allow the development team to have a flexible schedule if the requirements are expected to change dynamically. Only minor changes, which don't have the impact on the architecture of the software, can be changed dynamically. The major changes, requiring the change in architecture, cant be completed in the same time and cost. Therefore, if the customer requires or expects the dynamic changes in the requirement definition, it must expect a relatively higher cost and time to complete the project. [13]
iv. The development team must be familiar with the Enhanced Information Deployment [7] technique, to take care of the default requirements that are not explicitly mentioned by the customer.

B. Low Estimation And Time And Cost

i. The development team while bidding for the project must have a clear idea of the requirements that are explicitly stated and also of those that are expected by default. It is appropriate that the management acquires multiple estimates from different sources, and suggest a flexible schedule in terms of time and cost. Only the acquisition of estimates from multiple sources is not sufficient but a mechanism should be in place to identify the best possible estimates out of available. It is recommended that this process be governed by the team of experienced analysts, developers and managers, in order to make this exercise more effective and result oriented [14].
ii. It has been observed that if the funding and time are not flexible, the incremental model [4] of development may be a solution. As it grows in increments, if the funding or time collapse, at least there is something presentable to the customer, rather than having nothing at all. Although the product may be incomplete yet the time and cost incurred can be presented to customer to grab the future funding for development purposes.
iii. The development team must try to find the maximum amount of reusable code, the availability of reusable code will have three dimensional positive effects. First it will decrease the time required for the software development by making available the code that was to be developed if the reusable code were not available. Secondly, it will decrease the cost of development as less development is required in the presence of reusable code, the higher the usage of re-usable code the lower the cost of software development comes. Thirdly, the re-usable code is already tested component and hence does not require re-testing, therefore, saving time of testing the component.
iv. The team of experienced developers and management may decide, in consultation with the customer, that if there are any scrubable requirements that may not harm the overall working of the software. Such requirements may be eliminated to save time and cost[15].
v. Clean room engineering may not be implemented in the projects that have tight time and cost schedule.

C. More Stress Of Users Than Expected

The developer must always expect and consider that the customer is not capable of describing all the requirements. The developer, if possible, must design and implement the system in a way that it can tolerate with the extra burden as well.

The developers must also do the extensive stress testing to ensure that the software is capable of handling the load and stress of the users. The development teams can stress test the software at component level, environment level, architecture level and end-to-end level. In *component level* we assume that although unit testing has its existence yet it has a disadvantage that in the domain of web services, it can't work to check the concurrency and deadlock of the simultaneous requests, adequately. Therefore it is necessary that each component residing on the web server is tested through the stress testing, in order to check that no deadlock occurs during the simultaneous access, and the consistent position of data is maintained and also no deadlock occurs while the records are being accessed and updated. In *environment level* and after the completion of the requirement engineering phase, the development team decides the hardware and software infrastructure that they plan to provide for the development life cycle. The infrastructure may include a database application, a front end application, a hardware platform and a load balancer. This infrastructure helps in determining the scalability, reliability and cost of application. Hence, all available

infrastructural option are to be reviewed categorically in order to identify and estimate the performance and the cost of performance. *Architectural level* stress testing is also called benchmarking. The basic purpose of the stress test on the application's architecture is to measure the cohesiveness of the component residing at the different levels. A well responsive application would ensure that all components at all tires are well associated and working properly. During the development process, the sample components may be taken from each tire of the cohesive modules to detect any flaws during analysis and design of the application. *End-to-End* stress test has a flavor of real test that may be prolonged to several hours and in some cases even to some days. These End-to-End test (if accurately designed) test the application as whole and at length [8].

D. Less Reuse Than Expected

i. While estimating for the projects cost and resource requirement, the developers must know that what amount of software is available for re-use, this should be an rational decision as, if the reusable code is not available the effort to develop such code will be duplicated. As not only code is to be developed, but also the component is to be tested before integration with other components. The person investigating for the availability of re-usable code must have adequate knowledge of existing libraries of components and must also know about the active libraries being updated. The active knowledge of web is also essential in this regard.
ii. If the component is to be developed, it is necessary that a clean room engineering approach is applied is the development so that the time required for testing the component is minimized if not completely eliminated [16].
iii. The best developer, among the available lot, should be deployed to develop the components so that the expected time on development and testing is minimized.

E. Delivery Deadline Tightened Or Manager Change Circumstances

i. The managers somehow try changing the circumstances because of the deadline pressure or because of the orientation of new requirements. The absolute definition of requirements at the beginning ensures that circumstances remain constant and deadlines are not tightened.
ii. The development team and management of the development firm must have the foreseeing capability, and should try adhering to the dynamic circumstances without disturbing the firm itself. For this purpose the firm must try and maintain the experienced staff who can use their intuition at the required time and contribute for the betterment of the firm.
iii. The FAST approach may be used to speed up the requirement acquisition, thus decreasing the negative impact of tightened deadlines. Although FAST session has the build-in capability to speed up the requirement acquisition process yet it is necessary that the FAST session is conducted with the sincerity, spirit and motivation. A FAST session that can't deliver positively causes the wastage of extra time that is very hard to manage in the coming time if the project is already behind the already agreed schedule.

F. Funding Will Be Lost

i. Inorder to ensure that funding issues remain in order, the development team must first ensure that the software is developed within time, developing within time will not only help to improve the revenues and profits but would also ensure that the funding remains available throughout the software development lifecycle. This is the win-win situation in which neither the development firm seeks extra time nor the customer is to pay anything extra for any requirement change.
ii. Its important that friendly relationship is maintained with the funding agency. A state of trust should be estabiliehed between both parties and they should be able to communicate with each other which utmost ease and without involving any other third party channel. The informal meetings of both parties at social events may be of great help in improving the warmness of the relation.
iii. Along with the cordial relationship with the funding agency, it is also important that the funding agency is kept updated regarding the progress of the software development process, and also any problem that is faced during the process. Being informed about the problems and achievements, the funding agency will be in a better place to help the development firm with the continuation of the funding.

G. Technology Does Not Meet Expectations

i. The decision about the choice of technology should be taken only after a very through consideration of the available tools and technologies and only by the experienced practitioners. The customer in some cases may allow the change in technology, but this change must not have any negative effect on the quality of the software, it is also important that any change in the already agreed tools and technologies is done only after the mutual consultation of the development team and the customers. It is the moral responsibility of the development firm to advise the most suitable solution to the customer if he does not have the adequate knowledge of the possible tools and technologies that are available to choose from the available lot [17].

ii. If the change of tool, is agreed between the customer and the development team the development team must try to choose the best available tool in consultation with the customer. The development team should choose the tool in which they have very good expertise so that the expertise in tool may be translated into the company's revenues and profits.
iii. The tool chosen should not only be acceptable to the customer but the customer should have necessary training on the tool. It is also important for the customer to argue with the development firm about the future acceptability of the product being developed by using that specific tool. The choice of tool must not only meet the current needs of the customer but should also be able to meet the future expectations of the customer.

H. Lack Of Training On Tool Or Staff Inexperience

i. The rapid advancement in the current tools and technologies force the developer to remain up-to-date. The development firm can keep its employees updated by offering them training on the emerging tools and technologies. Along with the training on the emerging tool, it is also important that the employees be also provided the advanced knowledge of the current tools in which the firm is doing the development currently. It is also important that someone in the organization have the vision and wisdom to use his intuition about the arrival of future technologies, so that the training can be arranged and provided to the employees in advance and market benefits can be obtained by having this advanced availability of the usage of technology [18].
ii. The firm may hire the new graduates from the leading universities, having some knowledge of the current tools. The firm can train them and provide them small assignments to do, in order to complete their training and making them a useful member of the firm, but all this requires a long planning and a visionary leadership at the firm, who can have the knowledge and wisdom about the emerging trends in technologies. In order to hire the graduates from the reputed universities, the firms may plan to schedule the seminar in the universities for the final term students and may opt to arrange on the spot job interviews to identify potential candidates for the possible hiring to meet the future needs of the development firm. This approach has been observed to be extremely helpful in not only fulfilling the industry-academia gap but in also producing the quality products for the industry by using the knowledge imparted by the academia [18].
iii. It is important that the teams are made for each project. Developing the team structure will help in not only promoting the efficiency of the work but will also help in providing experience to new members. This will also help the new members to learn about the smooth flow and effective handling of the tedious work. Such exercise will help them to learn the art of working in a team in also producing the outcome by doing smart work.

I. Staff Turnover

i. Staff, and particularly the experienced staff is an asset to any firm, and firms generally do their best to retain such individuals. But this is very obvious that learned individuals still want to change the jobs, although this trend may not be eliminated yet it can be reduced. The employer should keep the honest estimations of the salaries available in the market for experienced people. By giving less salary, the employer should not assume that the employee will work sincerely and with the best of his effort, rather the employee may keep on wasting his and firm's time by searching for other employment opportunities during the office hours [19].
ii. Proven experience show that employee enjoys working with an employer who has more care for the families of the employee. The employer may offer the services like, free family medical; children school fee, car allowance, house rent, etc in order to keep the employee attracted.
iii. The employer should provide other social gathering and meeting opportunities to the employees, inorder to help establish a family culture at the organization. This get-together is a good chance for the juniors to meet with the firm's top management and listen to their views and vision about the future of the firm's business strategy. The individual's must be encouraged to provide their view and their views must be considered valuable, so that each individual can feel his/her importance in the decision making of the firm[10].
iv. The employer must try to keep the employees updated and should provide the employees with chances to refresh their knowledge about the emerging tools and technologies [19]. This can be done by arranging the courses at their own site, or by sending the employees to the specialized institutes for training.
v. The employer may introduce a loan scheme to help the needy individuals and the return may be in easy installments, without or at a minimal interest rate [20].
vi. It is necessary that the employer try maintaining the respect and honor of the employees, and it is never compromised in any situation. It is obvious that the respect just does not come by paying the employee more, but it comes by having the friendly and trust oriented relationship. The employees must not be in a position of continuous tease; horror and

torture, a work done under such circumstances can hardly be productive and badly affects the mental and sociological health of the employee. The governing force for the employees to work should not be the threat and anger but the affiliation and desire. Therefore, the polite handling of the staff must be the top priority of the management. A specialized human resource (HR) department may be established in the organization to keep track of all the employee related affairs: including the salary increments, hiring and firing, leave and holidays, productivity, expenses vs. productivity ratio (EPR) etc. The employees having the high EPR must be given the salary rise according to their contribution in the firm's profits. The employees having the normal EPR may or may not be give some benefits, while the employees having low EPR should be warned properly in advance, according to the condition of the contract, before their contracts are terminated [21].

vii. The employer may introduce a bonus scheme to make the employees a part of the profit that the firm gains. This would give a sense of ownership to the employee and the employee will try to deliver according to the best of his capabilities [21].

J. Backup Not Taken & Actual Document/Data Loss

i. Backup must be taken at multiple sites, so that in case of any physical or technical damage the backup itself remains intact, the smaller software development firms may opt not to take backups as they may consider this effort as wastage of time and resources. Actually, they oversee the risk by just being over optimistic about the fact that data neither can be lost nor be stolen.
ii. The management must try to introduce the paperless environment in the firm; this would help in maintaining the efficient, secure, and traceable working environment.
iii. The backup sites may be frequently updated and the updates should be inspected regularly to reduce the chances of any data not being updated on the server. The firm may hire the services of reputed individuals to provide help in this regard, as this is considered the one of the most critical risk factor to be managed.
iv. The team strictures should be implemented in the development environment, this not only improves the working environment but also helps in decreasing the dependency on the individuals as the team members remain active and keep knowledge of the trends and patterns that someone uses in its development. This will not only help in introducing the harmony in the team members but would also increase the efficiency in the working environment [22].

K. Fire, Flood And Building Loss

i. The firm must ensure that the working environment across the organization is not only conducive but also safe for the employees. Proper smoke detectors and fire alarms must be installed in the building to detect the fire and the emergency exit should be provided in case of any emergency.
ii. The organization must also ensure that the building codes have been followed and the structure is according to the prescribed standards. With the orientation of more earthquakes recently in the world, it is also important that the building structure is developed in a way that it can absolve the earthquake shocks of an adequate level.

L. Too Many Development Error

i. Although testing techniques can help in identifying errors yet it is more appropriate to try enforcing the clean room engineering approach [23]. The cost to identify the errors in a relatively large amount of code can be both expensive and difficult at the same time. The cost of rectification of these errors is also very high as the schedule of the development is disturbed and many changes are to be made in iteration in order to bring the software on right track. Clean room engineering, although requires the development of error free code yet it can only be adopted when ample time is available for software development.
ii. For this purpose not only the development team must try working accurately but also the continuous inspections of the work being done by the developer must be reviewed by some senior colleague, so that the guideline may be provided early and correction are made without serious harm [9].
iii. Along with the availability of the inspections, the developer must unit test the piece of software that he is developing and must ensure that the code is free of errors and that it is according to the prescribed requirements [24].
iv. The small software houses, consider testing as a sole responsibility of the developers, and do not have a specific testing department. Although individual components may work fine but the integrated application may still not work, because of the run-time and integration errors. These types of errors are generally beyond the scope of the developer and are to be addressed by the specific testing team in the organization. Absence of dedicated testing team may cause serious problems for the organization in delivering the correct software in-time.
v. The organization must adopt the team structure in the software development. Along with the unit testing, that generally, the developer will do on his

own, the team can help each other to test the code and to ensure that the test cases are correctly designed and are efficiently handled in order to save time and improve the productivity of the resting process [25].

vi. A sudden jump to the new tools and technology adds the risk of too many errors. It is suggested that the jump to a new technology should not be made without adequate thinking and must be supported by the discussion and should be a result of a decision governed by the logical thinking. It should also be noted that adequate training on the tools must be available and provided before the actual shift in the technology is made.

vii. Sometimes there are so many errors identified in a piece of code that correction may not only cause the wastage of time but also the resources. In such circumstance, the re-development of that component may be easier than correcting the existing one. The decision of re-development is a very critical decision and should be supported by the logical discussion among the management governing the project. Before any such decision, a mathematical calculation should be done to logically represent that the re-development is in the benefit of the organization. A re-development must logically be completed in much higher speed as compared to the initial development [26].

viii. It is also important that the testing process works fine, i.e. identification of too many errors can still be less harmful as compared to the ignoring errors or un-identified errors [27], because the identified errors can somehow be tackled and addressed for correction, but an un-identified error may cause harm after all the bugs have been fixed. The errors become more harmful when they exist even after the release of the software. An error identified in external environment costs the firm much more to rectify that error. A released software is like a thrown arrow, once becomes public can't be brought back [28].

M. Developer Run Away With Code

i. At the time of appointment, the Human Resource (HR) department must ensure that the person they are hiring, is adequately trustable and owes a good employment history. His credibility can be checked from the previous employer. The contact details provided by the employer must be verified before the employee is hired permanently.

ii. The organization may also opt to take the employees from the accredited universities and resource providers so that only, already verified, individuals can find a place in the organization.

iii. The organization may also decide to hire the employees based upon the references or recommendation of their existing employees or someone may provide the guarantee for the employee for the purpose of reliability and trust [29].

iv. Backup must be taken at multiple sites, so that in case of any physical or technical damage the backup itself remains intact. The backup sites may be frequently updated and the updates should be inspected.

N. Lack Of Intuition

i. It has been observed that the experienced individuals can help in estimating the cost, budget and manpower of any project by just using their intuition [11]. The guess provided by them is generally accurate, and thus causes a huge benefit for the organization. The organization must do adequate effort to retain such people and should continue befitting from their experience.

ii. Talented individual must be attached to work with the experienced individuals so that they can learn that how the estimations can be made by using the previous knowledge and intuition [30].

IV Conclusion

Software development process is complex and requires efficient handling of the available resources. Poor planning invites risk factors that are very difficult to deal with. The paper unleashes the possible strategies to avoid or overcome risk, once they have been identified in a software process. Although a complete list of software risk factors is impossible to produce, as the risk factors keep on growing with the new tools and technologies, yet a comprehensive list has been considered for providing knowledge about the handling and avoidance mechanism. In the last three decades ample stress has been given on the identification, management, avoidance and handling of risk factors. This paper after having identified the risk factors, proposes the avoidance and mitigation strategies for each risk factor based on the frequency of their occurrence. The software houses that are developing the small and medium software can especially benefit by following the avoidance strategy.

V Acknowledgments

We are thankful to the Research Center (RC) at College of Computer & Information Science of King Saud University, for providing partial support for completion of this work.

VI References

1) Rothfeder, "It's Late, Costly, and incomplete-But Try Firing a Computer System, " Business Week, November 7, 1988, pp. 164-65
2) Coper Jones, "patterns of software success and failure", 1996
3) Roger S. Pressman, "Software engineering: a practitioner's approach", 5th ed, McGraw-hill, pp 151-159

4) Barry W. Boehm, "software risk management: principles and practices", pp 13
5) Basit Shahzad, Tanvir Afzal, "Enhanced risk analysis and relative impact factorization", 1st international conference on information and communication technology, IBA Karachi, August 27-28, 2005 ,pp 290-295.
6) Basit Shahzad, Javed Iqbal, ""Software Risk Management – Prioritization of frequently occurring Risk in Software Development Phases. Using Relative Impact Risk Model", 2nd International Conference on Information and Communication Technology (ICICT2007), December 16-17, 2007, IBA Karchi.
7) Roger S. Pressman, "Software engineering: a practitioner's approach", 5th ed, McGraw-hill, pp 151-159
8) Borland, the open alm company, A Load Testing Strategy, white paper, April 2006,pp6
9) Jiantao Pan, Software Testing, Carnegie Mellon University, Dependable Embedded Systems, spring 1999, pp 1-14.
10) Duport, "how to control and manage the staff turnover "http://www.duport.co.uk/guides/staff%20issues/Controlling%20and%20managing%20staff%20turnover.htm, May 2006.
11) Magic intuition, "definition of intuition "http://www.magicintuition.com/intuition.html", 2009
12) T.E. Bell, T. A. Thayer, Software Requirements-Are they really a problem?, International Conference on Software Engineering Proceedings of the 2nd international conference on Software engineering San Francisco, California, United States, Pages: 61 – 68, 1976.
13) Gursimran Singh Walia , Jeffrey C. Carver, A systematic literature review to identify and classify software requirement errors, Information and Software Technology, v.51 n.7, p.1087-1109, July, 2009
14) Longstreet Consulting Inc., Software development estimation, http://www.softwaremetrics.com/Articles/estimating.htm
15) Baskeles, B.; Turhan, B.; Bener, A. Software effort estimation using machine learning method, 22nd international symposium on, Computer and information sciences, Volume , Issue , 7-9 Nov. 2007 Page(s):1 - 6 Digital Object Identifier 10.1109/ISCIS.2007.4456863
16) Matsumura, K. Yamashiro, A. Tanaka, T. Takahashi, I,Modeling of software reusable component approach and its case study, Proceedings of 4th International Conference on Computer Software and Applications, (COMPSAC1990), 10/31/1990 - 11/02/1990, Chicago, IL, USA, pp 307-313
17) Jhon McManus, Risk Management in Software Projects, Computer Weekly professional series by Elsevier, pp 4-18.
18) Kim Man Lui1 and Keith C.C. Chan, Test Driven Development and Software Process Improvement in China, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 219-222, Friday, May 14, 2004
19) Sarah Wilson, CCRP, How to reduce turnover and manage employees, http://www.ehow.com/how_4495943_reduce-turnover-motivate-employees.html
20) How to motivate employees, http://www.ehow.com/how_2094622_motivate-employees.html
21) Dovinea, How to Hire, Manage and Motivate employees effectively, http://www.ehow.com/how_2377369_hire-manage-motivate-employees-effectively.html
22) Moore, R.W.; JaJa, J.F.; Chadduck, R. Mitigating risk of data loss in preservation environments, Proceedings. 22nd IEEE / 13th NASA Goddard Conference on Mass Storage Systems and Technologies , 11-14 April 2005 Page(s): 39 – 48, Digital Object Identifier 10.1109/MSST.2005.20
23) S. wayne Shere, Ara Kouchakdjian, Paul G. Arnold, Experience Using Cleanroom Software Engineering, IEEE Software, Volume 13 , Issue 3 (May 1996), Pages: 69 – 76,Year of Publication: 1996, ISSN:0740-7459.
24) IEEE Standards Board, "IEEE Standard for Software Unit Testing: An American National Standard, ANSI/IEEE Std 1008-1987" in IEEE Standards: Software Engineering, Volume Two: Process Standards; 1999 Edition; published by The Institute of Electrical and Electronics Engineers, Inc. Software Engineering Technical Committee of the IEEE Computer Society.
25) Carnegie Mellon-Software Engineering Institute, Overview of team software process and personal software process, http://www.sei.cmu.edu/tsp/
26) Sachidanandam Sakthivel, A decision model to choose between software maintenance and software redevelopment, Dept. of Accounting & MIS, Bowling Green State University, Bowling Green, OH 43403, USA
27) Bill Curtis, Top five reasons for poor software quality, http://itmanagement.earthweb.com/entdev/article.php/3827841/Top-Five-Causes-of-Poor-Software-Quality.htm, July 1 2009.
28) Leo King, Businesses fear lost revenues after poor software testing, http://www.infoworld.com/d/developer-world/businesses-fear-lost-revenues-after-poor-software-testing-134, April 08, 2008.
29) Dr. Dobb's, Investigating software and source-code theft,

http://www.ddj.com/windows/184406134, July 19, 2005.

30) Naur, Intution in software development, Proceedings of the International Joint Conference on Theory and Practice of Software Development (TAPSOFT) on Formal Methods and Software, Vol.2: Colloquium on Software Engineering (CSE) , Berlin, Germany Pages: 60 – 79, Year of Publication: 1985 , ISBN:3-540-15199-0
31) Danny Lieberman,http://74.125.153.132/search?q=cache:4VSx1A9wqVUJ:www.software.co.il/downloads/EnterpriseSoftware_RiskReduction.pdf+reducing+operational+risk+by+improving+production+software+quality&cd=1&hl=en&ct=clnk&gl=pk&client=firefox-a.
32) 2005 Breach Analysis, April 2006 http://www.software.co.il/downloads/breachAnalysis2005.xls
33) Privacy Rights Clearinghouse, http://www.privacyrights.org
34) Developing Secure Software, Noopur Davis, http://www.softwaretechnews.com/stn8-2/noopur.htm
35) J.H. Persson, L.Mathiassen, T.S. Madsen, and F. Steinson, "Managing risks in distributed software projects: An integrative franework", IEEE Transactions on engineering management, Vol. 56, No. 3, August 2009.
36) B.J. Alge, C.Witheoff, and H.J. Klein, "When does the medium matters? Knowledge building experiences and opportunities in decision making teams", Organ.Behav, Hum, Dects, Process, vol. 91, no. 1, pp 26-37, 2003.
37) BRADNER Erin ,MARK Gloria,HERTEL Tammie D." Team size and technology fit: Participation, awareness, and rapport in distributed teams", EEE transactions on professional communication ISSN 0361-1434, 2005, vol. 48, no 1 (104 p.) , pp. 68-77
38) R. N. Burns, A. R. Dennis, "Selecting the appropriate application development methodology ",SIGMIS Database, Vol. 17, No. 1. (1985), pp. 19-23.
39) Robert N. Charette, Software engineering risk analysis and management, McGraw-Hill, Inc. New York, NY, USA,Pages: 325,1989,ISBN:0-07-010719-X.
40) Ronald P. Higuera, Yacov Y. Haimes, "Software Risk Management", Technical Report CMU/SEI-96-TR-012 ESC-TR-96-012, June 1996.

Recognition of the Making on Integrated Circuit Chips Based On the Hybrid Fourier-AFMT

Yee Ming Chen[1] Jen-Hong Chiang[2]

***Abstract*-In this paper, an automatic optical inspection system incorporating optical character recognition (OCR) is employed to inspect markings on integrated circuit (IC) chips. Here, we present a novel OCR-based method for character recognition based on the features extracted from the integrated discrete wavelet and the hybrid Fourier-AFMT framework (DWF-AFMT) which is fast and yields better results as compared to basic Fourier Mellin Transform. We focus on the computation of a new set of invariant features allowing the classification of multi-oriented and multi-scaled patterns. Discrete Wavelet transform was used to smooth and preserve the local edges after image decomposition, making the character images less sensitive to shape distortion whilst Fourier-AFMT served to produce a translation, rotation and scale invariant feature, Multiple DWF-AFMT features could be used to form a reference invariant feature through the linearity property of Fourier-AFMT and reduce the variability of input images. The experiments showed the recognition rate reached over 100% when multiple DWF-AFMT features were used.**

Keywords-Printed marking recognition, analytical Fourier-Mellin transform (AFMT), Wavelet transforms, Optical Character Recognition (OCR)\

I. Introduction

Automatic optical inspection (AOI) system vision has many important applications in the electronics manufacturing industry, including the evaluation and testing of electronic components, inspection of IC chips. In a manufacturing system, the IC chips are lined up in a running conveyor for marking inspection. Markings on the IC chips are captured as a movie clip by a charged couple device (CCD) (Figure1). In machine vision inspecting markings on IC chips, an incorrect decision on marking may result in inappropriate placement of chip on printed circuit board during assembly process. The software then identifies whether the IC marking is of good quality. Thus automatic inspection of IC markings attracted considerable interest (Cho et al., 2005). The software, are used to convert characters in the images to their corresponding ASCII values for subsequent operations. However, poor lighting conditions during image capture can seriously degrade the quality of the digitized image; especially images captured in non-uniform illumination conditions are often noisy and low in contrast. Both these image acquisition methods degrade the quality of the digitized image, and directly influence the accuracy of character recognition

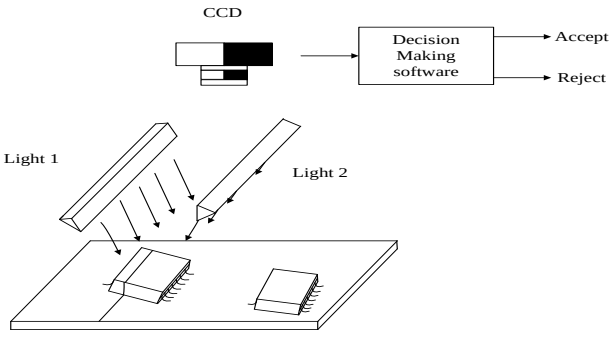


Figure 1. IC chip marking inspection system

The problem of finding reliable detection performance of the AOI for the character area for recognition of object IC chips marks is nowadays more present than ever. Since 1950, Optical Character Recognition (OCR) has been very active in the application of automatic pattern recognition, now it is used to read the printed characters at high speed (Mantas, 1986; Shunji et al, 1992). OCR is widely used in recognition of hand written characters and printed characters using Artificial Intelligence techniques (Mantas, 1986; Shunji et al, 1992; Adnan et al., 1996; Nagarajan et al., 2003). The OCR is applied for business card recognition (Wumo et al., 2001) where the manual input is optional, however, by scanning; the OCR is able to create an easier database. The document reading and analysis have reached an important position in certain markets. The application of OCR in the postal automation has followed into the banks and industrial recognition processes (Rafael and Richard, 1993; Mauritius and Christopher, 2001; Oivind et al., 1996). A popular technique to analyze images for the purpose of automatic recognition is supervised learning. In supervised learning a representative set of images is available from which relevant features that uniquely characterize the objects in the images can be extracted. The extracted features then allow the computer to recognize similar objects in images that are not contained in the representative set. Techniques for feature extraction in image analysis commonly use a basis of functions to generate a subspace over which images are projected. The coefficients resulting from the projection are then, either directly or the result of a combination of them used as features to characterize the image. Many different types of invariant features and invariant representations have been studied (Brandt and Lin, 1996) (Lin and Brandt, 1993). An approach is to represent

About[1] -Yee Ming Chen, Department of Industrial Engineering and Management
Yuan Ze University, Taoyuan, Taiwan, Republic of China
About -[2]Jen-Hong Chiang, Department of Industrial Engineering and Management
Yuan Ze University, Taoyuan, Taiwan, Republic of China

the images with frequency domain invariants that are derived from the Fourier-Mellin transform. The magnitude of the Fourier-Mellin transform is invariant with respect to rotation and scaling, but is incomplete. Although several sets of rotation and scaling invariant descriptors have been designed under the Fourier-Mellin transform framework (Milanese and Cherbuliez, 1999), the incompleteness property could not usually be satisfied. This is because the phase spectrum was always ignored. In order to overcome this problem, Ghorbel in Ref. (Ghorbel, 1994) proposed a complete set of rotation and scaling invariants under the analytical Fourier-Mellin transform (AFMT) based on the complete complex spectra. All above invariant transform have been extensively used in Image Analysis since they allow the extraction of rotation and scale invariant features for image recognition. Wavelets are also popular families of functions to build basis sets for feature extraction (Sastry et al., 2004; Li et al., 2004). Because of their efficiency to identify temporal/spatial features in different types of data sets, wavelet functions have gained popularity in a large domain of data classification applications. In (Lee and Pun, 2003) a log-polar wavelet transform is used to extract rotation and scale invariant features for texture analysis. In (Shen and Ip, 1999) also proposes the extraction of shape descriptors using wavelet analysis.In this paper, a novel method of character identification based on the IC chips features extracted from the integrated discrete wavelet and the hybrid Fourier-AFMT framework (DWF-AFMT) is proposed. By using DWF-AFMT features, the distortion and alignment problems are alleviated while retaining the advantages of the image-based approach. In this proposed technique, discrete wavelet transform preserves the local edges and reduce the noise in the low frequency domain after the image decomposition, and hence makes the character images less sensitive to shape distortion. In addition to that, the reduced dimension of the images also helps to improve the computation efficiency. Fourier-AFMT produces a translation, rotation in plane and scale invariant feature. The linearity property of Fourier-AFMT enables multiple DWF-AFMT features to be used to form a reference invariant feature and hence reduce the variability of the input images. Based on the method, a IC chip character identification system has been designed.We exemplify the use of our framework using wavelet functions and measure the quality of the extracted IC chips features as image descriptors in an OCR experiment. The rest of this paper is organized as follows. In Section 2, we introduce the pattern recognition system which includes a DWF-AFMT framework. In Section 3, we report our experiments. Finally in the last section, we draw the conclusion and point out some directions for future research.

II. Pattern Recognition System

A pattern recognition system includes three modules, i.e. the preprocessing stage, the extracted IC chip feature stage and the classification stage.

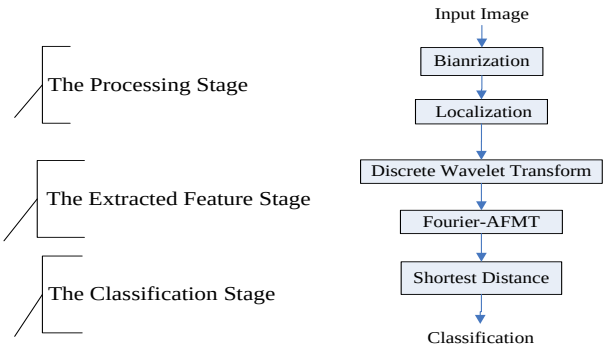


Figure 2. Block diagram of the proposed hybrid Fourier-AFMT and project profile combination framework

The framework of IC chip feature extraction is shown as Figure 2. Given an input face image, binarization converts the image pixels and the feature extractions are individually implemented with the hybrid Fourier-AFMT framework. Then classify an image feature by the shortest distance criterion.

A. The Processing Stage

The input image is initially processed to improve its quality by normalizing the image with some threshold which helps to offset the differences resulting form changing environmental and lighting conditions, etc. It increases the contrast for obtaining a suitable binary (black and white) image from the grey image obtained by camera for character identification. It involves

(1) Conversion of gray intensity image into a binary image by application of Canny method of segmentation (Canny, 1986). It finds edges by looking for local maxima of the gradient of the B & W image.

(2) Dilation using the linear morphological structuring element orthogonal to each other: Using this structuring element, holes and spots are filled, to make character localization easier.

(3) Sub-sampling: Averaging filter is applied, based on a fixed standard/threshold, in order to reduce the number of pixels herein in the image, leading to lesser computational complexity.

B. The Extracted Feature Stage

(1). Extract facial expression insensitive using wavelet transform

Wavelet transform is able to decompose image into a lower dimensional multi-resolution representation, which grants a compact hierarchical framework for interpreting the image information. The wavelet decomposition of a signal $f(x)$ can be obtained by convolution of signal with a family of real orthonormal basis, $\psi_{a,b}(x)$;

$$(W_\psi f(x))(a,b) = |a|^{-\frac{1}{2}} \int_{\Re} f(x)\psi\left(\frac{x-b}{a}\right)dx \qquad f(x) \in L^2(\Re) \quad (1)$$

where $a, b \in \Re, a \neq 0$ are the dilation parameter and the translation parameter respectively. The basis function

$\psi_{a,b}(x)$ is obtained through translation and dilation of a kernel function $\psi(x)$ known as mother wavelet as defined below:

$$\psi_{a,b}(x) = 2^{-a/2}\psi(2^{-a}x - b) \quad (2)$$

The mother wavelet $\psi(x)$ can be constructed from a scaling function, ψ(x). The scaling function ψ(x) satisfied the following two-scale difference equation

$$\phi(x) = \sqrt{2}\sum_n h(n)\phi(2x-n) \quad (3)$$

Where $h(n)$ is the impulse response of a discrete filter which has to meet several conditions for the set of basis wavelet functions to be orthonormal and unique. The scaling function ψ(x) is related to the mother wavelet $\psi(x)$ via

$$\psi(x) = \sqrt{2}\sum_n g(n)\phi(2x-n) \quad (4)$$

The coefficient of the filter $g(n)$ are conveniently extracted from filter $h(n)$ from the following relation

$$g(n) = (-1)^n h(1-n) \quad (5)$$

For 2D signal such as image, there exists an algorithm similar to the one-dimensional case for two dimensional wavelets and scaling functions obtained from one-dimensional ones by tensorial product. This kind of two-dimensional wavelet transform leads to a decomposition of approximation coefficients at level j-1 in four components: the approximations at level j, L_j and the details in three orientations (horizontal, vertical and diagonal), $D_{j\ vertical}$, $D_{j\ horizontal}$ and $D_{j\ diagonal}$:

$$L_j(m,n) = \lfloor H_x * \lfloor H_y * L_{j-1}\rfloor_{\downarrow 2,1}\rfloor_{\downarrow 2,1}(m,n) \quad (6)$$

$$D_{j\ vertical}(m,n) = [H_x * [G_y * L_{j-1}]_{\downarrow 2,1}]_{\downarrow 1,2}(m,n) \quad (7)$$

The original image is decomposed into four subband images. Similarly, we can obtain two levels of the wavelet decomposition by applying wavelet transform on the low-frequency band sequentially. According to wavelet theory, the low-frequency band is the smoothed version of original image with lower-dimensional space. It also contains the highest-energy content within the four subbands. The low-frequency band features are insensitive to_the expressions and small occlusion. Hence, if applying n level wavelet decomposition to the capital image, the recognition performance and space dimension were affected. In this paper, we let n = 2 applying in the letter images group.

(2). Analytical Fourier-Mellin transforms (AFMT)

This is then followed by our proposed hybrid complete invariant under the AFMT transform frameworks, respectively. For clarity, the basic notation used can be described as follows. We denote the Cartesian spatial domain by (x, y) and the Cartesian frequency domain by (u, v); we denote the log-polar coordinates by (ρ, φ) and the polar coordinates by (r, θ). When an image (or a spectra) is converted to the log-polar or polar coordinates, the intensity (or spectral) function name is preserved while the variable is replaced with (ρ, φ) or (r, θ), respectively.

(A). The analytical Fourier-Mellin transform

It is well known that the direct similarity group on the plane is equivalent to the space of polar coordinates:

$$\Pi = \{(r,\theta) | r > 0 and 0 < \theta < 2\pi\}$$

The Fourier-Mellin Transform on Π can be defined as:

$$\hat{f}(u,v) = M_f(u,v) = \int_0^{+\infty}\int_0^{2\pi} f(r,\theta)e^{-iu\theta}r^{-iv}\frac{dr}{r}d\theta, \quad (8)$$

for $u \in Z$ and $v \in R$

It is the Fourier-Mellin Transform of the irradiance distribution $f(r, \theta)$ in a two-dimensional image expressed in polar coordinates can be taken in the image center of gravity in order to obtain invariance under translations.

The integral (8) diverges in general, since the convergence is indeed under the assumption that $f(r, \theta)$ is equivalent to Kr^{α} (α>0 and K a constant) in a neighborhood of the origin (the center of gravity of the observed image). For this reason, the Analytical Fourier-Mellin transform (AFMT) was defined by

$$M_f(u, s=\sigma+iv) = \int_{R^+}\int_0^{2\pi} f(r,\theta)e^{-iu\theta}r^{\sigma+iv}\frac{dr}{r}d\theta, \quad (9)$$

For $u \in Z$, $v \in R$ and $\sigma > 0$. f is assumed to be square summable under the measure dθdr /r.

The AFMT of an object f can be seen as the usual FMT of the distorted object $f_\sigma(r,\theta)=r^\sigma f(r,\theta)$. The AFMT gives a complete description of gray-level objects since f can be retrieved by its inverse transform given by

$$f(r,\theta) = \int_R \sum_{v\in Z} M_f(u,\sigma+iv)e^{iu\theta}r^{\sigma+iv}dv$$

With a variable change on the integral (q=ln(r) instead of r), the equation (9) can be rewritten into Fourier transforms as follows:

$$M_{f_\sigma}(u,v) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_0^{2\pi} e^{q\sigma}f(e^q,\theta)e^{-i(u\theta+qv)}d\theta dq \quad (10)$$

A fast algorithm is obtained by computing a two dimensional Fast Fourier Transform on the log-polar distorted object $e^{q\sigma}f(e^q,\theta)$.The log-polar sampling is built from the points obtained by the intersection between N beans originating from the image centroid and M concentric circles with exponentially increasing radii. In this paper, we have chosen N=80 and M=80 for E, F, H and T Database, and σ= 0.5 (Fig.3)

Fig.3 The original image (left), The log-polar image of origin (middle), the magnitude image of AFMT (right).

Let us recall the transformation law of the analytical Fourier-Mellin transform for planar similarities. Let g be the

orientation and size change of an object f by the angle $\beta \in [0;2\pi]$ and the scale factor R_+^*, i.e. $g(r,\theta) = f(\alpha r, \theta + \beta)$. These two objects have the same shape and denoted similar objects. One can easily shows that the AFMT of g and f are related by :

$$M_{g_\sigma}(k,v) = \alpha^{-\sigma+iv} e^{ik\beta} M_{f_\sigma}(k,v) \quad (11)$$

For all *k* in Z, *v* in R and σ>0.

Equation (11) is called the shift theorem and suggests that the AFMT is well suited for the computation of global shape features which are invariant to the object position, orientation and size.

(B). A complete set of Fourier-Mellin features

Since the usual Fourier-Mellin transforms of two similar objects only differ by a phase factor (Eq. (11) without the $\alpha^{-\sigma}$ term), a set of global invariant descriptors regardless of the object position, orientation and size, is generally extracted by computing the modulus of some Fourier-Mellin coefficients (Crimmins, 1982). A set like this is not complete since the phase information is lost and it only represents a signature of the shape. Due to the lack of completeness, one can find distinct objects with identical descriptor values and a classification process may mix up objects, which is critical for content-based retrieval from image database (both false positive and true negative matches).

Recently, a complete family of similarity invariant descriptors based on the AFMT has been suggested (Ghorbel, 1994). This family can be easily written and applied to any strictly positive σ value as follows:

$$I_{f_\sigma}(k,v) = M_{f_\sigma}(0,0)^{-\sigma+iv} \cdot e^{ikArg(M_{f_\sigma}(1,0))} \cdot M_{f_\sigma}(k,v) \quad (12)$$

for all *k* in Z , *v* in R.

Each feature *I*(*k*, *v*) is constructed in order to compensate the $\alpha^{-\sigma+iv} e^{ik\beta}$ term that appears in the shift theorem (11). The compensation is achieved via the two Fourier-Mellin coefficients, M(0, 0) and M(1, 0), which are the normalization parameters. The set in Eq. (12) is complete since it is possible (i) to recover the FMT of an object from all of their invariant descriptors and the two normalization parameters by inverting Eq. (12); (ii) to reconstruct the original image by the inverse AFMT.

C. Proposed Fourier-AFMT transforms

When considering the translation, rotation and scaling together, we combine the translational invariant with the rotation and scaling invariant to construct a hybrid complete invariant under the Fourier-Mellin transform scheme. Indeed the translation property of the Fourier transform is the basis of these above invariants, which is also satisfied in the complex domain. Through the combination of Fourier invariant and AFMT invariant, we can construct a hybrid complete similarity invariant as followings:

$$S(\cdot) = AFMT(F(\cdot)), \quad (13)$$

Note that because AFMT(.) is directly applied to a complex spectrum but not separately applied to a magnitude spectrum and a phase spectrum to generate two invariant descriptors, the obtained hybrid invariant descriptor is complete.

But then, due to the reciprocal scaling property of the Fourier transform, when the property of Eq. (11) is applied in the polar domain of Fourier spectra, it needs to be modified as follows:

$$M_{g_\sigma}(k,v) = \alpha^{\sigma-2-ik} e^{iv\beta} M_{f_\sigma}(k,v) \quad (14)$$

The AFMT invariant of Eq. (12) is also modified as follows:

$$I_{f_\sigma}(k,v) = \left|M_{f_\sigma}(0,0)\right|^{(-\sigma+2+ik)/(\sigma-2)} \cdot \exp(-i \operatorname{var} g(M_{f_\sigma}(0,1))) \cdot M_{f_\sigma}(k,v) \quad (15)$$

In the same manner as Eq. (13), we can construct a hybrid complete invariant under the AFMT scheme as follows:

$$S(\cdot) = AFMT_M(F(\cdot)) \quad (16)$$

III. Classification Stage

For IC chips features recognition purposes, the classification of an unknown object into a set of reference patterns is achieved by several comparison methods. Besides them, the direct comparison of a couple of features, neural networks, or statistical classifiers, by means of intra- and inter-class similarity measures.

Since the invariant set of Eq.(12) is also convergent for square summable functions, it can be shown that the following function defines a true mathematical distance between shapes (Ghorbel, 1994):

$$d_2(I_{f_\sigma}, I_{g_\sigma}) = \left[\int_{-1}^{+1} \sum_{k \in z} \left|I_{f_\sigma}(k,v) - I_{g_\sigma}(k,v)\right|^2 dv\right]^{1/2} \quad (17)$$

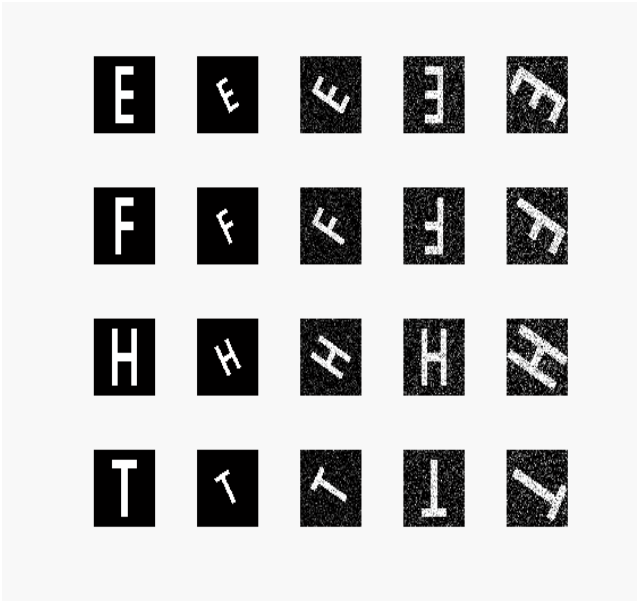
This distance is an Euclidean distance expressed in the invariant domain. Theoretically, it is zero if and only if the objects are identical up to a similarity transformation. Due to numerical sampling and approximation, we never have exactly zero and the value of the distance is used for the quantification for the similarity between objects, regardless of their pose, orientation and size in the image.

IV. Experiments and Results

Four binary images were used as IC chip feature prototypes, each containing a capital letter E, F, H, or T. Each letter has a uniform intensity equal to 1, and the uniform background of the images is set to 0. From each prototype, 96 versions are obtained by scale, orientation, or noisy changes of the letter (Fig.4). The resultant images show that the original feature image is low in contrast and has low gray levels. The 96 images containing the same letter are considered an image group. The invariant features M_{f_σ} for each image in the group should be identical.

Because translation will not change the relative position of the centre of mass of the letter, our major concern is the AOI system's performance on rotation angles are 30, 60, 90, 120, 180 and 240^0, and the four different scaling factors are 0.6, 0.8, 1.0 and 1.2. The noisy images are added at three levels with Gaussian noise having a mean value equal to 0 and a standard deviation(SD) equal to 0.04, 0.08 and 0.12. In this paper, 2 levels decomposition are performed on a image

with size 300 × 300 pixels. The 80 × 80 pixels sub-images have been chosen and it was interpolated and re-sampled in log-polar coordinates with 80 beans in the angular direction and 80 concentric circles in the radial direction.



(1) (2) (3) (4) (5)

Fig. 4. parts of four pattern groups: (1) with prototypes; (2) scaled by the factor 0.6 and with rotated by 30^0; (3) scaled by the factor 0.8, rotated by 60^0 and with noise SD 0.04; (4) with prototype, rotated by 180^0, and with noise SD 0.08; (5)

In the proposed method, Fourier-AFMT is based on Fourier Transform theory, which has a linear property. This implies that multiple m DWF-AFMT features can be used to form a reference DWF-AFMT feature and just only one representation per user needs to be stored in the database. The representation for each user, ($DWF\text{-}AFMT_{ui}$) can be formulated as follow:

$$DWF \cdot AFMT_{ui} = \frac{1}{m}\sum_{j=1}^{m} DWF \cdot AFMT_j^i$$

Where DWF-AFMT, is the invariance feature of the j^{th} view image of the i^{th} character. Producing a $DWF\text{-}AFMT_u$ feature from different training images, could relax various variability that occur during the acquisition process, such as sharp distortion and noise.

A. Classification in feature space

A database of 384 images of size 300×300 including 96 images per character from 4 characters have been used for experiments. We have chosen 80×80 sub-images such that the central sub-image to be entirely located in the pattern in all images. Each image is thus represented in a 441 dimensional feature space obtained by the magnitude of central analytical Fourier-Mellin harmonics (K=V=10). The recognition performances achieved by using the proposed DWF-AFMT features have been evaluated using a k-NN classifier. A number of k images from each character (for a total of 4k images) have been used as the training set, whereas the remaining 96-k images from each character (for a total of 4× (96-k)) have been used for testing. The interclass distances among the four prototypes are given in Table 1. The intraclass distances for each group of pattern points is defined by the mean-square error

$$\sigma = \frac{1}{N-1}\sum_{i=2}^{N}\left\|(M_f)_i - (M_f)_1\right\|^2,$$

Where N = 24 is the total number of images in the group and i=1 corresponds to the prototype. The root-mean-square errors for each group are given in Table 2, which are taken as clustering criteria for the images groups. By comparison with the interclass distances in Table 1, all four image groups are well clustered and well separated from one another.

Table 1. Interclass Distances among prototypes E, F, H

prototypes	E	F	H	T
E	0	0.4676	1.3515	0.7939
F	0.4676	0	1.3046	0.7674
H	1.3515	1.3046	0	1.5798
T	0.7939	0.7674	1.5798	0

Table 2. Intraclass Distances for the Image Groups E, F, H and T

prototype	E	F	H	T
σ	0.8683	0.7968	1.1424	0.5943

By comparison with the interclass distances of Table 1 and 2, an explanation for this may be as follows. The first one is that the interclass distance between capital letter E and F tend to be smaller than that between capital letter E and H. When interclass distance between two characters is small, any variation on digit will lead this digit to be easily confused with the other. The second reason why capital letter E and F digits are harder to classify than H and T digits is that the variance in image groups H and T are in general higher than that in case of E and F. The comparison characters images recognitions are done by studying the performances of different classification techniques. Five different nearest neighbor rule from k = 1 to k = 5, table 3 shows the results for higher nearest neighborhood classifiers of better recognition rate comparison.

Table 3. Comparison of 5 classification techniques on characters images recognitions

K-NN	1-NN	2-NN	3-NN	4-NN	5-NN
Recognition rate	0.73	0.92	0.92	0.96	0.97

The recognition rate from each noisy (SD=0.04, 0.08 and 0.12) and deformed image to all the four prototypes are shown in Table 4. By the nearest neighbor rule, these noisy images recognition rate showed in a descending order form top to bottom. It is clear that the decay of recognition rate of different classifiers for all the cases.

Table 4. Comparison of 5 classification techniques on characters noisy images recognitions

Recognition rate	SD=0	SD=0.04	SD=0.08	SD=0.12
1-NN	0.73	0.60	0.53	0.41
2-NN	0.92	0.72	0.52	0.5
3-NN	0.92	0.67	0.45	0.35
4-NN	0.96	0.64	0.44	0.32
5-NN	0.97	0.66	0.46	0.33

The recognition performances achieved by using the proposed DWF-AFMT features have been evaluated using different (K,V) pairs. The most interesting observation may be noticed from Table 5. It increases almost linearly with (K,V) pairs for different noise conditions. This result shows that, in general, that higher value (K,V) are more robust against noise.

Table 4. Comparison of different (K,V) of the proposed DWF-AFMT on characters noisy images recognitions

Recognition rate	SD=0	SD=0.04	SD=0.08	SD=0.12
K=V=7	0.97	0.65	0.41	0.30
K=V=10	0.97	0.66	0.46	0.32
K=V=15	0.97	0.72	0.54	0.43
K=V=20	0.97	0.73	0.55	0.46

B. Recognition results

We use 5 - NN classifier, have 3 misclassifies only. As the following experiment, 75% recognition rate for a misclassified. The recognition results for different scaling factors are given in Table 5, while the recognition results for a combination of rotation and scaling are shown in Table 6. We shall emphasis on testing for robustness to noise in Table 7. These results demonstrate the effectiveness of this feature extraction algorithm against geometric variation.

Table 5. Recognition result for different scaling factors

Scale factors	0.6	0.8	1.0	1.2
Recognition rate	75%	100%	100%	100%

Table 6. Recognition rate for different rotation angles and scaling factors

	30^0	60^0	90^0	180^0	240^0
0.6	100%	100%	100%	75%	100%
0.8	100%	100%	100%	100%	100%
1.2	100%	100%	75%	100%	100%

V. Conclusion

In this paper, we proposed the integrated discrete wavelet and the hybrid Fourier-AFMT framework (DWF-AFMT) for IC chips marking inspection. We focus our attention in this contribution on two particular points. The first one deals with the computation of a new set of invariant features allowing the classification of multi-oriented and multi-scaled patterns the estimation of characters orientation. From this point, the interests of this computation rely on the possibility to use this Analytical Fourier Mellin transform within a "filtering mode" that permits to solve the difficult problem of optical character recognition. The second point deals with the characters noisy images. These results are in agreement with the analysis of Section 3.1. They confirm that DWF-AFMT is more robust against noise. In this paper, we also present four different feature extraction methods have been studied for the marking inspection. The recognition accuracy of the proposed methods with 5-NN classifier is over 97%. As a conclusion, the proposed system is successfully developed and tested to solve the semiconductor industry problems

Acknowledgements

This research work was sponsored by the National Science Council, R.O.C., under project number NSC98-2221-E-155-023.

VI. References

1) Cho, C. S, Chung, B.M. and Park, M.J., (2005) Development of real-time vision-based fabric inspection system, IEEE Trans Industrial Electron, 52 , Pp.1073-1079.
2) Mantas, J.(1986) An Overview of Character Recognition Methodologies", Pattern Recognition, V0119(6), Pp.425-430.
3) Shunji Mori, Ching Y. Suen and Kazuhiko Yamamoto (1992) Historical Review of OCR Research and Development", Proceedings of the IEEE, vol80 (7), Pp. 185-201.
4) Adnan Amin, Humoud Al-Sadoun and Stephen Fischer (1996) Hand-Printed Arabic Character Recognition System Using An Neural Network", Pattern recognition, vol.29(4), Pp. 663-675.
5) Nagarajan, R., Sazali Yaacob, Paulraj Pandian, Karthigayan, M. Shamsudin Hj Amin and Marzuki Khalid (2003) A Neural Network Application for Inspecting the Marking of Symbols on Integrated

Circuit Chips". The 7th Triennial AEESEAP Conference Enhancing Engineering Education and Training, Malaysia, 8-9 August 2003,Pp. 1145-1161.

6) Pan, Wumo , Jin, Jianming, Shi, Guangshun and Wang, Q.R. (2001) A System for Automatic Chinese Business Card Recognition. Proceedings of the Sixth International on Document Analysis and recognition", IEEE Computer society, Pp. 577 – 581.
7) Gonzalez, Rafael C. and Woods, Richard E.(1993) Digital Image processing, Addison-Wesley Longman, Inc..
8) Mauritius, Seeger and Christopher, Dance. (2001). Binarising Camera Images for OCR. Proceedings of Sixth International Conference on Document Analysis and Recognition, 54 – 58.
9) Oivind, Due Trier , Anil, K.Jain and Torfinn, Taxt, (1996) Feature Extraction Methods for Character Recognition a Survey", Pattern Recognition, Vol.29(4), 641-662.
10) Brandt , R. D., Lin, Feng (1996) Representations that uniquely characterize images modulo translation, rotation and scaling", Pattern Recognition letters 17, Pp.1001-1015.
11) Lin, F. and R. D. Brandt. (1993) "Towards absolute invariants of images with respect to translation, rotation and scaling". Pattern Recognition letters 14 (5), Pp. 369-379.
12) Milanese , R., Cherbuliez , M., (1999) A rotation, translation, and scale invariant approach to content-based image retrieval, J. Visual Commun. Image Representation 10, Pp. 186-196.
13) Ghorbel, F. (1994) A complete invariant description for gray-level images by the harmonic analysis approach, Pattern Recognition Lett. 15, Pp. 1043-1051.
14) Sastry, C. S., Pujari, A. K., Deekshatulu, B. L., and Bhagvati C.(2004) A wavelet based multiresolution algorithm for rotation invariant feature extraction", Pattern Recognition Letters, Vol. 25, Pp. 1845-1855.
15) Li, C., Huang, J-Y., Chen, C-M.(2004) Soft computing approach to feature extraction", Fuzzy sets and systems, Vol. 147, Pp. 119-140.
16) Lee, M..C., and Pun, C.M.(2003) Rotation and Scale Invariant Wavelet Feature for Content-Based Texture Image Retrieval", Journal of the America Society: for Information Science and Technology, Vol 54. No. 1, Pp. 68-80.
17) Shen, D. and Ip H.(1999) Discriminative wavelet shape descriptors for recognition of 2D patterns", Pattern Recognition, Vol 32, Pp.. 151-165.
18) Canny, J (1986) A Computational Approach to Edge Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, Pp. 679-698.
19) Crimmins,T.R.(1982) A complete set of Fourier descriptors for two dimensional shapes", IEEE Trans. Syst., Man, Cybern. 12(6), Pp. 848-855.
20) Ghorbel, F.(1994) A complete invariant description for grey-level images by the harmonic analysis approach", Pattern Recognition Lett. 15, Pp. 1043-1051.

Global Journals Guidelines Handbook 2010

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 'AICCT' can 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.

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) Register yourself using top right corner of Home page then Login from same place twice. If you are already registered, then login using your username and password.

(II) Choose corresponding Journal from "Research Journals" Menu.

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

(B) If you are using Internet Explorer (Although Mozilla Firefox is preferred), 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 as an attachment.

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 Times New Roman.
- 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 two 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 welcome the submission of original paper, review paper, survey article relevant to the all the streams of Philosophy and knowledge. The Global Journals 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 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 and Editorial Board, will become the *copyright of the Global Journals.*

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 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, 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 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 homepage at the judgment of the Editorial Board.

The Editorial Board and Global Journals 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 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 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.

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 (Publication Prior to Print)

The Global Journals 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 .

Informal Tips for writing a Computer Science Research Paper to increase readability and citation

Before start writing a good quality Computer Science Research Paper, let us first understand what is Computer Science Research Paper? So, Computer Science Research Paper is the paper which is written by professionals or scientists who are associated to Computer Science and Information Technology, or doing research study in these areas. If you are novel to this field then you can consult about this field from your supervisor or guide.

Techniques for writing a good quality Computer Science Research Paper:

1. Choosing the topic- In most cases, the topic is searched by the interest of author but it can be also suggested by the guides. You can have several topics and then you can judge that in which topic or subject you are finding yourself most comfortable. This can be done by asking several questions to yourself, like Will I be able to carry our search in this area? Will I find all necessary recourses to accomplish

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

Please carefully note down following rules and regulation before submitting your Research Paper to Global Journals:

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.

Written Material: You may discuss with your guides and key sources.

- Do not copy or imitate anyone else paper. (Various 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

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.

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

M

N

O

P

R

S

T

V

W

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