

GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY

discovering thoughts and inventing future

Volume 10 Issue 14 Version 1.0

Online ISSN: 0975-4172

Print ISSN: 0975-4350

highlights

Knowledge Management Systems

Fuzzy Logic Techniques

Break Point Determination

World Wide Web

17 Technology
Reforming
Ideas

November 2010



Global Journal of Computer Science and Technology



Global Journal of Computer Science and Technology

Volume 10 Issue 14 (Ver. 1.0)

Global Association of Research

© Global Journal of Computer Science and Technology.2010.

All rights reserved.

This is a special issue published in version 1.0 of "Global Journal of Computer Science and Technology" By Global Journals Inc.

All articles are open access articles distributed under "Global Journal of Computer Science and Technology"

Reading License, which permits restricted use. Entire contents are copyright by of "Global Journal of Computer Science and Technology" unless otherwise noted on specific articles.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without written permission.

The opinions and statements made in this book are those of the authors concerned. Ultraculture has not verified and neither confirms nor denies any of the foregoing and no warranty or fitness is implied.

Engage with the contents herein at your own risk.

The use of this journal, and the terms and conditions for our providing information, is governed by our Disclaimer, Terms and Conditions and Privacy Policy given on our website <http://www.globaljournals.org/global-journals-research-portal/guideline/terms-and-conditions/menu-id-260/>.

By referring / using / reading / any type of association / referencing this journal, this signifies and you acknowledge that you have read them and that you accept and will be bound by the terms thereof.

All information, journals, this journal, activities undertaken, materials, services and our website, terms and conditions, privacy policy, and this journal is subject to change anytime without any prior notice.

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

Global Journals Inc.

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

Sponsors: Global Association of Research

Open Scientific Standards

Publisher's Headquarters office

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

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

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

Offset Typesetting

Global Journals Inc., City Center Office, 25200
Carlos Bee Blvd. #495, Hayward Pin: CA 94542
United States

Packaging & Continental Dispatching

Global Journals, India

Find a correspondence nodal officer near you

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

eContacts

Press Inquiries: press@globaljournals.org

Investor Inquiries: investors@globaljournals.org

Technical Support: technology@globaljournals.org

Media & Releases: media@globaljournals.org

Pricing (Including by Air Parcel Charges):

For Authors:

22 USD (B/W) & 50 USD (Color)

Yearly Subscription (Personal & Institutional):

200 USD (B/W) & 500 USD (Color)

Editorial Board Members (HON.)

John A. Hamilton, "Drew" Jr.,
Ph.D., Professor, Management
Computer Science and Software
Engineering
Director, Information Assurance
Laboratory
Auburn University

Dr. Henry Hexmoor
IEEE senior member since 2004
Ph.D. Computer Science, University at
Buffalo
Department of Computer Science
Southern Illinois University at Carbondale

Dr. Osman Balci, Professor
Department of Computer Science
Virginia Tech, Virginia University
Ph.D. and M.S. Syracuse University,
Syracuse, New York
M.S. and B.S. Bogazici University,
Istanbul, Turkey

Yogita Bajpai
M.Sc. (Computer Science), FICCT
U.S.A. Email:
yogita@computerresearch.org

Dr. T. David A. Forbes
Associate Professor and Range
Nutritionist
Ph.D. Edinburgh University - Animal
Nutrition
M.S. Aberdeen University - Animal
Nutrition
B.A. University of Dublin- Zoology

Dr. Wenying Feng
Professor, Department of Computing &
Information Systems
Department of Mathematics
Trent University, Peterborough,
ON Canada K9J 7B8

Dr. Thomas Wischgoll
Computer Science and Engineering,
Wright State University, Dayton, Ohio
B.S., M.S., Ph.D.
(University of Kaiserslautern)

Dr. Abdurrahman Arslanyilmaz
Computer Science & Information Systems
Department
Youngstown State University
Ph.D., Texas A&M University
University of Missouri, Columbia
Gazi University, Turkey

Dr. Xiaohong He
Professor of International Business
University of Quinipiac
BS, Jilin Institute of Technology; MA, MS,
PhD, (University of Texas-Dallas)

Burcin Becerik-Gerber
University of Southern California
Ph.D. in Civil Engineering
DDes from Harvard University
M.S. from University of California, Berkeley
& Istanbul University

Dr. Bart Lambrecht

Director of Research in Accounting and Finance
Professor of Finance
Lancaster University Management School
BA (Antwerp); MPhil, MA, PhD
(Cambridge)

Dr. Carlos García Pont

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

Dr. Fotini Labropulu

Mathematics - Luther College
University of 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., Eötvös Loránd University
Postdoctoral Training,
New York University

Dr. Söhnke M. Bartram

Department of Accounting and Finance
Lancaster University Management School
Ph.D. (WHU Koblenz)
MBA/BBA (University of Saarbrücken)

Dr. Miguel Angel Ariño

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

Philip G. Moscoso

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

Dr. Sanjay Dixit, M.D.

Director, EP Laboratories, Philadelphia VA
Medical Center
Cardiovascular Medicine - Cardiac
Arrhythmia
Univ of Penn School of Medicine

Dr. Han-Xiang Deng

MD., Ph.D
Associate Professor and Research
Department Division of Neuromuscular
Medicine
Davee Department of Neurology and Clinical
Neuroscience
Northwestern University
Feinberg School of Medicine

President Editor (HON.)

Dr. George Perry, (Neuroscientist)

Dean and Professor, College of Sciences

Denham Harman Research Award (American Aging Association)

ISI Highly Cited Researcher, Iberoamerican Molecular Biology Organization

AAAS Fellow, Correspondent Member of Spanish Royal Academy of Sciences

University of Texas at San Antonio

Postdoctoral Fellow (Department of Cell Biology)

Baylor College of Medicine

Houston, Texas, United States

Chief Author (HON.)

Dr. R.K. Dixit

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 & Chancellor (Asia Pacific)

deanind@computerresearch.org

Luis Galárraga

J!Research Project Leader

Saarbrücken, Germany

Er. Suyog Dixit

(M. Tech), BE (HONS. in CSE), FICCT

SAP Certified Consultant

CEO at IOSRD, GAOR & OSS

Technical Dean, Global Journals Inc. (US)

Website: www.suyogdixit.com

Email: suyog@suyogdixit.com

Pritesh Rajvaidya

(MS) Computer Science Department

California State University

BE (Computer Science), FICCT

Technical Dean, USA

Email: pritesh@computerresearch.org

Contents of the Volume

- i. Copyright Notice
 - ii. Editorial Board Members
 - iii. Chief Author and Dean
 - iv. Table of Contents
 - v. From the Chief Editor's Desk
 - vi. Research and Review Papers
-
- 1. A Congenial Access Control Technique for Knowledge Management Systems **2-6**
 - 2. Simplified AES for Low Memory Embedded Processors **7-11**
 - 3. Automatic Ontology Creation by Extracting Metadata from the Source code **12-18**
 - 4. A Simplified NMEA Sentence Generator for the Simulation of GPS Tracking **19 -26**
 - 5. Effectiveness of ICT in E-Governance With Special Reference To Jharkhand State **27-31**
 - 6. Various Security Attacks and Trust Based Security Architecture for MANET **32-36**
 - 7. A Modification on K-Nearest Neighbor Classifier **37-41**
 - 8. IEEE WLANs Standards for Mobile Ad-hoc Networks (MANETs): Performance Analysis **42-47**
 - 9. Analysis of Liver Disorder Using Datamining Algorithm **48-52**
 - 10. Performance Analysis of QoS in PMP Mode Wimax Networks **53-59**
 - 11. The Extraction of the Tooth Contour for Biometric Identification Device **60-65**
 - 12. A Proposed Novel Framework for Early Effort Estimation using Fuzzy Logic Techniques **66-72**
 - 13. Class Break Point Determination Using CK Metrics Thresholds **73-77**
 - 14. Fuzzy Logic Based Digital Image Edge Detection **78-83**
 - 15. Secure Mining of Association Rules over Horizontally Partitioned Data in Data Mining **84-88**
 - 16. Theory of Carry Value Transformation (CVT) and its Application in Fractal formation **89-98**
 - 17. Research on Automatic Fuzzy Ontology Generation for World Wide Web **99-102**
-
- vii. Auxiliary Memberships
 - viii. Process of Submission of Research Paper
 - ix. Preferred Author Guidelines
 - x. Index

From the Chief Author's Desk

We see a drastic momentum everywhere in all fields now a day. Which in turns, say a lot to everyone to excel with all possible way. The need of the hour is to pick the right key at the right time with all extras. Citing the computer versions, any automobile models, infrastructures, etc. It is not the result of any preplanning but the implementations of planning.

With these, we are constantly seeking to establish more formal links with researchers, scientists, engineers, specialists, technical experts, etc., associations, or other entities, particularly those who are active in the field of research, articles, research paper, etc. by inviting them to become affiliated with the Global Journals.

This Global Journal is like a banyan tree whose branches are many and each branch acts like a strong root itself.

Intentions are very clear to do best in all possible way with all care.

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

A Congenial Access Control Technique for Knowledge Management Systems

Julius Olatunji Okesola & Oluwafemi Shawn Ogunseye

GJCST Classification
H.2.8, K.6.5

Abstract-Usability is of extreme importance in any system design. In knowledge management systems, the need for usability is heightened by the inertia to use the system by workers. The current popular access control technique used by in KMS and portals is not exactly suitable for such a sensitive system because it does not amend to the fuzzy nature of a KMS and KM functions and ends up making the system difficult to use and violates the overall objective of the system. The research highlights usability issues as one of the problems of KMS and a potent cause of failure it was therefore treated with such seriousness. A more congenial access control technique was proposed which allows for the fuzziness inherent in KMS for large organizations. The model was evaluated through a real-world implementation – the dotCSC and the design proved viable. The system had a 0% false positive and an initial 2.1% false negative rate which was quickly corrected. It eliminated the stress of continuous role engineering and modifications. The system also recorded a high level of usability based on an online survey conducted through it. Overall, we achieved adequate security and usability, a goal which has been elusive to KMS and other systems.

Keywords-Knowledge management systems, Access control, Information security, semantic web, web mining, FOAF

I. INTRODUCTION

Saying knowledge is the world's most important resource is not entirely far from the truth. The status of an organization as far as competitiveness is concerned is dependent on the knowledge of the organization's workers. Knowledge Management describes how organization's manage and apply this knowledge. Interest in knowledge management has been very high and impressive with organizations deploying projects to manage knowledge worth millions of dollars[1]. As stated by [2], knowledge management projects would amount to a waste of time and money if organizations are not successful in its mediation and use. Knowledge management systems are information systems that help organizations acquire, manage and provide knowledge to solve their problems. They are systems that serve as mediators and are supposed to provide knowledge to users when, how and where they need them depending on the organization's policy. No organization would want their high cost knowledge management system to fail. One of the major causes of poor attitude to KMS leading to its eventual failure is the usability. This is true for most information systems[3]. Security has been known to have an inverse

relationship with usability of systems[4]. The problem is basically the more intense the security the less usable a system becomes. This effect is amplified in knowledge management systems due to the already present reluctance users exude towards knowledge sharing and knowledge management systems [5,6]. Once people start to get discouraged or the system proves too difficult to use, the apprehension is quickly transferred to others discouraging them from using the system. The most prominent type of security measure used in KMS is Access control. Access control involves users providing the system with their valid identity and the system verifies the supplied identity in order to determine eligibility for access and the allowed activities for such legitimate user. Access control techniques are primarily classified into discretionary access control (DAC) and Non Discretionary Access Control (NDAC). DAC is an access control policy controlled by the owner of an object [7]. NDAC is directly opposite to DAC and does not involve the owner specifying the access policy to his/her resource[8]. Most KMS portals use Role based access control (RBAC). RBAC is a NDAC and involves assigning roles to users that will determine how they can access the {knowledge} resources [9&10]. Despite its wide use, knowledge management systems are undoubtedly more sensitive than portals and other information system applications and the unsuitability of RBAC as the access control technique is beginning to show. RBAC restricts all users to one role or the other in a finite set of roles. It does not appreciate the fuzzy nature of user roles in organization and which might not be exhaustible by a predefined role or set of roles [11]. This actually becomes more pronounced in the globalized state of large corporations today. So many mergers, collaboration and acquisition leading to so many established communities and systems coming together. KMS use in such systems require hitch free integration despite the different organizational structure. The predefined role in RBAC makes it tasking adapting to such growth or development. Knowledge requirement can also be fuzzy, how do we say for certain that a particular staff does not need to know a certain thing. This can be easy atimes and can also be very difficult. We therefore propose a more congenial access control technique which will use the user's reputation to determine their access rights to knowledge resources. Access is determined based on semantic information used to generate a reputation/authorization score that determines a user's access rights. The perspective taken therefore involves looking at the user as a member of a community regardless of our dispersed in time and space the other members of the community might be. His right to a

About¹ - Department of computer science;University of Agriculture, Abeokuta, Ogun State, NigeriaStatetunji_okesola@yahoo.co.uk, ogunseyeoluwafemi@yahoo.com

particular knowledge resource is therefore built on this relationship he has with the creator or owner of the knowledge resource and the collaborative knowledge community.

II. RELATED WORK

[12] used a Friend Of A Friend (FOAF) architecture based web mining techniques to study relationship between users of a semantic web based social network site. They also were able to rank users by calculating trust values for them in the network. The EigenTrust [13] computes global trust values for peers based on their previous behavior. [14] proposed a strategy for calculating global trust for individuals in the network from the perspective of the designated seeds. The strategy uses group assertions for determining membership within a group. [15] created a method that involves certifying users at three levels. This method is very good in identifying “bad” nodes and resisting their effect in the computation of trust. [16] used a local metric to calculate trust values for individual on a network. A similar method was also used in [17] but defers in that it uses a probabilistic interpretation of global belief combinations. [18] proposed an efficient algorithm for generating locally calculated reputation ratings from a semantic network and applied it to a mail application to rate mail according to relevance. This method was mathematically tested and proved to be highly effective and one of the most realistic. All these approaches are highly effective but not suitable for this work because we need a technique that can compute local trust values and does not involve a user manually ranking other users. To achieve this, we therefore build on the algorithm proposed by [18] and [12].

III. METHOD

First we view the organization’s organogram as a multilevel tree structure (organizational tree structure). The chairman/president is at the root of the tree. We then assign values to each of the levels. Depending on the depth of the tree. The root has the highest value and highest weight and it decreases down the tree. Previous reputation networks are made of ratings given from one node to another [18] implying that each node (user) ranks each node on a particular scale e.g. 1-9. In this work we use a defined rank of 2 if there is a relationship between the users and 0 if there isn’t. we assign the value of 1 to indirect relationships. When a user wants to access a knowledge resource, the user is evaluated to see if they have a direct connection with the owner of the resource i.e. they are in the same group. If they do, they are allowed to use that resource. If they do not then they are evaluated to see if they have an indirect relationship (see figure 1). The nodes are therefore traversed to find out. The level a node is on the tree also has its effect. For instance, the president should be able to access any knowledge resource on the system therefore a score of 1 on level 3 is different from 1 on level 5. Since level 5 is higher, the relationship is stronger and the score will therefore be higher. The data is gotten through a customized FOAF

semantic data (figure 2) which applies to all user and keeps a record of their groups and affiliations. To detect relationship or authorization mark, the user is first probed using a web mining technique and then each member of their group is also probed to detect indirect relationship. The result is then used to calculate the user’s authorization score.

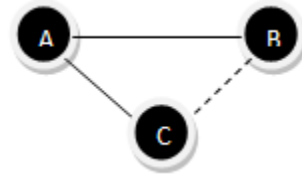


Figure 1: the possible relationships between nodes.

Direct relationship is shown with the straight thick line and indirect relationship is shown by the dashed line. C is the resource, and A has a direct link to it, and A has a direct link to C then B has an indirect link to C. from [12]

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:foaf="http://xmlns.com/foaf/0.1"
xmlns:unaab="http://www.dotcsc.com/
okesola/tasued/0.1">
<foaf:Person>
<foaf:mbox rdf:resource="Okesola@tasued.edu.ng"/>
<foaf:name>okesola</foaf:name>
<foaf:interest rdfs:label="Character agent"
rdf:resource="http://www.dotcsc.com"/>
<foaf:currentProject rdfs:label="CACT"
rdf:resource="www.dotcsc.com"/>
<foaf:workplaceHomepage rdfs:label="department of
computer science"
rdf:resource="www.dotcsc.com"/>
<dotcsc:group>programming
</position:seniorlecturer>
</level:5>
<foaf:Person>
<foaf:mbox rdf:resource="Ogunseye@tasued.edu.ng"/>
<foaf:name>ogunseyeoluwafemishawn</foaf:name>
</foaf:Person>
</group:linux group>
</group:programmers group>
</position:analyst>
</level:3>
</foaf:Person>
```

Fig. 2. An example of a FOAF file based on extracted information from users

1) The Algorithm

The algorithm for our proposed access control technique is as follows:

Step 1: Login the user logs in to his computer and the KMS [the KMS can use the system’s login so the user will not have to log in to the KMS as a separate entity. This is an optional feature]

Step 2: User profiling

The user is immediately profiled and his FOAF file pulled from the database

Step 3: Resource Requisition and Resource Owner Profiling
When the user requests a resource, the FOAF file of the owner is immediately pulled to an active memory.

Step 4: Processing Authorization and FOAF Mining

To achieve this step, the FOAF file is mined and the data it contains is extracted to be used for computation of reputation score.

This step involves four stages.

1. Get the prescribed threshold score for access authorization set by the KMS administrator.
2. Check if the user is directly connected to the resource owner, if yes, grant access if no then goto 3
3. Check if the members of the group are connected to the owner or a member of the owners group and compute their authorization score
4. If authorization score > threshold then access granted else, ask user to request permission from owner.

The FOAF data would contain, the person's name, the group(s) they belong to, their position {used in computing the level in the organogram scale}

The algorithm implies a member of a group can request resources belonging to someone or a group which he is not directly connected to but indirectly connected to {through a colleague that knows a colleague that knows a colleague.....} which would not be possible in a role based architecture. This is illustrated in figure 3.

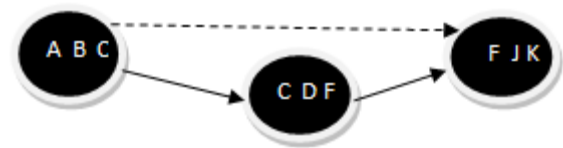


Figure 3. access control through user relationship and reputation

In figure 4, the assumption is that any user in the group made up of members A,B,C can request access to a knowledge resource from another group with members F,J,K which it might not have any knowledge of or direct connection to. This will be made possible through the use a link group which user C from the first group and user F from the owner's group both belong to. If this were defined in terms of roles a user A will not be able to access knowledge resources belonging to owner J in an independent group structure.

2) The Authorization Score Computation Model

The model used for computing the authorization score is built on the popular PageRank made popular and used by Google to rank web pages and sites based on links and has been established as an efficient model. This model had been used effectively in [12][19] to rank user based on trust and reputation. In explaining our application of the model we consider the authority value for a node v that requests a knowledge resource has an authorization value $A_n(v)$ on iteration n . The authorization value propagates to neighboring nodes in proportion to the node's relevance.

$$A_{n+1}(v) = c \sum_{v' \in Neighbor(v)} \frac{R(v, v')}{Rsum(v)} A_n(v') + cE(v)$$

$$Rsum(v) = \sum_{v'' \in Neighbor(v)} R(v, v'')$$

Where **Neighbor(v)** represent a set of nodes, each of which is connected to node v , c is a constant for normalization and E represents a source of authorization value in this case the level of node v . The owner of the document represents our target and is represented as v_{target} . The final score is then compared to a threshold value which determines if access is granted or not. Note that the score are rounded up to the nearest integer value before comparison. See [12] and [20] for more details of the mathematical model.

3) Assigning the threshold value

The flexible threshold value gives the KMS administrator an opportunity to vary the strictness of access to knowledge resources and therefore the strictness of security for the system. By default, the system could be set to a score which will equal the attainable score if the user and the owner was on the same team. In this case that would be a score of 2.

Our implementation is discussed in the next section to serve as a sample.

IV. APPLICATION OF THE MODEL

The original influence for this research was the failure of the RBAC to meet the need for the Social Networking/KMS dot CSC" a web based wireless intranet portal designed by the second author for the department of computer science UNAAB in 2008 based on issues already discussed in the Introductory section of this paper. The proposed congenial model was therefore implemented as a replacement for the RBAC. The figure x show the screenshot of the SNA.

The model is therefore evaluated based on the reasons for its creation, these are: Security and Usability

1) Security

To evaluate for security, after four months of full scale deployment of the dotCSC wireless intranet portal, the

portal had 346 registered users. Upon implementing the congenial access control model proposed in this paper, The security capabilities was tested by separating user access along class lines which could be overcome by the users due to group affiliations. Lecturers were given a rank of 5 and 100 level students 1. The users are grouped as follows:

Table 1: Groups according to class

Level	Number
100	50
200	63
300	121
400	106

There are also 5 lecturers registered on the site Total:346

Table 2: Groups according to Interest

AVERAGE LIKERT SCALE RATINGS FOR THE DOTCSC USING THE SCALE 1 = DISAGREE AND 4=AGREE	
Accessing knowledge resources on the dotCSC was easy	3.8
I enjoyed conversing and sharing ideas on the dotCSC	3.9
The flow of operation was straightforward	3.9
It was easy using the portal to solve my problems	3.6

below; 127 people participated in the polls. The results for the usability test is shown in table 3 below.

Group Name	Number
Linux Groups	15
Programmers Club	191
Information Systems Group	23
Artificial Intelligence Group	47
Number of users not in a group	70

From the system log after 5 weeks of active use the results are analyzed as follows (see figure 4).

Total amount of login session = 1722
 Total amount of distinct users = 313
 % of False positives based on affiliation = 0.00%
 % of False Negatives based on affiliation = 2.1%
 % accuracy level of the system = 96%

Figure 4: the results of access log analysis

2) Usability

To estimate this, we polled the users through the portal: Questions relevant to this survey are shown in the table

V. DISCUSSION

The sample implementation the viability and accuracy of our design. Though not perfect the false negative was quickly corrected by increasing the threshold. The system also proved to be acceptable to users who hardly realized any difference. The system was designed to monitor user login and also the attempts to access resources which triggers the authorization score computation. The system performed better than RBAC which was about 89% accurate in terms false positive and false negative computation and peaked at 91% when we used it. It also did not deny anyone access to resources who should have been allowed table 3: results of the usability poll

limitation

The model assumed a collaborative society where knowledge is allowed to be shared with almost everybody and where they are work groups and units. The implementation also showed that the initial computation of authorization score can be processor and memory intensive as the organizations grows. This we hoped can be solved by caching computation result in a fast memory.

VI. CONCLUSION AND FUTURE WORK

In this work, we have proposed a congenial access control for knowledge management system based on the peculiarities of knowledge management and the unsuitability of the commonly used role based access control technique to address those needs. Through our methodology and evinced in the sample, the simplicity this model brings to access control for KMS is unique and achieves the desired and erstwhile elusive goal of accurately blending security and usability. This is highly due to the user focused design. This work will serve as a model for KMS developers to follow in their search for a more usable KMS. The effect of an increase in usability of a system is positive and multiplier and can be adopted even beyond the field of KM.

VII. REFERENCE

- 1) Awad E.M. and Ghaziri H.M. (2004), Knowledge Management", Pearson Education, Inc.
- 2) Eva Maaninen-Olsson May Wismén and Sven A. Carlsson, (2009), Permanent and temporary work practices: knowledge integration and the meaning of boundary activities, Knowledge Management Research & Practice (2008) 6, 260–273
- 3) Wing Lam & Alton Chua (2005) KNOWLEDGE MANAGEMENT PROJECT ABANDONMENT: AN EXPLORATORY EXAMINATION OF ROOT CAUSES, Communications of the Association for Information Systems (Volume 16, 2005) 723-743

- 4) Lorrie Faith Cranor, Simson Garfinkel (2005) Security and Usability: Designing Secure Systems that People Can Use, O'Reilly Media
- 5) Jean-Grégoire Bernard, (2006) "A Typology of Knowledge Management System Use by Teams," *hicss*, vol. 7, pp.155a, Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06) Track 7, 2006
- 6) Kamla Ali Al-Busaidi, Lorne Olfman, Terry Ryan, and Gondy Leroy (2010), Sharing Knowledge to A Knowledge Management System: Examining the motivators and the benefits in an Omani organization, *Journal of Organizational Knowledge Management*, IBIMA Publishing Vol. 2010 Article ID 325835, [online] <http://www.ibimapublishing.com/journals/JOKM/jokm.html>
- 7) Cavale, M., & McPherson, D. (2003). *Cu, Role-based access control using Windows Server 2003 Authorisation Manager*. Microsoft Corporation. Retrieved from <http://www.microsoft.com/technet>
- 8) Ferraiolo, D., Kuhn, D., & Chandramouli, R. (2003). *Role-based access control*. Artech House, Computer Security Series.
- 9) Kamvar, Sepandar D. Mario T. Schlosser, Hector Garcia-Levien, Raph and Alexander Aiken. (1998) "Attack resistant trust metrics for public key certification." *7th USENIX Security Symposium*, San Antonio, Texas.
- 10) Hu, V. C., Ferraiolo, D. F., & Kuhn, D. R (2006). *Assessment of access control systems*. NIST Interagency Report 7316.
- 11) Mori J., Matsuo Y. Hashida K. Ishizuka M., Web Mining Approach for a User-centered Semantic Web, [online] <http://swoogle.umbc.edu>. Accessed 9/august/2010
- 12) Molina, The EigenTrust Algorithm for Reputation Management in P2P Networks", *Proceedings of the 12th International World Wide Web Conference*, May 20-24, 2003, Budapest, Hungary.
- 13) The Advogato Website: <http://www.advogato.org>
- 14) Golbeck, Jennifer, Bijan Parsia, James Hendler, (2003), Trust Networks on the Semantic Web," *Proceedings of Cooperative Intelligent Agents 2003*, August 27-29, Helsinki, Finland.
- 15) Kumar, Ravi, Prabhakar Raghavan, Sridhar Rajagopalan, D. Sivakumar, Andrew Tomkins, and Eli Upfal. (2000) "The web as a graph". *Proceedings of the Nineteenth ACM SIGMODSIGACT-SIGART Symposium on Principles of Database Systems*.
- 16) Richardson M, Agrawal R, Domingos P. Trust Management for the Semantic Web", available at <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.3.6539&rep=rep1&type=pdf>. Accessed 19/august/2010.
- 17) Golbeck J. and Hendler J. () Inferring Reputation on the Semantic Web
- 18) Y. Matsuo, H. Tomobe, K. Hasida, and M. Ishizuka. Finding Social Network for Trust Calculation, *In Proc. 16th European Conf. on Artificial Intelligence (ECAI2004)*, 2004.
- 19) S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine, *In Proc. 7th WWW conf*, 1998.

Simplified AES for Low Memory Embedded Processors

GJCST Classification
E.3,C.3,C.1.M

Suhas J Manangi, Parul Chaurasia, Mahendra Pratap Singh

Abstract-Embedded Processors have 2 main constraints^[1]: Memory and Processing Power. Encryption algorithm running on such processors should be specially customized to overcome these 2 constraints. AES^[2] is a well-known standard algorithm for encryption but its memory and processing power requirements are very high. This paper proposes modifications to AES to optimize it to overcome above 2 constraints.

Keywords-AES, Encryption, Decryption, Embedded Processors.

I. INTRODUCTION

Embedded systems are those which are complete devices often including hardware and mechanical parts. For Eg: Digital watches, MP3 players, mobile phones, videogame consoles, GPS receivers, Wireless Sensor Nodes, dedicated routers and network bridges etc. Embedded systems have mainly 2 constraints^[1]: Low memory and Low Processor Capacity. In this paper a simplified version of Advanced Encryption Standard is designed for low memory embedded processors. This algorithm can be used in cases where data or network security is required involving embedded systems like Wireless Sensor Nodes, Mobile Phones etc. The proposed algorithm tries to reduce amount of memory needed by AES and also increasing amount of processing needed for encryption and decryption. AES^[2] is Advanced Encryption Standards designed from larger collection of Rijndael algorithm. AES has 3 flavors AES-128, AES-192, AES-256 with block size 128 and number of rounds is 10, 12, and 14 respectively.

II. PROPOSED ALGORITHM (LMEP-S-AES)

LMEP-S-AES is a simplified version of AES, it has all the functions AES has but made suitable for embedded systems by optimizing on its memory and processor requirements. Encryption part of LMEP-S-AES has 5 sub functions: Key Expansion, Adding Key Round, Substitution Function, Row Transformation function and Mix Column function. Decryption part of LMEP-S-AES has 5 sub functions: Key Expansion, Inverse Mix Column functions, Inverse Row Transformation function, Inverse Substitution function and Adding Key Round

1) Key Expansion

Key length here is not fixed but can vary between 1 Byte to 16 Bytes^[3]. Initial key is expanded to 16 bytes expanded

key. This same key is applied in each round during encryption and also during decryption. Key is basically XORed with data blocks during encryption as well as decryption. Initial key is expanded using substitution S Box mentioned in algorithm in Figure 1.

```

KeyExpansion ( ByteString InputKey, Int KeyLen )
{
    RoundKey = InputKey

    for ( i=0; i<16; i++ )
    {
        RoundKey[i] = SubByte ( RoundKey[i%KeyLen] )
    }

    return RoundKey
}

```

Figure 1: Key Expansion

2) Encryption

Block size is same as AES, 128 bits arranged in 4x4 bytes called State Box. Encryption algorithm is explained in Figure 2. Number of rounds here is reduced to three and transformation functions are applied in the order of adding round key, substitution, row rotation and mix column functions.

```

Encryption ( ByteString Input, ByteString Output, ByteString InputKey, int ByteKeyLen )
{
    ByteString State[4][4] = Input
    ByteString RoundKey = KeyExpansion ( InputKey, ByteKeyLen )

    for ( i=0; i<3; i++ )
    {
        State = AddRoundKey ( State, RoundKey )
        State = SubByte ( State )
        State = ShiftRows ( State )
        State = MixColumns ( State )
    }

    State = AddRoundKey ( State, RoundKey )

    Output = State
}

```

Figure 2: Encryption[2] [3]

a) Add Round Key

For each round key used is same, and is XORed with data in State Box. Complete algorithm is shown in Figure 3 below.

```

void AddRoundKey ( ByteString RoundKey )
{
    k=0;
    for ( i=0; i<4; i++ )
    {
        for ( j=0; j<4; j++ )
        {
            State[i][j] = State[i][j] XOR RoundKey[k++]
        }
    }
    return ;
}

```

Figure 3: Add Round Key[2] [3]

b) Substitution Transformation

S Box is modified into 4x4 Matrix from 16x16 Matrix in AES. Each byte in State Box is split into 2 Nibble. Each Nibble is replaced by another nibble from S-Box in same manner of AES. First 2 bits of nibble represents row number and next 2 bits represents column number. For Eg: if Nibble value is 1010 then present nibble is replaced with 2nd row, 2nd column nibble of S-Box. Substitution of bytes is shown in Figure 4, breaking of each Byte into 2 nibbles is shown in figure 5, substituting each nibble with nibble from S-Box is shown in Figure 6. S-Box is shown in Table 1.

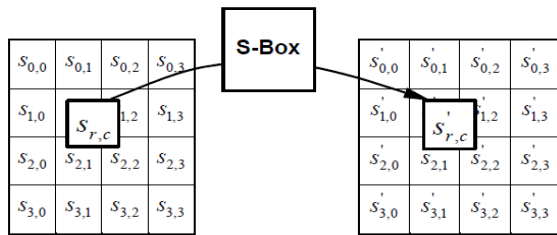


Figure 4: Byte Substitution[2] [3]

```

Byte SubByte ( Byte B )
{
    Nibble N1, N2
    N1, N2 = B
    N1 = SubNibble ( N1 )
    N2 = SubNibble ( N2 )
    B = N1, N2
    return B;
}

```

Figure 5: Substitution Byte Function

```

Nibble SubNibble ( Nibble N )
{
    Base4 r,c
    r,c = N
    N = S-Box[r][c]
    return N
}

```

Figure 6: Substitution Nibble Function

	0	1	2	3
0	21	20	23	10
1	03	32	00	01
2	12	33	30	02
3	11	22	13	31

Table 1: S-Box

c) Row Transformation

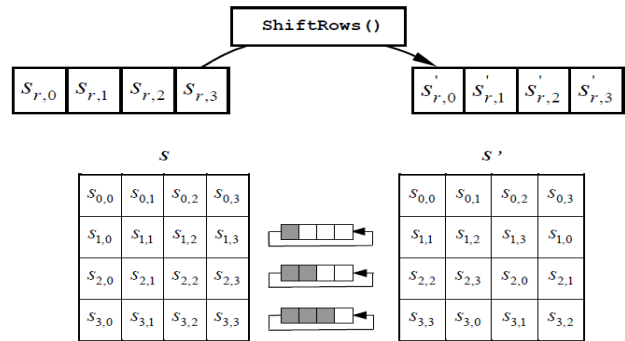


Figure 7: Shift Rows[2] [3]

d) Column Transformation

Mix Columns function is exactly same as in AES algorithm. This is represented as matrix multiplication as shown below.

$$\begin{bmatrix} s'_{0,c} \\ s'_{1,c} \\ s'_{2,c} \\ s'_{3,c} \end{bmatrix} = \begin{bmatrix} 02 & 03 & 01 & 01 \\ 01 & 02 & 03 & 01 \\ 01 & 01 & 02 & 03 \\ 03 & 01 & 01 & 02 \end{bmatrix} \begin{bmatrix} s_{0,c} \\ s_{1,c} \\ s_{2,c} \\ s_{3,c} \end{bmatrix}$$

Figure 8: Matrix Multiplication[2] [3]

Above matrix multiplication is expanded as below. This includes scalar multiplication thus ensuring final value not exceeding 8 bits length.

$$\begin{aligned}
 s'_{0,c} &= (\{02\} \bullet s_{0,c}) \oplus (\{03\} \bullet s_{1,c}) \oplus s_{2,c} \oplus s_{3,c} \\
 s'_{1,c} &= s_{0,c} \oplus (\{02\} \bullet s_{1,c}) \oplus (\{03\} \bullet s_{2,c}) \oplus s_{3,c} \\
 s'_{2,c} &= s_{0,c} \oplus s_{1,c} \oplus (\{02\} \bullet s_{2,c}) \oplus (\{03\} \bullet s_{3,c}) \\
 s'_{3,c} &= (\{03\} \bullet s_{0,c}) \oplus s_{1,c} \oplus s_{2,c} \oplus (\{02\} \bullet s_{3,c}).
 \end{aligned}$$

Figure 9: Scalar Multiplication[2] [3]

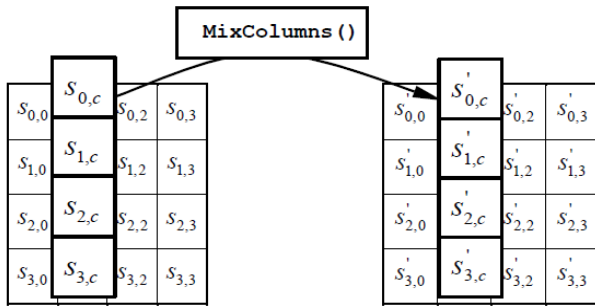


Figure 10: Column Transformation[2] [3]

3) *Decryption*

```

Decryption ( ByteString Input, ByteString Output, ByteString InputKey, int ByteKeyLen )
{
    ByteString State[4][4] = Input
    ByteString RoundKey = KeyExpansion ( InputKey, ByteKeyLen )

    State = AddRoundKey ( State, RoundKey )
    for ( i=0; i<3; i++)
    {
        State = InvMixColumns ( State )
        State = InvShiftRows ( State )
        State = InvSubByte ( State )
        State = AddRoundKey ( State, RoundKey )
    }

    Output = State
}

```

Figure 11: Decryption Function[2] [3]

a) *Add Round Key*

```

void AddRoundKey ( ByteString RoundKey )
{
    k=0;
    for ( i=0; i<4; i++)
    {
        for ( j=0; j<4; j++)
        {
            State[i][j] = State[i][j] XOR RoundKey[k++]
        }
    }

    return ;
}

```

Figure 12: Add Round Key Function[2] [3]

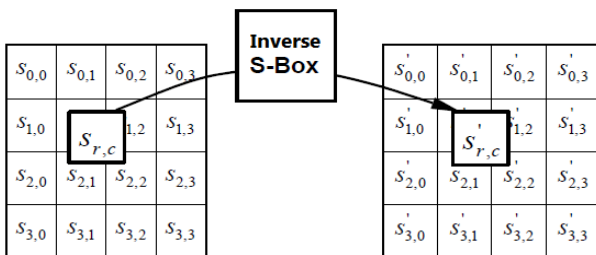
b) *Inverse Substitution Transformation*

Figure 13: Inverse Byte Substitution[2] [3]

```

Byte InvSubByte ( Byte B )
{
    Nibble N1, N2
    N1, N2 = B
    N1 = InvSubNibble ( N1 )
    N2 = InvSubNibble ( N2 )
    B = N1, N2
    return B;
}

```

Figure 14: Inverse Byte Substitution Function

```

Nibble InvSubNibble ( Nibble N )
{
    Base4 r,c
    r,c = N
    N = InvS-Box[r][c]
    return N
}

```

Figure 15: Inverse Nibble Substitution Function

	0	1	2	3
0	12	13	23	10
1	03	30	20	32
2	01	00	31	02
3	22	33	11	21

Table 2: Inverse S-Box

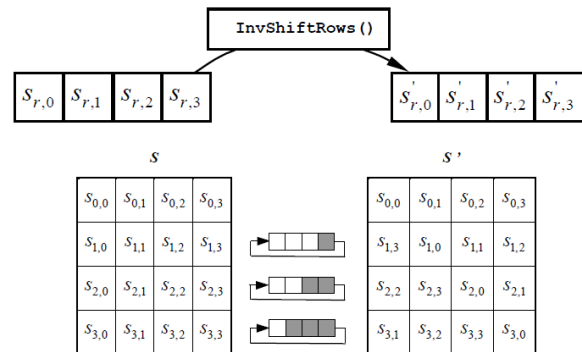
c) *Inverse Row Transformation*

Figure 16: Inverse Row Transformation[2] [3]

d) *Inverse Column Transformation*

Inverse Mix Columns function is exactly same as in AES algorithm. This is represented as matrix multiplication as shown below.

$$\begin{bmatrix} s'_{0,c} \\ s'_{1,c} \\ s'_{2,c} \\ s'_{3,c} \end{bmatrix} = \begin{bmatrix} 0e & 0b & 0d & 09 \\ 09 & 0e & 0b & 0d \\ 0d & 09 & 0e & 0b \\ 0b & 0d & 09 & 0e \end{bmatrix} \begin{bmatrix} s_{0,c} \\ s_{1,c} \\ s_{2,c} \\ s_{3,c} \end{bmatrix}$$

Figure 17: Matrix Multiplication[2] [3]

Above matrix multiplication is expanded as below. This includes scalar multiplication thus ensuring final value not exceeding 8 bits length.

$$\begin{aligned} s'_{0,c} &= (\{0e\} \bullet s_{0,c}) \oplus (\{0b\} \bullet s_{1,c}) \oplus (\{0d\} \bullet s_{2,c}) \oplus (\{09\} \bullet s_{3,c}) \\ s'_{1,c} &= (\{09\} \bullet s_{0,c}) \oplus (\{0e\} \bullet s_{1,c}) \oplus (\{0b\} \bullet s_{2,c}) \oplus (\{0d\} \bullet s_{3,c}) \\ s'_{2,c} &= (\{0d\} \bullet s_{0,c}) \oplus (\{09\} \bullet s_{1,c}) \oplus (\{0e\} \bullet s_{2,c}) \oplus (\{0b\} \bullet s_{3,c}) \\ s'_{3,c} &= (\{0b\} \bullet s_{0,c}) \oplus (\{0d\} \bullet s_{1,c}) \oplus (\{09\} \bullet s_{2,c}) \oplus (\{0e\} \bullet s_{3,c}) \end{aligned}$$

Figure 18: Scalar Multiplication[2] [3]

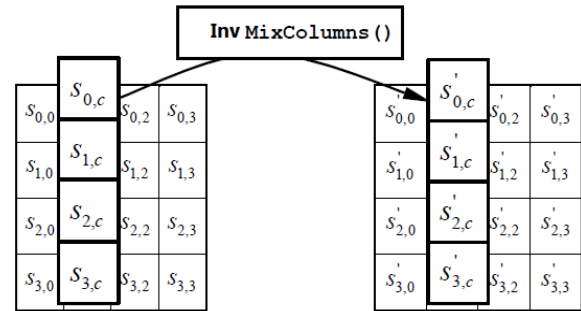


Figure 19: Inverse Column Transformation[2] [3]

III. COMPARISON OF AES WITH PROPOSED ALGORITHM (LMEP-S-AES)

	AES	LMEP-S-AES
Key Length	128, 192, 256 Bits	Variable Length (Max 128 bits)
Number of Rounds	10, 12, 14	3
Round Key	Different for each round	Same for all rounds
Key Expansion	Complex combination of SubByte() and Scalar Multiplication	Simplified
S-Box size	16x16	4x4
S-Box Memory	256 Bytes	8 Bytes
Inverse S-Box Memory	256 Bytes	8 Bytes

Table 3: Comparison between AES and LMEP-S-AES

IV. ANALYSIS OF LMEP-S-AES

1) Memory Optimization

Embedded processors need applications to be running in low memory constraints.

- The proposed AES optimizes memory requirements by reducing S-Box and Inverse S-Box. Together S-Box needs 16 Bytes in comparison to 512 Bytes of actual AES algorithm.

- Same key is used for all rounds, thus reducing memory consumption of expanded key like in AES algorithm.

2) Processing Optimization

Embedded processors are generally low end processors with processing power constraints.

- The proposed algorithm optimizes processing power requirements by reducing number of rounds to 3.
- Key Expansion algorithm is simplified.

- III. Additional computations needed in choosing round key for each round is reduced, since same key is used for all rounds.

V. CONCLUSION

Embedded systems like Mobile phones, GPS receivers, Wireless Sensor Nodes etc handle sensitive data, hence requires data security mechanisms. AES algorithm which is a standard algorithm for data encryption is unsuitable for such scenarios where memory and processing power constraints are very high. The proposed Simplified AES for Low Memory Embedded Processors" algorithm is optimized to overcome these 2 constraints.

VI. REFERENCES

- 1) ~~AA~~ Embedded Software", Edward A. Lee, Advances in Computers, Academic Press London 2002
- 2) ~~AA~~ Advanced Standard Encryption" Federal information Processing Standards Publication 197
- 3) ~~AA~~ Merging of RC5 with AES - Incorporating more Flexibility and Security in AES" ICIT 2009, ISBN: 978-0-07-068014-2

Automatic Ontology Creation by Extracting Metadata from the Source code

Gopinath Ganapathy¹, S. Sagayaraj²

GJCST Classification
D.4.3, D.2.3, D.2.4

Abstract-Semantic Web can be created by developing Ontologies. For every new project Software companies are going for designing new code and components, by new developers. If the company archives the completed code and components, which can be used with no need for testing it unlike open source code and components. File metadata and file content metadata can be extracted from the Application files and folders using API's. The extracted components can be stored in the Hadoop Distributed File System along with the application environment. Extracted metadata will be in XML format. XML deals with syntactic level and the Web Ontology Language (OWL) supports semantic level for the representation of domain knowledge using classes, properties and instances. This paper converts the data model elements of XML to OWL Ontology that implements the mapping the standard XML technology XSLT.

Keywords-Metadata, Parsing, Extensible Markup Language, Hadoop Distributed File System, Hbase, Web Ontology Language and Ontology.

I. INTRODUCTION

Most of the today's Web content is suitable for human consumption. An alternative approach is to represent Web content in a form that is more easily machine-processable by using intelligent techniques. The machine processable Web is called the Semantic Web. Semantic Web will not be a new global information highway parallel to the existing World Wide Web; instead it will gradually evolve out of the existing Web [1]. Ontologies are built in order to represent generic knowledge about a target world [2]. Ontology is more than a rule base of terminological problems and is worth to consider a promising methodology in the Semantic Web. In the semantic web, ontologies can be used to encode meaning into a web page, so that intelligent agents can understand what the web page is about. Ontologies increase the efficiency and consistency of describing resources, by enabling more sophisticated functionalities in development of knowledge management and information retrieval applications. Knowledge management concerns itself with acquiring, accessing, and maintaining knowledge within an organization. It has emerged as a key activity of large businesses because they view internal knowledge as an intellectual asset from which they can draw greater

productivity, create new value and increase their competitiveness. From the knowledge management perspective, the current technology suffers in searching, extracting, maintaining and viewing information. The aim of the Semantic Web is to allow much more advanced knowledge management system. Software firms work on various projects for many years with little amount of archiving. If the projects are achieved properly the reusability of the available components will make the software development much easier and faster. This paper proposing an automatic generation of knowledge management by extracting metadata from the source code of the project files by using the Source code Documenter. The output will be converted into XML. Today XML has reached a wide acceptance as data exchange format. The aim of this paper is to bridge the gap from metadata, XML and OWL. A strategy for how OWL ontologies may be generated automatically out of existing XML data is proposed. This has to be done by establishing suitable mappings between the different data model elements of XML and OWL. All the project files, XML and OWL files are stored in Hbase repository. By storing code components the new project development can search for component in the OWL ontology and retrieve from the repository and using that will reduce the cost, manpower, time, testing, etc., in the software development. The paper begins with the discussion on the related technologies for this paper in Section 2. Then, detailed features and framework for the automatic metadata extraction is found in Section 3. The conversion from XML to OWL Ontology implements the mapping in the standard XML technology XSLT is presented in section 4. The implementation scenario is presented in Section 5. Section 6 deals with the findings of the paper.

II. RELATED WORK

1) Metadata

Metadata is defined as data "about data" or descriptions of stored data. Metadata definition is about defining, creating, updating, transforming, and migrating all types of metadata that are relevant and important to a user's objectives. Some metadata can be seen easily by users, such as file dates and file sizes, while other metadata can be hidden. Metadata standards include not only those for modeling and exchanging metadata, but also the vocabulary and knowledge for ontology [3]. A lot of efforts have been made to standardize the metadata but all these efforts belong to some specific group or class. The Dublin Core Metadata Initiative (DCMI) [4] is perhaps the largest candidate in

About¹- F.A. Author is Professor & Head with the Department of Computer Science, Bharathidasan University, Trichy, Tamil Nadu, India (e-mail: gganapathy@gmail.com)

About²- S.B Author is Associate Professor with the Department of Computer Science, Sacred Heart College, Tirupattur, Vellore, Tamil Nadu, India (Mobile: +91 9443035624; fax: +91 4179 220553; e-mail: sagi_sara@yahoo.com)

defining the Metadata. It is simple yet effective element set for describing a wide range of networked resources and comprises 15 elements. Dublin Core is more suitable for document-like objects. IEEE LOM [5], is a metadata standard for Learning Objects. It has approximately 100 fields to define any learning object. Medical Core Metadata (MCM) [6] is a Standard Metadata Scheme for Health Resources. MPEG-7 [7] multimedia description schemes provide metadata structures for describing and annotating multimedia content. Standard knowledge ontology is also needed to organize such types of metadata as content metadata and data usage metadata. These standards may be adopted in full or in part. Further, appropriate procedures need to be defined and followed within the enterprise in documenting the capture, update, transformation, migration, replication of metadata and relevant transformation rules.

2) HDFS & HBASE

The Hadoop project promotes the development of open source software and it supplies a framework for the development of highly scalable distributed computing applications [8]. Hadoop is designed in such a way to efficiently process large volumes of information. It connects many commodity computers together so that they could work in parallel. It is a simplified programming model which allows the user to write and test distributed systems quickly. In a Hadoop cluster even while, the data is being loaded in, it is distributed to all the nodes of the cluster. The Hadoop Distributed File System (HDFS) will break large data files into smaller parts which are managed by different nodes in the cluster. Each part is replicated across several machines, so that a single machine failure does not lead to non-availability of any data. Even though the file parts are replicated and distributed across several machines, they form a single namespace, so their contents are universally accessible. MapReduce [9] is a functional abstraction which provides an easy-to-understand model for designing scalable, distributed algorithms. HDFS, Hbase is the Hadoop application to use when real-time random accesses to very large datasets are required. It supports both batch-style computations using MapReduce and point queries. It is built from the ground-up to scale linearly just by adding nodes. Hbase is not relational and does not support SQL, but given the proper problem space it is able to do what an RDBMS cannot. The reason for using HDFS and Hbase in this paper is that the software size in a project is growing exponentially and also takes care of volumes data for processing. To handle data and code is stored in Hbase and it is distributed using HDFS. Also the size of a project and the number of project for a three decade company needs terabytes of volume to be stored which needs Hbase.

3) Ontology

The key component of the Semantic Web is the collections of information called ontologies. Ontology is a term borrowed from philosophy that refers to the science of describing the kinds of entities in the world and how they are related. Gruber defined ontology as a specification of a

conceptualization [10]. Ontology defines the basic terms and their relationships comprising the vocabulary of an application domain and the axioms for constraining the relationships among terms [11]. This definition explains what an ontology looks like [12]. The most typical kind of ontology for the Web has taxonomy and a set of inference rules. The taxonomy defines classes of objects and relations among them. Classes, subclasses and relations among entities are a very powerful tool for Web use. A large number of relations among entities can be expressed by assigning properties to classes and allowing subclasses to inherit such properties. Inference rules in ontologies supply further power. Ontology may express rules on the classes and relations in such a way that a machine can deduce some conclusions. The computer does not truly "understand" any of this information, but it can now manipulate the terms much more effectively in ways that are useful and meaningful to the human user. More advanced applications will use ontologies to relate the information on a page to the associated knowledge structures and inference rules.

III. EXTRACTION FRAMEWORK

After the completion of a project all the project files are sent to the metadata extraction framework. The files of a project can be asked to give in form of zip file which contains various types of files. The framework will extract files and give it to corresponding Document Generator tools and it will convert the file metadata into XML file. This paper deals with java files only. To extract file content metadata the source code is processed by Javadoc. The retrieved components will be stored in the HBase, which is the subproject of Hadoop. The project may contain files in various languages and tools. Similarly there are many parser tools available to parse the source code from various languages and tools to either HTML or XML output. So the framework will identify the type of project file and supply the file to the corresponding parser. Two types of metadata will be extracted. First is the file metadata and it contains file attributes, file size, parent, path, etc. The second type of metadata contains the contents of the source file. The file content metadata will be different for different types of languages and tools. Here the framework presents the general form for all type of project files. The framework output shown is for the extracted metadata into XML file. The available Document Generator converting programs from one language or tool to HTML or XML file is presented in Table 1. But the framework takes care of the languages and tools that support XML format. The framework needs only few tools which will be available in the HDFS to support to handle various types of files. After the converting the source files into XML, files they are stored in HBase along with the Project files. The purpose of storing the XML files, source file and file attributes and file components in Hbase is to retrieve the code components and to reduce the software development cost. The components will be stored in Hbase by creating database design with the fields project name, project leader, period of project and code components. Several Sample Document Generator

strategies for extracting metadata from source file have been proposed. Document generator which takes care of XML

alone is discussed and used in this framework.

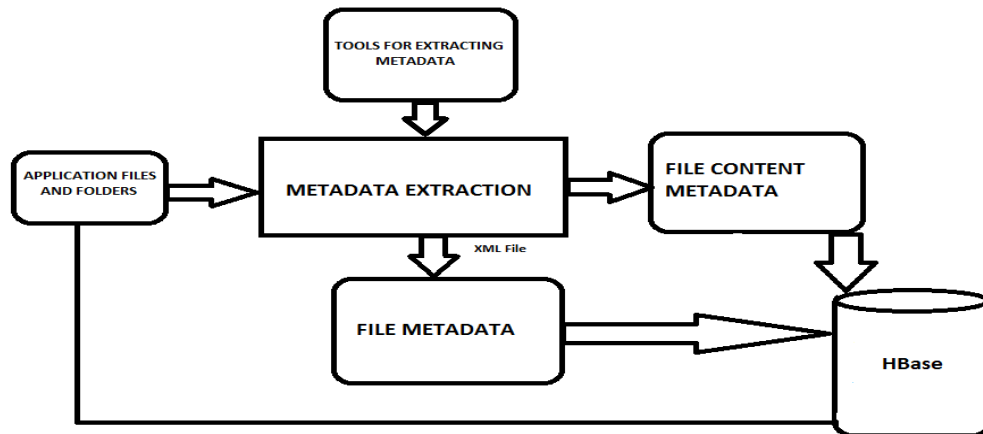


Fig.1. Metadata extraction Framework

1) Doxygen

Doxygen is a documentation for C++, C, Java, Objective-C, Fortran, VHDL, PHP and C# [13]. Doxygen supports the documentation tags used in the Qt toolkit and can generate output in HTML as well as in CHM, RTF, PDF, LaTeX, PostScript or man pages.

2) Ddoc

Ddoc is the embedded for the D programming language designed by Walter Bright. Its emphasis is on being able to write documentation in code comments in a natural style, minimizing the need for embedded markup and thus, improving the legibility of the code comments[14]. Code comments are associated with symbols in the code, and Ddoc uses the semantic and syntactic information available from the D compiler to fill in routine information such as parameters and return types automatically. Ddoc can generate output in XML and XHTML.

3) FPDdoc

FPDdoc is a tool that combines a Pascal unit file and a description file in XML format and produces reference documentation for the unit. The reference documentation contains documentation for all of the identifiers found in the unit's interface section[15]. FPDdoc does not require the presence of formatted comments in the source code. It takes a source file and a documentation file (in XML format) and merges these two together to a full documentation of the source. This means that the code doesn't get obfuscated with large pieces of comment, making it hard to read and understand.

4) JSDoc

JSDoc is syntax for adding inline API documentation to JavaScript source code. The JSDoc syntax is similar to the

Javadoc syntax, used for documenting Java code, but is specialized to work with JavaScript's more dynamic syntax

and therefore unique, as it is not completely compatible with Javadoc[16]. However, like Javadoc, JSDoc allows the programmer to create doclets and tags which can then be translated into published output, like HTML or RTF.

5) PHPDocumentor

PHPDocumentor is an open source documentation written in PHP. It automatically parses PHP source code and produces readable API and source code documentation in a variety of formats. PHPDocumentor generates documentation based on PHPDoc-formatted comments and the structure of the source code itself. PHP Documentor can create documentation in HTML, PDF, CHM or Docbook formats. PHPDocumentor is able to parse all PHP syntax and supports PHP4 and PHP5.

IV. CONVERTING XML TO OWL

The conversion process elevates the XML Schema Definition (XSD) to the level of OWL ontology. XML documents contain relational structure and represent those using OWL classes, properties and instances. The XML data model describes a node labeled tree [17], on the other hand OWL's data model is based on the subject-predicate-object triples from RDF[18]. RDF-Schema defines a vocabulary for creating class hierarchies, attaching properties to classes and adding instance data [19]. The framework creates a corresponding class hierarchy by exploiting the tree structure of XML. Restrictions like cardinality constraints and properties can be expressed easily when OWL is on the top of RDF and RDFS. This enables the representation of Source code data in OWL. The mapping is implemented in XML Stylesheet Language Transformations (XSLT) and it is interoperable with different programming languages [20]. As for as XML data is concerned no XML schema is attached. The conversion process generates a suitable intermediate XML schema definition for the XML. The overall architecture of the framework is shown in Fig. 2.

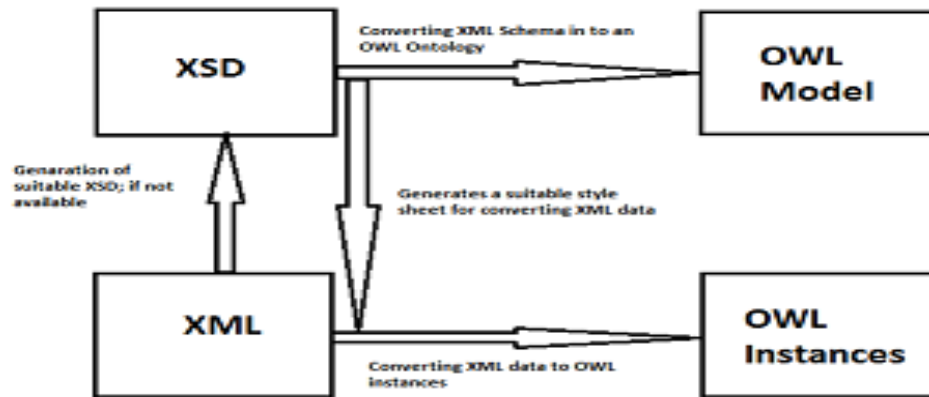


Figure.2. Conversion process from XML to OWL.

The conversion process requires at the most three steps for XML instance data and at least one step for XML Schema only. To process XML Schema only, the model of the ontology is created along with classes and properties. Intermediate step is created for XML instance data only. The first step extracts an XML Schema out of the XML instance data. [21] states that in every XML instance document an XML Schema is exist implicitly. Since XML instance document do not contain as much information about constraints as in manually created XML Scheme, automatically generated XML Schema could be incomplete. There are also several XML Schema components and they cannot be discovered via stylesheet driven extraction. XML instance documents can contain optional elements or attributes, which are not in the sample document. Thus they are not in the generated XML Schema and OWL ontology. In XML Schema extraction a preferably representative XML instance document is needed, so that the XML Schema can serve as a good basis. This XML Schema extraction is based upon an XSLT stylesheet [22], which is extended and adapted to this framework. The majority of existing XML to OWL mapping tools use XSLT to accomplish the mapping [23]. A stylesheet is created simultaneously while converting XML to OWL and this stylesheet will converts XML instance data into the instances part of the ontology. This stylesheet will be configured automatically to adjust the transformation process of the instances to the OWL model. It determines whether elements become classes or properties. That is necessary, because the XML instance data can have optional elements or attributes and the created stylesheet will be their common denominator. To support the separation of model and data, the OWL model will be stored separately from the OWL instances. The OWL instances will be connected to their model using the owl:

import property. Therefore every OWL instance, which refers to the OWL model, will obtain an adjustable

namespace prefix. A further stylesheet is generated automatically for the conversation of XML instance data to OWL instances. The framework is designed to be easily extensible, so that the support for the missing XSD components can be included and a better support for document oriented XML, can be integrated. The conversion from XML to OWL uses Classes (owl:Class) which will emerge from xsd:complexType and xsd:elements according to the following rules: For the case, that an element in the source XML tree is always a leaf, containing only a literal and no attributes. This element will be mapped to an owl:DatatypeProperty having as domain the class which represents the surrounding element. XML attributes will be handled equally which means it is mapped to owl:DatatypeProperties, too. XML Schema also can contain arity constraints like xsd:minOccurs or xsd:maxOccurs, which is mapped to the equivalent cardinality constraints in OWL, owl:minCardinality and owl:maxCardinality. Table 2 summarizes the mapping based on the correspondence between XML schema elements and OWL classes and properties. OWL ontology created by this process is a full proof ontology without any discrepancy for the given source code without leaving a single piece of information. The code will also be stored in the database for future code extraction. The entire conversion process uses the concepts of XSD and OWL to convert XML to XSD using a transform tool and XSD to OWL Ontology manually. Conversion of XSD to OWL Ontology is also possible to using a tool.

V. CASE STUDY

To evaluate the present frame work a simple java code is used, to extract the file and file contents metadata into a XML file.

Source code in java

```

public class Calculator
{
    public int add(int i1, int i2)
    {
        return i1 + i2;
    }
    public int subtract(int i1, int i2)
    {
        return i1 - i2;
    }
}

```

The java sample code is given to Doxygen Document generator tool that extracts the metadata extracted from file and file content. The contents of the file metadata are Title, Author, Owner, Year, Month, Pages, Number of Bytes, File permission attributes, Date, time, etc., are stored in an XML file. In the same way Javadoc is used to extract the metadata from the source code contains Function, Definition, Type, Arguments, Brief Description, Member Definition, Declaration Name, Parameters, etc., are stored in HBase

```

<compounddef id="classCalculator"
kind="class" prot="public">

<compoundname>Calculator</compoundname>

    <sectiondef kind="public-func">
        <memberdef kind="function"
<type>int</type>
        <definition>int
Calculator::add</definition>
        <argsstring>(int i1, int
i2)</argsstring>
        <name>add</name>
        <param>
            <type>int</type>
            <declname>i1</declname>
        </param>
        <param>
            <type>int</type>
            <declname>i2</declname>
        </param>
        <location
        </memberdef>
        <memberdef kind="function"
<type>int</type>
        <definition>int
Calculator::subtract</definition>
        <argsstring>(int i1, int
i2)</argsstring>
        <name>subtract</name>
        <param>
            <type>int</type>
            <declname>i1</declname>
        </param>
        <param>
            <type>int</type>
            <declname>i2</declname>
        </param>
        <location
        </memberdef>
    </sectiondef>

```

```

<location
    <listofallmembers>

<scope>Calculator</scope><name>add</na
me></member>
    <member
<scope>Calculator</scope><name>subtrac
t</name></member>
    </listofallmembers>
</compounddef>
</doxygen>

```

The goal is to transform this syntactic XML file into an OWL file by creating instances of the classes with the XSLT Stylesheet. The conversion from XML to XSD is carried out using Trang, which is open source software. The XSD for the given XML file is shown below

```

<xs:complexType>
    <xs:sequence>
        <xs:element ref="compounddef"/>
    </xs:sequence>
    <xs:attribute name="version"
use="required" type="xs:NMTOKEN"/>
    <xs:attribute
ref="xsi:noNamespaceSchemaLocation"
use="required"/>
    </xs:complexType>
</xs:element>
<xs:element name="compounddef">
    <xs:complexType>
        <xs:sequence>
            <xs:element ref="compoundname"/>
            <xs:element ref="sectiondef"/>
            <xs:element ref="briefdescription"/>
            <xs:element
ref="detaileddescription"/>
            <xs:element ref="location"/>
            <xs:element ref="listofallmembers"/>
        </xs:sequence>
        <xs:attribute name="id" use="required"
type="xs:NCName"/>
        <xs:attribute name="kind"
use="required" type="xs:NCName"/>
        <xs:attribute name="prot"
use="required" type="xs:NCName"/>
    </xs:complexType>
</xs:element>
    <xs:element name="compoundname"
type="xs:NCName"/>
    <xs:element name="sectiondef">
        <xs:complexType>
            <xs:sequence>
                <xs:element maxOccurs="unbounded"
ref="memberdef"/>
            </xs:sequence>
            <xs:attribute name="kind"
use="required" type="xs:NCName"/>
        </xs:complexType>
    </xs:element>
    <xs:element name="memberdef">
        <xs:complexType>
            <xs:sequence>
                <xs:element ref="type"/>
                <xs:element ref="definition"/>
                <xs:element ref="argsstring"/>
                <xs:element ref="name"/>
            </xs:sequence>
        </xs:complexType>
    </xs:element>

```

```

    <xs:element maxOccurs="unbounded"
ref="param"/>
    <xs:element ref="briefdescription"/>
    <xs:element
ref="detaileddescription"/>
    <xs:element ref="inbodydescription"/>
    <xs:element ref="location"/>
  </xs:sequence>
  <xs:attribute name="const"
use="required" type="xs:NCName"/>
  <xs:attribute name="explicit"
use="required" type="xs:NCName"/>
  <xs:attribute name="id"
use="required"/>
  <xs:attribute name="inline"
use="required" type="xs:NCName"/>
  <xs:attribute name="kind"
use="required" type="xs:NCName"/>
  <xs:attribute name="prot"
use="required" type="xs:NCName"/>
  <xs:attribute name="static"
use="required" type="xs:NCName"/>
  <xs:attribute name="virt"
use="required" type="xs:NCName"/>
</xs:complexType>
</xs:element>
<xs:element name="definition"
type="xs:string"/>
<xs:element name="argsstring"
type="xs:string"/>
<xs:element name="param">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="type"/>
      <xs:element ref="declname"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="declname"
type="xs:NCName"/>
<xs:element name="inbodydescription">
  <xs:complexType/>
</xs:element>
<xs:element name="listofallmembers">
  <xs:complexType>
    <xs:sequence>
      <xs:element maxOccurs="unbounded"
ref="member"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="member">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="scope"/>
      <xs:element ref="name"/>
    </xs:sequence>
    <xs:attribute name="prot"
use="required" type="xs:NCName"/>
    <xs:attribute name="refid"
use="required"/>
    <xs:attribute name="virt"
use="required" type="xs:NCName"/>
  </xs:complexType>
</xs:element>
<xs:element name="scope" type="xs:NCName"/>

```

```

<xs:element name="type" type="xs:NCName"/>
<xs:element name="name" type="xs:NCName"/>
<xs:element name="briefdescription">
  <xs:complexType/>
</xs:element>
<xs:element name="detaileddescription">
  <xs:complexType/>
</xs:element>
<xs:element name="location">
  <xs:complexType>
    <xs:attribute name="bodyend"
use="required" type="xs:integer"/>
    <xs:attribute name="bodyfile"
use="required"/>
    <xs:attribute name="bodystart"
use="required" type="xs:integer"/>
    <xs:attribute name="file"
use="required"/>
    <xs:attribute name="line"
use="required" type="xs:integer"/>
  </xs:complexType>
</xs:element>
</xs:schema>

```

The implementation transforms XML to OWL using XSLT [24]. Applying the previous XSD Stylesheet to the XML file, the OWL Ontology presented below is created manually

```

<Project rdf:ID="CalPrj">
<hasPackages>
  <Package rdf:ID="default">
<hasClasses>
  <Class rdf:ID="Calculator">
    <hasMethods>
<Method rdf:ID="add">
<hasParameters>
  <Parameter rdf:ID="i1">
    <Name >i1</Name>
<DataType >int</DataType>
</Parameter>
</hasParameters>
<hasParameters>
<Parameter rdf:ID="i2">
<DataType >int</DataType>
<Name >i2</Name>
</Parameter>
</hasParameters>
<Identifier >public</Identifier>
<Name >add</Name>
</Method>
</hasMethods>
<hasMethods>
<Method rdf:ID="subtract">
<Name >subtract</Name>
<hasParameters rdf:resource="#i1"/>
<hasParameters rdf:resource="#i2"/>
<Returns >int</Returns>
<Identifier >public</Identifier>
</Method>
</hasMethods>
<Name
>Calculator</Name>
<Identifier
>public</Identifier>

```

```

</Class>
</hasClasses>
<Name >default</Name>
</Package>
</hasPackages>
</Project>
</rdf:RDF>

```

The code components stored in the HBase will be linked to the method signature in the OWL ontology for retrieval purpose. The components will be reused for the new project appropriately. The obtained OWL Ontology successfully loads on both Protégé Editor and Altova Semantics Works.

VI. CONCLUSION

The paper presents an approach for generating ontologies automatically out of existing XML data from source code. This approach helps to integrate conventional XML and source code into the Semantic Web. OWL is semantically much more expressive than needed for the results of our mapping. With these sample tests the paper argues that it is indeed possible to transform XML structures into OWL ones using XSLT. As the source XML structure and Ontology complexity increases, the XSLT stylesheet will get more and more complex. The purpose of the paper is to achieve the code reusability for the software development. By creating OWL ontology for the source code the future will be to search and extract the code and components and reuse to shorten the software development life cycle. Open source code can also used to create OWL Ontology so that there will be huge number of components which can be reused for the development.

VII. REFERENCES

- 1) Grigoris Antoniou and Frank van Harmelen, A Semantic Web Primer", PHI Learning Private Limited, New Delhi, 2010, pp 1-3
- 2) Bung. M, Treatise on Basic Philosophy. Ontology I. The Furniture of the World. Vol. 3, Boston: Reidel.
- 3) Won Kim: On Metadata Management Technology Status and Issues", in *Journal of Object Technology*, vol. 4, no. 2, 2005, pp. 41-47
- 4) Dublin Core Metadata Initiative. <
<http://dublincore.org/documents/>>,2002.
- 5) IEEE Learning Technology Standards Committee, <http://ltsc.ieee.org/wg12>, IEEE Standards for Learning Object Metadata (1484.12.1)
- 6) Darmoni, Thirion, Metadata Scheme for Health Resources" *American Medical Informatics Association*, 2000 Jan-Feb; 7(1): 108-109.
- 7) MPEG-7 Overview: ISO/IEC JTC1/SC29/WG11 N4980, Kla-genfurt, July 2002.
- 8) Jason Venner, Pro Hadoop : Build Scalable, distributed applications in the cloud, Apress, 2009.
- 9) Tom White, Hadoop: The Definitive Guide, O'Reilly Media, Inc., 2009.
- 10) Gruber, T. What is an Ontology? (September, 2005): <http://www.ksl-stanford.edu/kst/what-is-an-ontology.html>.

- 11) Yang, X. Ontologies and How to Build Them. 2001. (March, 2006): <http://www.ics.uci.edu/~xwy/publications/area-exam.ps>.
- 12) Bugaite, D., O. Vasilecas. Ontology-Based Elicitation of Business Rules. In A. G. Nilsson, R. Gustas, W. Wojtkowski, W. G. Wojtkowski, S. Wrycza, J. Zupancic Information Systems Development: Proc. of the ISD'2004. Springer- Verlag, Sweden, 2006, pp. 795-806.
- 13) Peter Toft, Introduction to Doxygen 1994
- 14) Gray Davis, REVISED PRESIDING MEMBERS PROPOSED DECISION,2000
- 15) Michaël Van Canneyt, Free Pascal code documenter: Reference manual, 2000
- 16) Bastian Feder, The Beauty and the Beast, 2009
- 17) Bert Bos. The XML data model. <http://www.w3.org/XML/Datamodel.html>, 1997.
- 18) Graham Klyne, Jeremy J. Carroll, and Brian McBride. Resource Description Framework (RDF): Concepts and Abstract Syntax. Technical report, W3C, <http://www.w3.org/TR/2002/WD-rdf-concepts-20021108/>, 2002.
- 19) Dan Brickley and R.V. Guha. RDF Vocabulary Description Language 1.0: RDF Schema. Technical report, W3C,<http://www.w3.org/TR/2002/WD-rdf-schema-20021112/>, 2002.
- 20) James Clark. XSL Transformations (XSLT). Technical report, W3C, <http://www.w3.org/TR/xslt>, 1999.
- 21) Stefan Mintert. Schlussselqualifikation; XML jenseits des Mainstreams. iX, 8:48-51, 2005.
- 22) Charlie Halpern-Hamu. Transform a sample instance to a schema. <http://incrementaldevelopment.com/papers/xslttrick/>,1999.
- 23) XSLT Quickly, Bob DuCharme, Manning Publications, ISBN 1- 930110-11-1.
- 24) Mastering XML Transformations, Doug Tidwell, O'REILLY, ISBN 0596000537.

A Simplified NMEA Sentence Generator for the Simulation of GPS Tracking

G. Rajendran¹, Dr. M. Arthanari², M. Sivakumar³

GJCST Classification
J.7

Abstract-The applications of GPS (Global Positioning System) tracking in real time have found its place in almost all walks of life, for instance, navigation, map making, land surveying, fishing and trekking. GPS tracking is defined as the measurement of object position and orientation in a given coordinate system using GPS data. GPS data can be used for tracking objects in real-time. When fixed with moving objects, GPS receivers are used to track them by generating GPS data in the form of standard NMEA (National Marine Electronics Association) sentences. One of the most important NMEA sentences is the recommended minimum sentence C, \$GPRMC, which provides the essential GPS PVT (position, velocity, time) data. This \$GPRMC sentence is used for the tracking of objects. But for doing research with GPS data, multiple GPS receivers are needed to generate GPS data in NMEA format, which involves huge cost. The simplified NMEA sentence generator is a software which generates the NMEA formatted GPS data i.e., \$GPRMC sentence without using a GPS receiver. This generator is useful to create a database of GPS data and simulate an environment of GPS tracking without using actual GPS receivers.

Keywords-GPS receiver, GPS simulation, GPS data, NMEA sentence, Real time tracking.

I. INTRODUCTION

In recent years, many companies are developing GPS [1] Enabled applications and systems, because GPS has many technical and economical benefits. Countries are spending millions of dollars for research and development of GPS applications as it has enabled them more powerful in the war and defence front. More research is being done in this domain as it has a unique capability of locating any GPS enabled system over the earth in terms of latitude, longitude and even altitude with good accuracy. Researchers need GPS databases which contain huge volume of GPS data generated by thousands of GPS receivers. But the presently available GPS databases are owned by private players and are not available for use by the researchers. This introduces a new problem domain of non-availability of sample data to develop and test GPS applications. Hence the need for a GPS data generator which simulates a GPS guided tracking environment is vital. Such a simulator will greatly reduce the expenses in research area and enable us to have more researches carried out in the allocated budgets. GPS simulation has already been carried

out by some of the researchers, but with limitations like hardware dependency, involvement of certain cost, complexity, lack of provision for GPS database creation and lack of provision for track replaying. These limitations have been addressed in this work. This paper presents a simplified NMEA sentence generator which produces GPS data for positioning of moving objects. This NMEA sentence generator generates the sentence \$GPRMC. Since the NMEA sentence, \$GPRMC is the only GPS data dealt in this paper, the words \$GPRMC and GPS data are used interchangeably. The generated GPS data is used to simulate a GPS guided tracking environment. This generator produces GPS PVT data and plots the position using 2D graphics. The remainder of this paper is organised as follows. Section 2 of this paper describes the GPS guided tracking systems. Previous work in this area is discussed in Section 3. In Section 4, the GPS data generator and the simulation scenario are introduced. The results of the simulation along with a comparison of output data items with standard and real-time data are dealt in Section 5. The work is concluded and the possible improvements are discussed in Section 6.

II. GPS GUIDED TRACKING SYSTEMS

1) Global Positioning System

GPS is a Satellite Navigation System which is funded and controlled by the U. S. Department of Defence [2, 3]. The GPS system consists of three segments viz., satellites that transmit the position information, the ground stations that are used to control the satellites and update the information, and finally there is the receiver that computes its location anywhere in the world based on information it gets from the satellites[4]. The satellite segment of GPS consists of a minimum of 21 satellites and 3 working spares totalling 24. There are often more than 24 operational satellites as new ones are launched to replace older satellites. The satellites are travelling around the world 20,200 kilometres high in carefully controlled orbits at a speed that means they will make a complete orbit twice a day. The GPS satellites broadcasts two signals, PPS (Precise Positioning Service) and SPS (Standard Positioning Service)[5]. The PPS signal is an encrypted military access code which is available for use by military and government. The SPS signal is an unencrypted, spread spectrum signal broadcasts at 1575.42 MHz and is available for public use.

The Control Segment consists of a system of tracking stations located around the world. The Master Control facility is located in Colorado. These control stations measure signals from the satellites which are incorporated

About -¹Assistant Professor of Computer Science, Government Arts College (Autonomous), Salem-636007, Tamilnadu, India. (email : guru.rajendran@yahoo.com)

About-²Director, Bharathidasan School of Computer Applications, Ellispettai-638116, Tamilnadu, India. (email : arthanarimsvc@gmail.com)

About-³Doctoral Research Scholar, Anna University, Coimbatore, Tamilnadu, India. (email : sivala@gmail.com)

into orbital models which in turn compute precise orbital data and satellites clock corrections for each satellite. The Master Control station uploads ephemeris and clock data to the satellites. The satellites then send subsets of the orbital ephemeris data to GPS receivers over radio signals. The Receiver Segment consists of GPS receivers which are used for navigation, positioning, time dissemination, and other research. The GPS receivers listen to signals from a minimum of either three or four satellites at a time and triangulate a position fix using the interval between the transmission and reception of the satellite signal. Any given receiver receives signals from more satellites than are

actually needed for a position fix. The reason for this is that if one satellite becomes unavailable, the receiver knows exactly where to find the best possible replacement. Three satellites are required for two dimensional positioning. Two dimensional positioning reports position only in terms of latitude and longitude. Four satellites are required for three dimensional positioning, that is to say position and elevation in terms of latitude, longitude and altitude. In general, an SPS receiver can provide position information with an error of less than 10 meters, and velocity information with an error of less than 5 meters per second.



Figure 1. GPS tracking with the help of a GPS receiver fixed in a moving object

Most of the tracking system combines GPS. In less than ten years since the first GSM network was commercially launched, it became the world's leading and fastest growing mobile standard, spanning over 200 countries. There is at least one cell tower in every 900m-1000m radius in the high traffic regions (city limits) and in the low traffic regions (high ways) a single cell tower can cover a radius up to 10 km. The operation of GPS tracking is explained in Fig. 1. The GPS receiver captures position data from the satellites, computes the position of the object, say, a vehicle, and sends this information to a central base station, using SMS. If the optional storage module is installed, then location data can be stored when the vehicle is out of range of the cellular operator and retrieved later. GSM technology is used to transmit this information which in turn is collected by the server at the base station. The geographical position of the object can be displayed at the base station using a suitable application.

2) NMEA Specification

GPS receiver communication is defined with NMEA specification. The NMEA has developed a specification that defines the interface between various pieces of marine electronic equipments. The NMEA standard permits marine electronics to send information to computers and to other marine equipments [6] in predefined formats. Most computer programs that provide real time position

information recognize data that are in NMEA format which includes the complete latitude, longitude, velocity and time computed by the GPS receiver. In NMEA specification system, data is sent as a line of text, called a sentence which is totally self contained and independent from other sentences. The precision of data may vary in the message. For example latitude or longitude may contain 3 or even 4 digits after the decimal point. There are standard sentences for each device category and there is also the ability to define proprietary sentences for use by the individual company. All of the standard sentences have a two letter prefix that defines the device that uses that sentence type which is followed by a three letter sequence that defines the sentence contents. The NMEA sentences for GPS receivers have the prefix GP. Each sentence contains a maximum of 80 characters of visible text, begins with a '\$' and ends with a carriage return/line feed sequence. The data is contained within this single line and the data items are separated by commas. The commas act as terminators for the sentences and the programs that read the data should only use the commas to determine the end of a data item.

3) The NMEA Sentence \$GPRMC

NMEA has its own version of essential GPS data. The most important NMEA sentences include the \$GPGLA which provides the current fix data, the \$GPRMC which provides the minimum GPS sentences information, and the \$GPGSA

which provides the Satellite status data. The recommended minimum sentence C, \$GPRMC, provides the essential GPS PVT data which is used to locate moving objects in terms of latitude and longitude. All GPS receivers output this sentence along with some other sentences. The \$GPRMC data format is given in Table 1.

This \$GPRMC NMEA sentence will look similar to:

```
$GPRMC,060447.644,A,1119.1450,N,07740.3855,E,22.4,216.39,030410,,,A*62
```

Where

```
$GPRMC      : Recommended Minimum sentence C
060447.644  : Fix taken at 06:04:47.644 UTC
A           : Status A=active or V=Void.
1119.1450,N : Latitude 11 deg 19.1450' N
07740.3855,E : Longitude 77 deg 40.3855' E
22.4        : Speed over the ground in knots
216.39      : Course over the ground
030410      : Date - 3rd of April 2010
A           : Autonomous mode
*62         : The checksum data, always begins with *
```

The moving object, if attached with a GPS receiver, can be located with the help of this NMEA sentence.

Table 1. Data format of \$GPRMC Sentence

Data Item	Format	Description
Message ID	\$GPRMC	RMC protocol header.
UTC Time (Coordinated Universal Time)	hhmmss.sss	Fix time to 1ms accuracy.
Status	Char	A Data Valid. V Data invalid.
Latitude	Float	Degrees * 100 + minutes.
N/S Indicator	Char	N=north or S=south.
Longitude	Float	Degrees * 100 + minutes.
E/W Indicator	Char	E=East or W=West.
Speed over Ground	Float	Speed Over Ground in knots
Course over Ground	Float	Course Over Ground in Degrees
Date	Ddmmyy	Current Date
Magnetic Variation	Blank	Not Used
E/W Indicator	Blank	Not Used
Mode	Char	A Autonomous
Checksum	*xx	2 Digits
Message Terminator	<CR><LF>	ASCII 13, ASCII 10

III. PREVIOUS WORK

The idea of many researchers for developing GPS simulators is to carry out their work with minimal cost and with accurate data which they need. Though some work has been done in this area, more GPS simulators which generate NMEA sentences are being developed to suit the custom needs of the hour. Some of the researchers [1,7] have

developed GPS simulators with hardware devices by generating pseudo GPS signals. These devices are dependent on hardware for the generation of GPS data. A device for generating NMEA sequences for testing embedded GPS reception firmware and hardware is described by Sinivee V [8]. This work describes a prototype GPS data simulator designed and built in Department of Physics of Tallinn University of Technology. Device can work in standalone mode and also in conjunction with control software. Configuration program can be used to generate test strings without tester hardware as well. First version of the device was limited to generating only one NMEA message and enabled simulation of communication errors. Later versions were developed to a more universal device with control via a GUI running on an ordinary PC. But there is no option provided in this software to create a database of NMEA sentences. Virace GPS Simulator[9] V0.01 is a keyboard or mouse controlled NMEA sentence generator and it can produce 3 COM port outputs. This simulator supports NMEA sentences like \$GPRMC, \$GPGLA, \$GPGSA and \$GPGSV. A lot of defined keys for steering and speed are available in this simulator. It supports three display and input formats of latitude and longitude. The disadvantage of using this is that it does not support GPS track replaying. The GPS Generator PRO [10] application is designed for providing assistance in developing, testing and debugging programs and equipment working with the NMEA-0183 protocol. This GPS device emulator generates NMEA messages from different data inputs. It can operate in 2 modes: 1) User select map, start point, heading, speed; 2) User select NMEA log file. The user can change speed and heading during simulation. The generated NMEA sentences can be used by other mapping software. But buying this software involves certain cost. Thus a few number of NMEA sentence generators are available but with some limitations. Mostly these generators involve some cost and hardware dependent. Some of the generators require keyboard or mouse control for path creation. In some generators, there is no provision for database creation. The NMEA sentence generator proposed in this work addresses these problems.

IV. MODELLING DYNAMIC ENVIRONMENT OF MOVING OBJECTS WITH NMEA SENTENCE GENERATOR

In order to simulate the 2D graphical tracking environment, the \$GPRMC-NMEA sentence which contains the random latitude and longitude positions of virtual moving objects are generated. The \$GPRMC sentence also contains other data items like UTC time, Status, speed over ground, course over ground, date, mode and checksum. The block diagram of the simplified NMEA sentence generator is given in Fig.2. The role of the NMEA sentence generator is to create the path of a moving object by generating sentences that contain random latitudes and longitudes. It also includes provision for random change of directions of moving objects. The source code for NMEA sentence generator is written in Matlab 7.6. The Flow Chart of NMEA sentence generator is given in Fig. 3 and the vital steps of the Flow

Chart are explained below. Date and Time: The system date and time is converted into UTC date and time. Status: The status of the GPS receiver is A', i.e., 'Active', when the receiver gets signals from satellites. It is assumed that the status of the receiver is A'. Latitude and Longitude: The starting latitude and longitude values of the virtual moving object are initialized. A graph is generated by plotting longitude on x axis and latitude on y axis. Two random functions are used to generate the random displacement of the virtual object in x axis and y axis respectively. Meanwhile another random function is used to generate a random number for direction. The direction of the object viz., North East, South East, South West and North West is generated based on the range of the generated random number. Based on the direction, the random displacement

values in x axis and y axis are added with or subtracted from previous latitude and longitude values and the new latitude and longitude positions of the object are calculated.

$$x(i)=x(i)+xdsign(i)*xrand;$$

$$y(i)=y(i)+ydsign(i)*yrand;$$

where

$x(i)$ x coordinate of current plot ; $i=1,2,3,\dots,n$.

$y(i)$ y coordinate of current plot ; $i=1,2,3,\dots,n$.

$xdsign$ controlled random direction sign (+ or -) generated for x axis.

$ydsign$ controlled random direction sign (+ or -) generated for y axis.

$xrand$ random displacement on x axis.

$yrand$ random displacement on y axis.

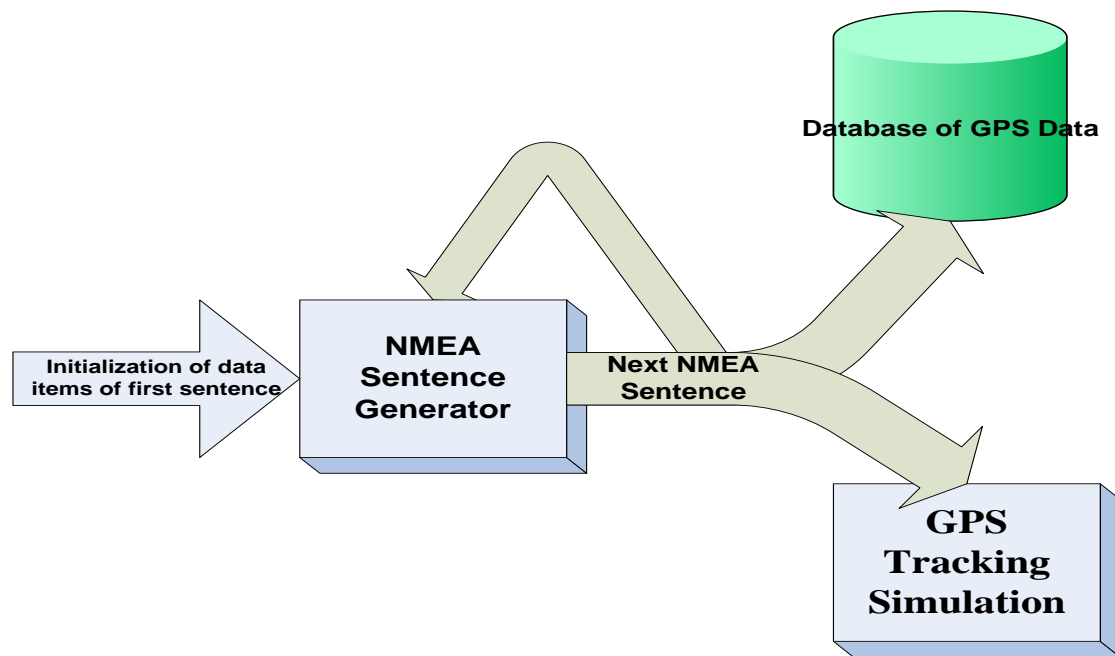


Figure 2. Block Diagram of NMEA Sentence Generator

The controlled random direction signs are generated in such a way that they change only after a random number of sentences and not for each and every sentence. Speed Over Ground: This data item is generated based on the

displacement of latitude and longitude position from the previous latitude and longitude position and it was converted into knots.

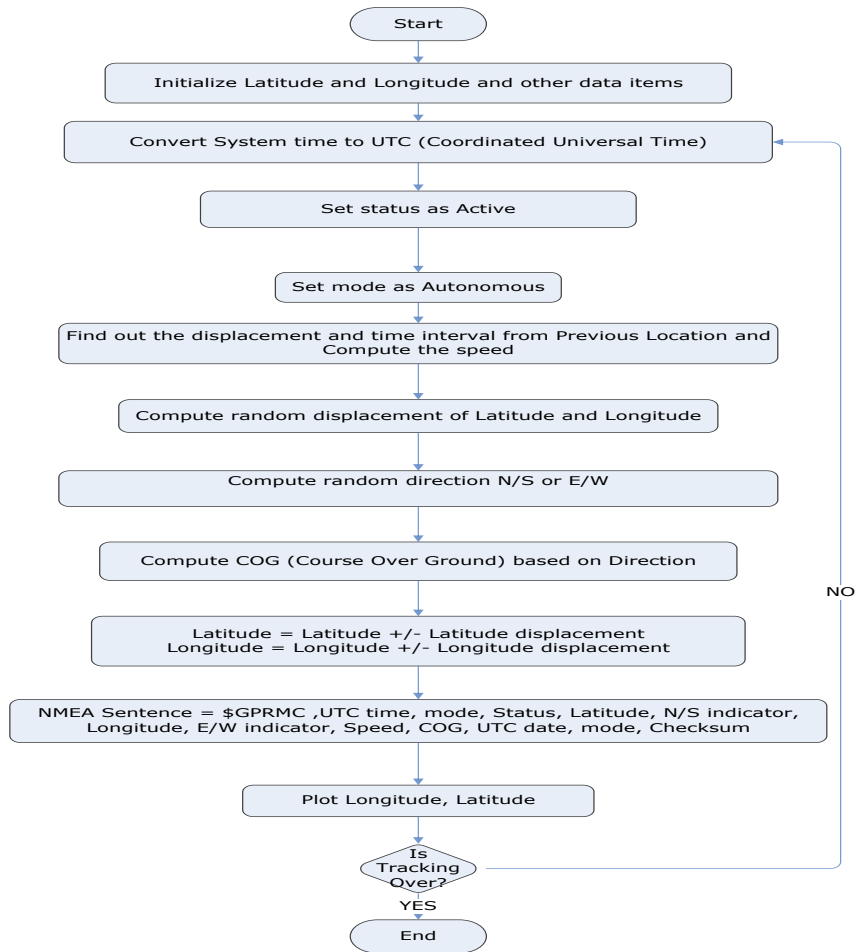


Figure 3. Flow Chart for NMEA Sentence Generator

Course Over Ground: This data item is generated based on the direction of the virtual moving object. The reading is always expressed in degrees to the true North. The range is between 0 degree and 359 degree. Mode: The mode is always fixed as Autonomous, i.e., 'A'. This field is of less importance in the simulation environment. Checksum: The checksum of the sentence is calculated for the self contained errors within the sentence. The object is now plotted in the current latitude and longitude position on the graph thus providing a simulated environment of a moving object. This process is iterated to get required number of subsequent latitudes, longitudes and other data items for the object and plotting the same on the graph.

V. GPS TRACKING SIMULATION RESULTS AND COMPARISON WITH STANDARD AND REAL TIME DATA

The generated \$GPRMC-NMEA sentences are compared with the data received from the actual GPS receiver. For the purpose of comparison, Wonde-X series GPS receiver (ZX4125) was used to generate real time GPS data. The GPS receiver is fixed in a moving car and the NMEA

sentence generated by it are stored in the database. For instance the first sentence received by the GPS receiver is, \$GPRMC,051142.950,A,1119.0211,N,07740.3135,E,0.00,135.00,110410,,A*69

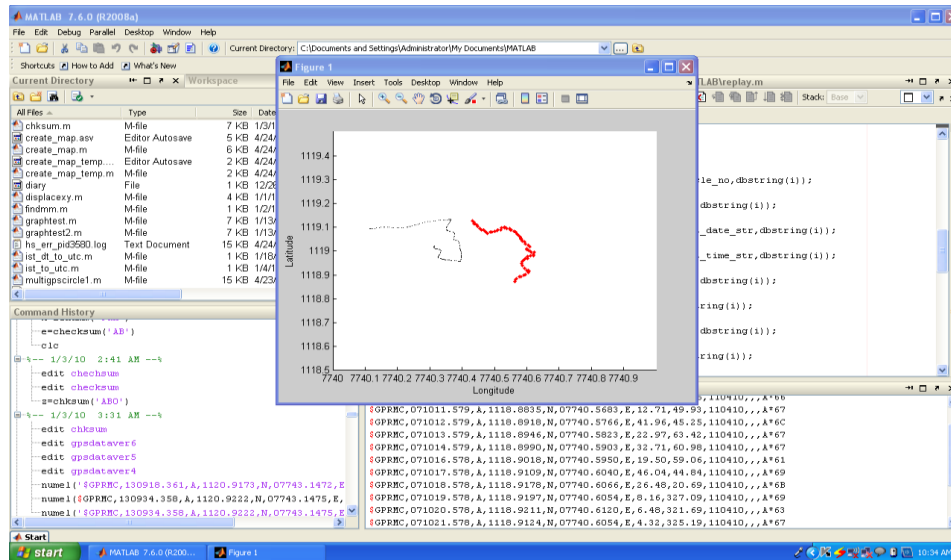
The following is the list of ten consecutive \$GPRMC sentences generated by ZX4125 when it is used in a moving car.

```

$GPRMC,051142.950,A,1119.0211,N,07740.3135,E,0.00,135.00,110410,,A*69
$GPRMC,051143.950,A,1119.0211,N,07740.3135,E,0.00,135.00,110410,,A*68
$GPRMC,051144.950,A,1119.0210,N,07740.3135,E,0.00,135.00,110410,,A*6E
$GPRMC,051146.949,A,1119.0201,N,07740.3134,E,2.48,182.66,110410,,A*67
$GPRMC,051147.949,A,1119.0189,N,07740.3133,E,4.61,183.94,110410,,A*63
$GPRMC,051149.949,A,1119.0155,N,07740.3131,E,6.00,183.10,110410,,A*67
$GPRMC,051150.949,A,1119.0141,N,07740.3134,E,5.08,169.10,110410,,A*60
$GPRMC,051152.948,A,1119.0120,N,07740.3155,E,5.79,125.53,110410,,A*6A
  
```

\$GPRMC,051153.948,A,1119.0113,N,07740.3173,E,7.13,1
10.62,110410,,,A*65

\$GPRMC,051154.948,A,1119.0105,N,07740.3197,E,8.93,1
07.04,110410,,,A*6E



..... Path Generated by GPS receiver

..... Path Generated by NMEA Sentence Generator

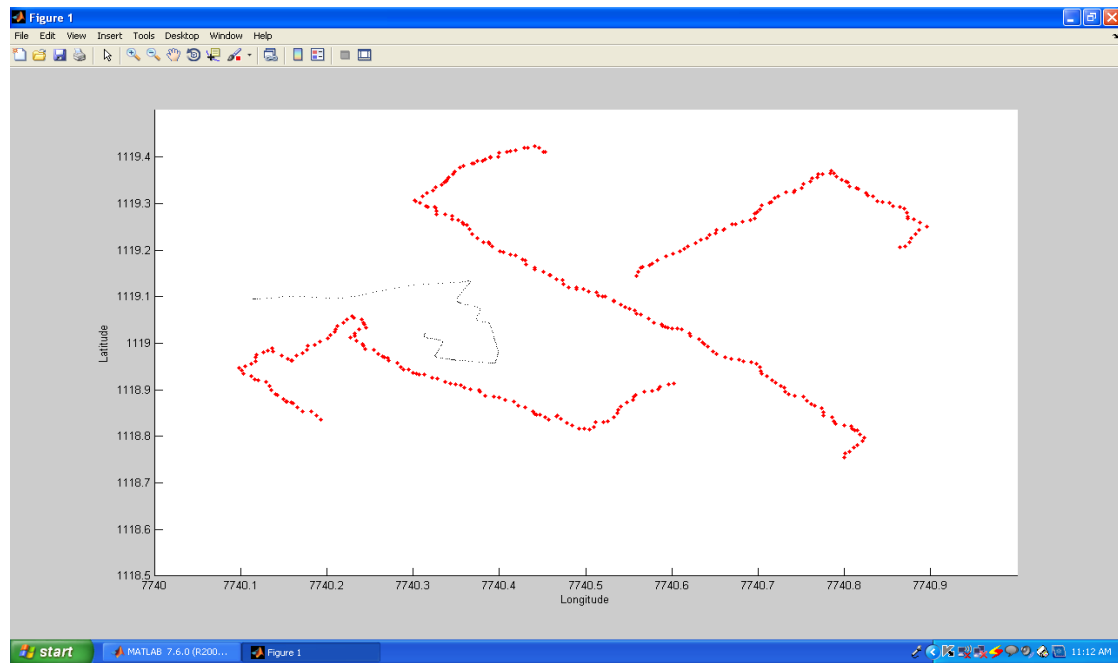
Figure 4. Latitudes and longitudes generated by Actual GPS receiver and NMEA Sentence Generator plotted in a graph

The \$GPRMC NMEA sentence generator discussed in this paper produces sentences that resemble those produced by ZX4125. The following is the list of ten consecutive sentences generated by the simplified NMEA sentence generator.

\$GPRMC,071009.579,A,1118.8728,N,07740.5641,E,43.23,
49.43,110410,,,A*65
\$GPRMC,071010.579,A,1118.8813,N,07740.5656,E,30.95,
10.15,110410,,,A*66
\$GPRMC,071011.579,A,1118.8835,N,07740.5683,E,12.71,
49.93,110410,,,A*67
\$GPRMC,071012.579,A,1118.8918,N,07740.5766,E,41.96,
45.25,110410,,,A*6C
\$GPRMC,071013.579,A,1118.8946,N,07740.5823,E,22.97,
63.42,110410,,,A*67
\$GPRMC,071014.579,A,1118.8990,N,07740.5903,E,32.71,
60.98,110410,,,A*67

\$GPRMC,071016.578,A,1118.9018,N,07740.5950,E,19.50,
59.06,110410,,,A*61
\$GPRMC,071017.578,A,1118.9109,N,07740.6040,E,46.04,
44.84,110410,,,A*69
\$GPRMC,071018.578,A,1118.9178,N,07740.6066,E,26.48,
20.69,110410,,,A*6B
\$GPRMC,071019.578,A,1118.9197,N,07740.6054,E,8.16,3
27.09,110410,,,A*69

The plotted graph for the latitude and longitude outputs of the GPS receiver ZX4125 and of the NMEA sentence generator is given in Fig. 4. The sentences generated by NMEA generator gives a simulated environment of moving object. These sentences are stored in the database to create a database of moving paths of virtual objects. From Fig. 5, it is observed that the NMEA sentence generator produces random path and random direction of the virtual object which is very similar to the moving path of a real object, like a car.



..... Path of a real object Generated by GPS receiver

..... Paths of 3 virtual objects generated by NMEA sentence generator

Figure 5. Paths of three virtual moving objects generated by NMEA Sentence Generator and a path generated by Actual GPS receiver plotted in a graph

The NMEA Sentence generator can be used to create paths of a number of virtual moving objects. The number of objects that can be created is limited to the main memory available in the system. Thus thousands of paths of independent moving objects can be generated by this NMEA generator without using GPS receivers. A comparison of the path generated by GPS receiver and paths of three virtual objects created by the simplified NMEA sentence generator is given in Fig. 5. Thus the created GPS database consists of paths of a number of virtual objects and this database can be used as a source of data by the researchers. The data items in the sentences produced by the actual GPS receiver and the NMEA sentence generator are compared against the standard values [11] in Table 2. It is observed from Table 2 that the values generated by NMEA sentence generator are similar to the values generated by ZX4125. Besides, these values are within the range of standard values. Hence the NMEA sentence generator can be used instead of an actual GPS receiver to generate tracking data. This saves cost and time of generation of GPS database to develop and test GPS applications.

Table 2. Comparison of the data items generated by NMEA Sentence Generator with Standard and Real Time Data

Field / Data item	Standard for data item		Generated by ZX4125 GPS receiver		Generated by NMEA Sentence Generator	
	Min chars	Max chars	Min chars	Max chars	Min Chars	Max chars
Message ID	6	6	6	6	6	6
UTC Time	1,2,2,1	2,2,2,3	2,2,2,3	2,2,2,3	2,2,2,3	2,2,2,3
Status	1	1	1	1	1	1
Latitude	1,2,1	3,2,4	1,2,4	3,2,4	1,2,4	3,2,4
N/S Indicator	1	1	1	1	1	1
Longitude	1,2,1	3,2,4	1,2,4	3,2,4	1,2,4	3,2,4
E/W Indicator	1	1	1	1	1	1
Speed over Ground	1.1	5.3	1.2	5.3	1.2	5.3
Course over Ground	1.1	3.2	1.1	3.2	1.1	3.2
Date	2,2,2	2,2,2	2,2,2	2,2,2	2,2,2	2,2,2
Magnetic Variation	(0)	(0)	(0)	(0)	(0)	(0)
E/W Indicator	(0)	(0)	(0)	(0)	(0)	(0)
Mode	1	1	1	1	1	1
Checksum	3	3	3	3	3	3
Message Terminator	2	2	2	2	2	2

VI. CONCLUSION AND FUTURE WORK

This paper introduces a simplified NMEA sentence generator which generates a sequence of NMEA sentences that are used to simulate a GPS tracking environment. The results of the NMEA Sentence generator are compared with the real time data and the available standard. It is found that the generated sentences are similar to the data generated in real time by a GPS receiver. Besides, the generated sentences are found to fit within standards. This NMEA sentence generator has eliminated the limitations of the previous work carried out in this area. The data generated by this software are used to create a database of sample GPS data which can be used by researchers to develop and test GPS applications. This generator can be used not only by software developers, but also by users, who want to learn navigation software before buying GPS receiver. This system generates only \$GPRMC sentences. In future this work can be extended to generate the entire set of NMEA sentences generated by a GPS receiver. This work can further be extended to generate GPS data of multiple GPS receivers. Besides, this system, at present can generate directions, but not based on the existing routes available on ground. The generation GPS data based on existing routes can also be an enhancement to this system.

VII. ACKNOWLEDGMENT

The authors would like to thank Hashprompt Logistics Management India Pvt. Ltd., for their support to perform initial demonstration on GPS tracking using their resources.

VIII. REFERENCES

- 1) Yun Young-sun, Park Sung-min and Kee Chang-don. (2003) "Test of GBAS Integrity Monitoring System Using GPS Simulator", Wuhan University Journal of Natural Sciences, vol. 8, no.2B, pp. 697-704.
- 2) Parkinson, B. W. and Spilker, J. J. (1996) Global Positioning System: Theory and Applications, American Institute of Aeronautics and Astronautics, Washington.
- 3) Interface control document. (2000) Navstar GPS Space Segment (Navigation User Interfaces).
- 4) Asoke K Talukder and Roopa R Yavagal. (2005) Mobile Computing-Technology, Applications and Service Creation, Tata McGraw Hill Publishing Company.
- 5) Goran M. Djuknic and Robert E. Richton. (2001) "Geolocation and Assisted GPS", Computer, vol. 34, no. 2, pp.123-125.
- 6) <http://www.gpsinformation.org/dale/nmea.htm> accessed on 15-02-2010.
- 7) Kou Yanhong, Yang Dongkai and Zhang Qishan. (2005) "GPS Satellite Simulator Signal Estimation based on ANN", Journal of Electronics(China), vol. 22 no.5, pp.458-464.
- 8) Sinivee V. (2010) "Simple yet efficient NMEA sentence generator for testing GPS reception

firmware and hardware", Novel Algorithms and Techniques in Telecommunications and Networking, Springer Netherlands, pp.207-210.

- 9) http://www.gpspassion.com/forumsen/topic.asp?TOPIC_ID=114933 accessed on 15-02-2010.
- 10) <http://avangardo.com> accessed on 15-02-2010.
- 11) Wonde-X series GPS module ZX4125 Specification Document. (2009) Wonde Proud Technology Co., Ltd., Taiwan.

Effectiveness of ICT in E-Governance with Special Reference to Jharkhand State

BhatnagarTanushree¹, Jha, A.N.², Singh, H.K.³

GJCST Classification
J.1

Abstract-This paper describes the effectiveness of the role of Information communication Technology in electronic governance in India and its genesis with special reference to Jharkhand State. It Highlights essentials of e-governance in current scenario, Influences of ICT in governance process & Implementation of e-governance in Jharkhand are also discussed. This article is based on the secondary sources like Journals, Newspaper Article, through Govt. Websites & personal communication. This paper throws light on the present scenario of e-governance in Jharkhand

Keywords-ICT, E-Governance, e-Governance in Jharkhand

I. ICT – INFORMATION COMMUNICATION TECHNOLOGY

ICT has become an indispensable part of our daily life. It has the potential to streamline procedures, inculcate discipline among workforce, and bring transparency and accountability and reduction of costs. The benefits are tremendous provided we have the will, insight and knowledge to use it. Information and Communication Technology (ICT) is a mixture of hardware (equipment), software (operating system, applications, etc.), and communications facilities (local area networks, wide area and backbone networks, communication protocols, etc.

II. E-GOVERNANCE INTRODUCTION

E-governance is used as a synonym to describe an it driven system of governance that works better, costs less and is capable of servicing the needs of the citizens and businesses as never before. Its goal is to create a more responsive, productive and effective administration. E-governance is also referred to as smart governance because it aims at using it to the processes of government functioning to bring about simple, moral, accountable, responsive and transparent governance. The citizens to government (c2g) and government to citizen (g2c) are the core pillars of not only e-governance but electronic commerce (e-commerce) as well. The most common interactions in e-governance are

- G2C – Government to Citizen
- G2B – Government to Business
- G2E - Government to Employee
- G2G – Government to Government

1) Importance of e-Governance

a) Introduction

India has been harnessing the benefits provided by the Information and Communication Technologies (ICT) to provide integrated governance, reach to the citizens faster, provide efficient services and citizen empowerment through access to information. To this effect, the Union Budget for the year 2008 had an increased allocation to the Department of Information Technology, amounting to INR 1,680 crore in 2008-09, up from INR1,500 crore in 2007-08. The National e-Governance Plan of Indian Government provides the impetus for long-term growth of e-Governance within the country. Several significant and successful e-Governance initiatives have already been taken at the centre and the state level in this direction. Some of these include: MCA 21 Project, Passport Project, Indian Railway Catering and Tourism Corporation computerization Project, at national level; and e-Seva and e-Cops, government of Andhra Pradesh project, Bhoomi, government of Karnataka project, Akhaya project of government of Kerala, Gyandoot project of government of Madhya Pradesh; among others at state level. There is an enthusiastic participation from the private sector in taking forward the e-Governance agenda of the government. The applications that have been implemented are targeted towards providing G2B, G2C and B2C services.

About¹ -Tanushree Bhatnagar, Junior Information Officer, Birla Institute of Technology, Deemed University, Mesra, Ranchi, Jharkhand, tanushreebhatnagar@gmail.com

About² -Dr. A.N. Jha, Professor & I/C BIT Extension Centre Lalpur, Ranchi, dranjha@rediffmail.com

About³ -Dr. H.K. Singh, Xavier Institute of Social Sciences(XISS) Ranchi, Jharkhand, hkxiss@rediffmail.com

Indian e-Governance Genesis

Year	Mile-stones for E-governance	Goal
1970	Origin in India	In house development
1984	New Computer Policy	Spread of computer Use
1986	Policy on Software Export, Development and Training	To promote Sectoral growth in IT's. Business process outsourcing
1987	Setting up of NICNET, DISNIC	Setting up of IT infrastructure in government Sector
1994	Policy on National Telecommunication (NTP 94)	To ensure better Tele-density focus on Rural Telephony.
1995	Launching of Internet Allocation and Release.	Web Access and bandwidth allocation for use
1997	Establishment of Telecom Regulatory Authority (TRAI)	To unbundled telecommunication services (last mile)
1998	National Task Force on IT	To formulate an IT policy document.
1999	Creation of Ministry of IT	To oversee implementation of IT Policy
1999	Policy on National Telecommunication (NTP 99)	To accelerate Tele-Density
2000	Formulation of IT Act	To provide legal status to use of IT in business, government and governance systems.
2000	Telecom Disputes settlement and Appellate Tribunal (TDSAT)	Fair and Transparent Telecom services
2000	Corporatization of Dot formation of BSNL	Unbundling of the telecommunication sector, private sector investments and managing USO
2004	Formulation of Broadband Policy	To implement broadband services in the last mile.
2006	National E-Governance Plan (NEGP)	To formulate, Plan, design and deploy e-government solutions and establish citizen interfaces.
2007	Mission 2007	To consider each village a knowledge centre.

Table: 1 Genesis of e-Governance

III. ORIGINS IN INDIA

E-governance originated in India during the seventies with a focus on in-house government applications in the areas of defense, economic monitoring, planning and the deployment of ICT to manage data intensive functions related to elections, census, tax administration etc. The efforts of the National Informatics Center (NIC) to connect all the district headquarters during the eighties was a watershed. From the early nineties, e-governance has seen the use of IT for wider sectoral applications with policy emphasis on reaching out to rural areas and taking in greater inputs from NGOs and private sector as well. Practically Mr. Rajiv Gandhi (based on the expensive IT proposal sent by bureaucracy from each ministry and state Govt during eighties) has created NIC to bring uniformity and transparency in the investment in IT.

Following are the ranking of Asian Countries according to UN survey on e-government

Table 2 UN survey on e-government of Asian countries

Country	2010	2008 Ranking	2005 Ranking
Maldives	92	95	77
Srilanka	111	101	94
Iran	102	108	98
India	119	113	87
Pakistan	146	131	136
Bhutan	152	134	130
Bangladesh	134	142	162
Nepal	153	150	126
Afganistan	168	167	168

IV. E-GOVERNANCE IN JHARKHAND

Jharkhand was formed as a separate state on November 15th, 2000 with Ranchi as its capital. Government of Jharkhand took conscious decision to make greater use of IT in governance. Thus in 2002 process of e-governance started.

Genesis of e-governance in Jharkhand at a glance

Year	Mile-stones for E-governance	Goal
2000	Origin in Jharkhand	Decision to make greater use of IT in governance
2000	Setting up of NIC	To provide Intranet and Internet connectivity to the State Administration through NICNET.
2002	New Computer Policy	Process of e-governance started
2002	Setting up of JharNet	State wide High Speed, Reliable & Scalable Communication links will be set up by the state government.
2003	Passport Control and Issuance System	Online passport System
2005	Treasury Computerization	Computerize the functioning of all the treasuries departments.
2007	CSC	plans to set-up 4562 Common Service Centres (CSC) through out all the panchayats in the State in first phase a
2007	E-Nagrik Seva	To provide good governance to the common citizens through the usage of ICT, for providing G2C services from the Kendra

Table 3 Genesis of e-Governance in Jharkhand

1) Objectives

IT policy of the State of Jharkhand will provide guidance to all agencies involved in the goal of overall development of the state using IT as the enabler. Agencies in the state who are likely to use the IT policy are Citizens, various ministries, legislative, IT entrepreneurs, business & professional associations and other stakeholders.

2) Initiatives for "Improved Communications & Infrastructure"

- JharNet State wide High Speed, Reliable & Scalable Communication links will be set up by the state government.
- Internet connectivity across the state
- Seamless connectivity among various govt. organizations
- Internet connectivity across the state

During the past few year's with the active support of various agencies, following are the some of the achievement of Govt. of Jharkhand towards e-governance.

- Treasury Computerization
- Transport Computerization
- Land Records Information System
- Election Support System
- Prison Management System
- Video Conferencing based Public
- Justice in jails of Jharkhand

- City Civil Court Computerization
- High Court Computerization
- Computerization Support to the Raj
- Bhawan and CM Secretariat
- Training Support
- Public Grievances Monitoring System
- Human Resource Management System (e-Personnel)
- DM's Court Monitoring System
- Online Project Monitoring
- Online Passport
- Other District Activities E-Nagrik Seva
- Pragya Kendra

According to Dataquest IDC Survey in 2008 Jharkhand comes in a last position. Jharkhand is a worst State in Satisfaction level of Business with e-governance services and also the satisfaction level of citizens with e-governance services in the Agriculture, Healthcare, Transport Services, Employment Exchange & Land Property Department.

V. CONCLUSION

Information and Communication technology is no longer an Alice in Wonderland" in any part of the world. Some ray of hope can be found in successful projects like MCA21 and modernisation of intellectual property rights offices in India. Other projects are also in pipeline but in the absence of transparency and accountability they are taking twice the time originally allotted to them. We hope the Government of

India would do the needful as soon as possible. We have to empower the citizens with the power of ICT.

VI. REFERENCE AND BIBLIOGRAPHY

- 1) Bhatnagar Subhash, and Robert Schware, eds. Information and Communication Technology in Rural Development: Case Studies from India." [2000] Working paper. World Bank Institute, Washington, D.C.
- 2) Bhatnagar Tanushree, Singh H.K. Role of ICT with special reference of Jharkhand State" National conference on recent advance in Information technology. ISM Dhanbad 2009.
- 3) Baweja Paul The International Information & Library Review, [2007] 4 September-December Volume 39, Issues 3-, Pages 176-184
- 4) <http://egovindia.wordpress.com/2006/07/06/what-is-happening-at-nic-some-background-info>
- 5) <http://www.i4donline.net> Volume VI no-2 Feb 2008
- 6) <http://www.dqindia.com/>, Yogesh Does E-governance mean Good Governance? [2005] YOJANA
- 7) Bedi, Kiran New governance opportunities in India" [2001] Sage Publication page 374
- 8) Gupta, B. et al., Adoption of ICT in a government organization ..., J. Strategic Information System. [2008], doi:10.1016/j.jsis.2007.12.004 <http://sciencedirect.com>
- 9) Jhunjhunwala, Ashok. Unleashing Telecom and Internet in India." [2000] Paper presented at the India Telecom Conference at Stanford University's Asia/Pacific Research Center, November, Stanford, Calif.
- 10) Karwal, Anita E-governance and citizen's charter" [2005] Journal of Rural Development, Vol. 24(3) pp 409-427
- 11) <http://www.indiatogether.org/2009/sep/gov-csegov.htm>
- 1) J. Bruno and P. Downey. Complexity of task sequencing with deadlines, set-up times and changeover costs. SIAM Journal on Computing, 7(4):393-404, 1978.
- 2) G.N. Frederickson and B. Wittman. Approximation algorithms for the traveling repairman and speeding deliveryman problems with unit-time windows. Proceedings of APPROX-RANDOM,, LNCS 4627:119-133, 2007.
- 3) M. Desrochers, J. Desrosiers, and M. Solomon. A new optimization algorithm for the vehicle routing problem with time windows. Operations Research, 40:342-354, 1992.
- 4) M. Desrochers, J. Lenstra, M. Savelsbergh, and F. Soumis. Vehicle routing with time windows: optimization and approximation. In B.L. Golden, A.A. Assad (eds.). Vehicle Routing: Methods and Studies, North-Holland, Amsterdam, pages 65-84, 1988.
- 5) B. Golden, L. Levy, and R. Vohra. The orienteering problem. Naval Research Logistics, 34:307-318, 1987.
- 6) M. Gravel, W. Price, and C. Gagn. Scheduling jobs in a aluminum factory using a genetic algorithm. International Journal of Production Research, 38(13):3031-3041, 2000.
- 7) C. Jordan. A two-phase genetic algorithm to solve variants of the batch sequencing problem. International Journal of Production Research (UK), 36(3):745-760, 1998.
- 8) M. Kantor and M. Rosenwein. The orienteering problem with time windows. Journal of the Operational Research Society, 43:629-635, 1992.
- 9) Y. Karuno and H. Nagamochi. A 2-approximation algorithm for the multi-vehicle scheduling problem on a path with release and handling times. In Proceedings of the European Symposium on Algorithms, pages 218-229, 2001.
- 10) A. Kolen, A. R. Kan, and H. Trienekens. Vehicle routing with time windows. Operations Research, 35:266-273, 1987.
- 11) S. Thangiah. Vehicle Routing with Time Windows using Genetic Algorithms, Application Handbook of Genetic Algorithms: New Frontiers, Volume II. Lance Chambers (Ed.). CRC Press, 1995.
- 12) K. Tan, L. Lee, and K. Zhu. Heuristic methods for vehicle routing problem with time windows. In Proceedings of the 6th AI and Math, 2000.
- 13) M. Savelsbergh. Local search for routing problems with time windows. Annals of Operations Research, 4:285-305, 1985.
- 14) S. R. Thangiah, I. H. Osman, R. Vinayagamorthy and T. Sun. Algorithms for the vehicle routing problems with time deadlines. American Journal of Mathematical and Management Sciences, 13(3&4):323-355, 1993.
- 15) J. Tsitsiklis. Special cases of traveling salesman and repairman problems with time windows. Networks, 22:263-282, 1992.
- 16) J. Wisner and S. Siferd. A survey of u.s. manufacturing practices in make-to-order machine shops. Production and Inventory Management Journal, 1:1-7, 1995.
- 17) W. Yang and C. Liao. Survey of scheduling research involving setup times. International Journal of Systems Science, 30(2):143-155, 1999.
- 18) N. Bansal, A. Blum, S. Chawla, and A. Meyerson. Approximation algorithms for deadline-TSP and vehicle routing with time-windows. In Proceedings of the 36th Annual ACM Symposium on Theory of Computing, pages 166-174. ACM New York, NY, USA, 2004.
- 19) R.K. Ahuja, T.L. Magnanti, and J.B. Orlin. Network Flows: Theory, Algorithms, and

- Applications. Prentice Hall, Upper Saddle River, New Jersey, 1993.
- 20) E.M. Arkin, J.S.B. Mitchell, and G. Narasimhan. Resource-constrained geometric network optimization. In Symposium on Computational Geometry, pages 307–316, 1998.
 - 21) S. Arora and G. Karakostas. A $2 + \epsilon$ approximation algorithm for the k-MST problem. Mathematical Programming A, 107(3):491–504, 2006. Preliminary version in Proc. of ACM-SIAM SODA, 754–759, 2000.
 - 22) B. Awerbuch, Y. Azar, A. Blum, and S. Vempala. Improved approximation guarantees for minimum weight k-trees and prize-collecting salesmen. SIAM J. Computing, 28(1):254–262, 1998. Preliminary Version in Proc. of ACM STOC, 277–283, 1995.
 - 23) E. Balas. The prize collecting traveling salesman problem. Networks, 19(6):621–636, 1989.
 - 24) A.M. Frieze, G. Galbiati, and F. Maffioli. On the worst-case performance of some algorithms for the asymmetric traveling salesman problem. Networks, 12(1):23–39, 1982.
 - 25) R. Bar-Yehuda, G. Even, and S. Shahar. On approximating a geometric prize-collecting traveling salesman problem with time windows. Journal of Algorithms, 55(1):76–92, 2005. Preliminary version in Proc. Of ESA, 55–66, 2003.
 - 26) N. Garg. A 3-approximation for the minimum tree spanning k vertices. In Proceedings of the 37th Annual Symposium on Foundations of Computer Science, pages 302–309. IEEE Computer Society, 1996.
 - 27) A. Blum, R. Ravi, and S. Vempala. A Constant-Factor Approximation Algorithm for the k-MST Problem. Journal of Computer and System Sciences, 58(1):101–108, 1999. Preliminary version in Proc. of ACMSTOC, 1996.
 - 28) K. Chaudhuri, B. Godfrey, S. Rao, and K. Talwar. Paths, trees, and minimum latency tours. In 44th Annual Symposium on Foundations of Computer Science, pages 36–45. IEEE Computer Society, 2003.
 - 29) C. Chekuri and A. Kumar. Maximum coverage problem with group budget constraints and applications. Proceedings of APPROX-RANDOM, pages 72–83, 2004.
 - 30) C. Chekuri and M. P'al. A recursive greedy algorithm for walks in directed graphs. In Proceedings of the 46th Annual Symposium on Foundations of Computer Science, pages 245–253. IEEE Computer Society, 2005.
 - 31) C. Chekuri and M. P'al. An $O(\log n)$ Approximation for the Asymmetric Traveling Salesman Path Problem Theory of Computing, 3:197–209, 2007. Preliminary version in Proc. of APPROX, 95–103, 2005.
 - 32) N. Garg. Saving an epsilon: a 2-approximation for the k-MST problem in graphs. In Proceedings of the 37th Annual ACM Symposium on Theory of computing, pages 396–402. ACM, 2005.
 - 33) M. X. Goemans and D. P. Williamson. A general approximation technique for constrained forest problems. SIAM Journal on Computing, 24:296–317, 1995.
 - 34) B.L. Golden, L. Levy, and R. Vohra. The orienteering problem. Naval Research Logistics, 34(3):307–318, 1987.
 - 35) G. Gutin and A.P. Punnen, editors. The traveling salesman problem and its variations. Springer, Berlin, 2002.
 - 36) C. Chekuri, N. Korula, and M. P'al. Improved Algorithms for Orienteering and Related Problems. In Proc. of ACM-SIAM SODA, January 2008.
 - 37) E.L. Lawler, A.H.G. Rinnooy Kan, J.K. Lenstra, and D.B. Shmoys, editors. The Traveling salesman problem: a guided tour of combinatorial optimization. John Wiley & Sons Inc, 1985.
 - 38) V. Nagarajan and R. Ravi. Poly-logarithmic approximation algorithms for directed vehicle routing problems. Proceedings of APPROX-RANDOM, LNCS 4627:257–270, 2007.
 - 39) P. Toth and D. Vigo, editors. The vehicle routing problem. SIAM Monographs on Discrete Mathematics and Applications. Society for Industrial Mathematics, Philadelphia PA, 2001.
 - 40) J.N. Tsitsiklis. Special cases of traveling salesman and repairman problems with time windows. Networks, 22(3):263–282, 1992.
 - 41) K. Chen and S. Har-Peled. The orienteering problem in the plane revisited. SIAM J. on Computing, 38(1):385–397, 2008. Preliminary version in Proc. of ACM SoCG, 247–254, 2006.

Various Security Attacks and Trust Based Security Architecture for MANET

Ankit Jain¹, Arnika Jain², Pramod Kumar Sagar³

GJCST Classification
C.2.0

Abstract-A Mobile Ad hoc Network is a group of wireless mobile computers in which nodes cooperate by forwarding packets to each other allowing them to communicate beyond direct wireless transmission range. Mobile Ad hoc Networks (MANET) has become an exciting and important technology in recent years because of the rapid proliferation of wireless devices. Security is an important issue for all kinds of networks including the Wireless Ad Hoc Networks. In this paper, we are presenting some of the reasons that have made MANETs more vulnerable to attacks than the traditional wired network. This paper also covers the security attributes and the various challenges to security design. This paper sheds light on some of the security attacks that exists in MANETs. This Paper also proposes Trust Based Security Architecture for MANET.

Index Term-s Blackhole attacks, DoS and Flooding, Eavesdropping, Grayhole attacks, Impersonation, MANETs (Mobile Ad hoc Networks), Rushing, Wormhole attacks

I. INTRODUCTION

MANET is a collection of independent mobile nodes that can communicate to each other via radio waves. The mobile nodes that are in radio range of each other can directly communicate whereas others need the aid of intermediate nodes to route their packets. These networks are fully distributed and can work at any place without the help of any infrastructure. The system may operate in isolation, or may have gateways to interface with a fixed network. This property makes MANET highly robust. The characteristics of these networks are summarized below:

- Autonomous and Infrastructure-less
- Dynamic Network Topology
- Self-organization and Self-administration
- No Centralized controller and Infrastructure: Intrinsic Mutual Trust
- Device Heterogeneity
- Bandwidth-Constrained
- Energy-Constrained Operation
- Multi-hop Routing
- Network Scalability
- Nodes can be both host or router
- Frequent Routing updates
- Limited Physical Security

About¹- Technical Leader Software Development Department IBM India; SRM University ankitjain@in.ibm.com

About²- Assistant Professor CSE Department NCR Campus, Modinagar, jain.arnika2009@gmail.com

About³- Assistant Professor IT Department Ghaziabad (UP), India pksagar75@rediffmail.com

II. SECURITY PROBLEMS IN MANET

Security is one of the important issues for wired networks as well as for Wireless Ad Hoc Networks. MANETs are much more vulnerable to attacks than the wired networks due to system constraints in mobile devices which include low-power, open medium, dynamically changing network topology, cooperative algorithms, lack of centralized monitoring and low battery power. The unique characteristics of MANETs present a new set of nontrivial challenges to security design which includes open network architecture, shared wireless medium, stringent resource constraints, highly dynamic network topology and some more. Some of the important ones are discussed below:

- 1) **Open Medium:** Eavesdropping is much easier in wireless networks than in wired networks.
- 2) **Dynamically changing Network Topology:** Mobile nodes come and go from network thereby allowing any malicious node to join the network without being detected.
- 3) **Cooperative Algorithms:** The routing algorithm of MANET requires mutual trust between nodes which violates the principle of network security.
- 4) **Lack of Centralized Monitoring:** Absence of any Centralized Infrastructure prohibits any monitoring agents in the system.
- 5) **Lack of Clear Line of Defense:** The line of Defense i.e. prevention, Detection and Response are needed. Security research in wired networks needs to deploy layered security mechanism because security is a process that is as secure as its weakest link.

III. SECURITY ATTRIBUTES

Security has become a primary concern in order to provide protected communication between mobile nodes in a hostile environment and the ultimate goal of the security solutions for MANETs is to provide security services such as Confidentiality, Integrity, Availability, Non-Repudiation and Authentication, Authorization and Anonymity.

- 1) **Confidentiality** ensures that Secret information or data is never disclosed to unauthorized devices.
- 2) **Integrity** tells that a received message is not corrupted.
- 3) **Availability** permits the survivability of network services despite Denial-of-Service attacks.

- 4) **Non-repudiation** ensures that the sender of a message cannot deny having sent the message.
- 5) **Authentication** enables a node to ensure the identity of the Peer node it is communicating with.
- 6) **Authorization** is a process in which an entity is issued a credential, which specifies the privileges and permissions it has and cannot be falsified, by the certificate authority.

Anonymity ensures that the information used to identify the owner or the current user of the node should default be kept private and not be distributed by the node itself or the System Software.

IV. Security Challenges

- 1) **Channel Vulnerability:** Broadcast Wireless channels allow message Eavesdropping and Injection easily.
- 2) **Node Vulnerability:** Nodes do not reside in physically protected places, thus easily fall under attack.
- 3) **Absence of Infrastructure:** Certification/Authentication Authorities are absent.
- 4) **Dynamically Changing Network Topology** puts security of routing protocols under threat.
- 5) **Power and Computational Limitations** prevent the use of complex Encryption Algorithms.

V. SECURITY ATTACKS

MANETs vulnerabilities and lacks give rise to attacks at network layer of ISO/OSI stack. There are basically two types of attacks: Active and Passive. In an active attack, information is inserted to the network and thus the network operation or some nodes may be harmed. Examples are Impersonation, Modification, Fabrication and Disclosure attack. Nodes that perform active attacks with the aim of damaging other nodes by causing network outage are considered to be malicious. In a passive attack, a malicious node either ignores operations supposed to be accomplished by it (examples: silent discard, partial routing information hiding), or listens to the channel, attempting to retrieve valuable information (example: eavesdropping). Nodes that perform passive attacks with the aim of saving battery life for their own communications are considered to be selfish. In this paper, we are discussing some of the active attacks given below:

1) Impersonation

An Act whereby one Entity assumes the identity and privileges of another Entity without restrictions and without any indication visible to the recipients of the impersonator's calls that delicately has taken place as shown in Fig 5.1.

2) Blackhole attacks:

A Blackhole is a malicious node that falsely replies for route requests without having an active route to the destination and exploits the Routing Protocol to advertise itself as

having a good and valid path to a destination node. As shown in Fig 5.2, a malicious node tries to become an element of an active route, if there is a chance and it has bad intention of disrupting data packets being sent to the destination node or obstructing the route discovery process.

3) Grayhole attacks:

A Grayhole may forward all packets to certain nodes but may drop packets coming from or destined to specific nodes as shown in Fig 5.3. In this type of attack, node may behave maliciously for some time but later on it behaves absolutely normally. This type of attacks is more difficult compared to black hole attack.

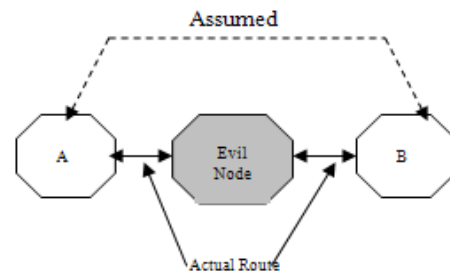


Fig 5.1 Impersonation

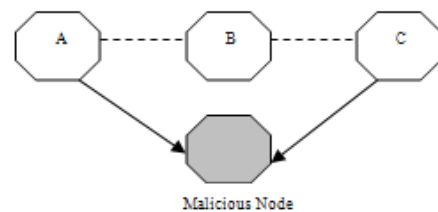
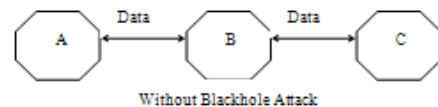


Fig 5.2 Blackhole Attack

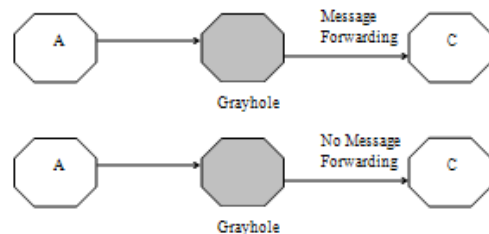


Fig 5.3 Grayhole Attack

4) Wormhole attacks

In a wormhole attack, a malicious node can record packets (or bits) at one location in the network and tunnel them to another location through a private network shared with a colluding malicious node. Wormhole attack can be done with one node also, but generally two or more attackers connect via a link called wormhole link. Wormhole attack is of three types: Closed Wormhole, Half Open Wormhole,

and Open Wormhole. All of these have been shown in Fig 5.4.

5) Rushing

When a neighbor of the target receives the rushed REQUEST from the attacker, it forwards that REQUEST, and will not forward any further REQUESTs from this Route Discovery. When non-attacking REQUESTs arrive later at these nodes, they will discard these legitimate REQUESTs as shown in Fig 5.5.

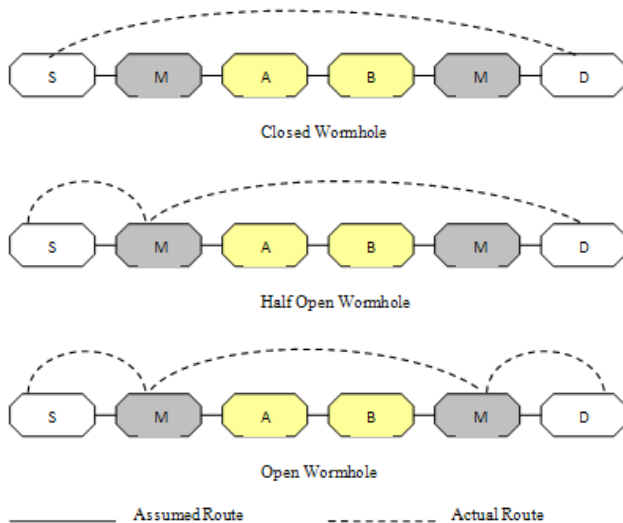


Fig 5.4 Wormhole

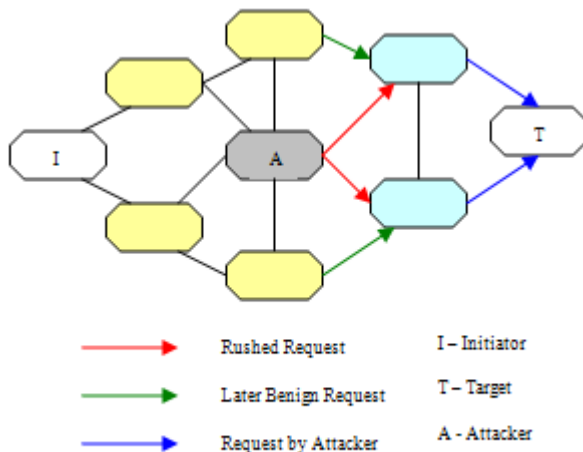


Fig 5.5 Rushing

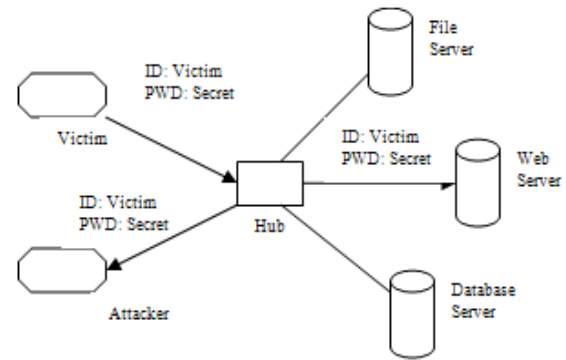


Fig 5.6 Eavesdropping

6) Eavesdropping

Eavesdropping is another kind of attack that usually happens in the Mobile Ad hoc Networks and is shown in Fig 5.6. The goal of eavesdropping is to obtain some confidential information that should be kept secret during the communication. This confidential information may include the location, public key, private key or even passwords of the nodes. Because such data are very important to the security state of the nodes, it should be kept away from unauthorized access.

7) DoS and Flooding

In a Denial-of-Service (DoS) attack, an attacker attempts to prevent legitimate users from accessing information or services. By targeting client computer and its network connection, or the computers and network of the sites client is trying to use, an attacker may be able to prevent the client from accessing email, websites, online accounts, or other services that rely on the affected computer. The most common and obvious type of DoS attack occurs when an attacker floods a network with information as shown in Fig 5.7. When client types a URL for a particular website into browser, the server can only process a certain number of requests at once, so if an attacker overloads the server with requests, it can't process the legitimate request.

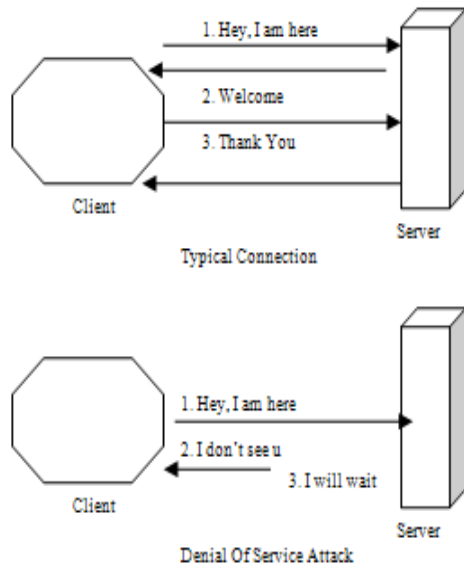


Fig 5.7 Denial Of Service Attack

VI. TRUST BASED SECURITY ARCHITECTURE FOR MANET

Since in MANETs, there is no centralized authority for trust relationship between the nodes to communicate, this Paper proposes a Security Architecture for the secure transmission of data between MANET nodes named Trust Based Security Architecture for MANET. The objective of the architecture presented in this paper is to provide security in a Collaborative, Self-organized manner. It provides tolerance to compromised nodes, have the ability to detect and remove adversaries from network and handle routing misbehaviors. The model defined in our architecture differs from related models [1] and [2, 3] by defending against both Flooding and Packet drop attacks as is shown in fig 6.1. The Trust Infrastructure module corresponds to the trust relationship between nodes to communicate and it has various security mechanisms that are constructed in a distributed manner and are the basic building blocks of the entire security system. Communications Security module refers to the secure transmission of data from one node to another. Since Routing is an important part for a network so every node should participate in routing activity and thus leads to network topology so that the data packets can be routed to the correct destination. Routing Security refers to security corresponding to Routing Protocols. The Routing Security is categorized into two parts: Secure Routing and Secure data forwarding. In Secure Routing, a node cooperate to have correct routing information and thus keeps the network connected efficiently and in Secure Data Forwarding, data packets, while on the route, should be protected from tampering, dropping, and altering by any unauthorized party. The secure routing protocol is responsible to discover secure paths and to have secure data transmission through the secret associations established and maintained by the key management mechanism.

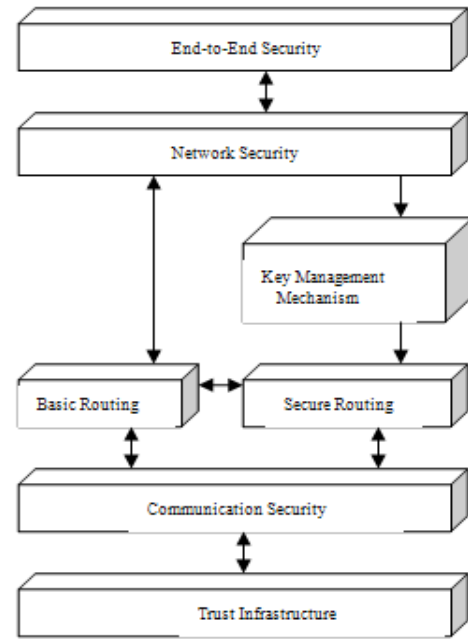


Fig 6.1 Trust Based Security Architecture

In this architecture, the Secure Routing Protocol forwards evidence for fabrication and modification attacks to the Trust Infrastructure module, and takes advantage of the Trust Infrastructure module feedback to make better routing decisions. Although the Key Management Mechanism [8] does not forward any evidence to Trust Infrastructure module, it relies heavily on its feedback to dynamically manage the keys. Network Security refers to the security mechanisms used by the network protocols which perform sub-network access operations from end system to end system. Secure Message Transmission [3] working at network security is a good example of security efforts done in the end systems as a remedy for the unreliable routing protocol. End-to-End Security refers to end system security. The security protocols at this level is independent of the underlying networking technology since the related security mechanisms are restricted to only intended parties. In sequence, this model is used to defend against both flooding and packet drop attacks.

VII. CONCLUSION

MANETs require a reliable, efficient, scalable and most importantly, a secure protocol as they are highly insecure, self-organizing, rapidly deployed and they use dynamic routing. In this paper, we discussed the vulnerable nature of the mobile ad hoc network, and owe to that; there are numerous security threats that disturb the development of it. This paper also covers the security attributes and the various challenges to security design. This paper also presents the security issues. Then it presents the main attack types that exist in it. This Paper has focused on designing security architecture in tackling security challenges that mobile ad hoc networks are facing. In this paper, we have proposed Trust Based Security Architecture for MANETs. We expect

this security architecture can be used as a framework when designing system security for ad hoc networks.

VIII. REFERENCES

- 1) Balakrishnan, V. Varadharajan, U. K. Tupakula, and P.Lucs, "*Trust Integrated Cooperation Architecture for Mobile Ad-hoc Networks*". Proceedings of 4th IEEE International Symposium on Wireless Communication Systems (ISWCS 2007), Trondheim, Norway, 2007.
- 2) L. Buttyan and J. Hubaux, "*Nuglets: A Virtual Currency to Stimulate Cooperation in Self-organized Ad hoc Networks*". Technical Report (DSC/2001/001), Swiss Federal Institute of Technology, 2001.
- 3) P. Papadimitratos, Z. Haas, *Secure Data Transmission in Mobile Ad Hoc Networks*, ACM Workshop on Wireless Security, 2003.
- 4) M.S. Corson, J.P. Maker, and J.H. Cernicione, *Internet-based Mobile Ad Hoc Networking*, IEEE Internet Computing, pages 63–70, July-August 1999.
- 5) Lidong Zhou and Zygmunt J. Hass, *Securing Ad Hoc Networks*, IEEE Networks Special Issue on Network Security, November/December 1999.
- 6) Balakrishnan, and V. Varadharajan, "*Fellowship in Mobile Ad hoc Networks*". Proceedings of 1st IEEE International Conference on Security and Privacy for Emerging Areas in communications Networks (SecureComm 2005), Athens, Greece, pp. 225-227, 2005
- 7) F. Stajano and R. Anderson: *The Resurrecting Duckling: Security Issues for Ad-Hoc Wireless networks*. In Proceedings of the 7th International Workshop on Security Protocols, 1999.
- 8) S. Capkun, L. Buttyan, J. Hubaux: *Self-Organized Public-Key Management for Mobile Ad Hoc Networks*, IEEE Transactions on Mobile Computing, VOL.1, NO.1, 2002.
- 9) W. Lou, W. Liu, Y. Fang: *SPREAD: Enhancing Data Confidentiality in Mobile Ad Hoc Networks*, IEEE INFOCOM, 2004.
- 10) H. Yang, X. Meng and S. Lu: *Self-Organized Network-Layer Security in Mobile Ad Hoc Networks*, ACM, 2002.

A Modification on K-Nearest Neighbor Classifier

Hamid Parvin, Hoseinali Alizadeh, Behrouz Minati

GJCST Classification
1.2.6

Abstract-K-Nearest Neighbor (KNN) classification is one of the most fundamental and simple classification methods. When there is little or no prior knowledge about the distribution of the data, the KNN method should be one of the first choices for classification. In this paper a modification is taken to improve the performance of KNN. The main idea is to use robust neighbors in training data. This modified KNN is better from traditional KNN in both terms: robustness and performance. The proposed KNN classification is called Modified K-Nearest Neighbor (MKNN). Inspired from the traditional KNN algorithm, the main idea is to classify an input query according to the most frequent tag in set of neighbor tags. MKNN can be considered a kind of weighted KNN so that the query label is approximated by weighting the neighbors of the query. The procedure computes the frequencies of the same labeled neighbors to the total number of neighbors. The proposed method is evaluated on a variety of several standard UCI data sets. Experiments show the excellent improvement in the performance of KNN method.

Keywords- MKNN, KNN Classification, Modified K-Nearest Neighbor, Weighted K-Nearest Neighbor, Neighbor Validation.

I. INTRODUCTION

Nowadays, recognition system is used in many applications which are related to different fields that have different nature. Pattern recognition is about assigning labels to objects which are described by a set of values named attributes or features. Current research builds upon foundations laid out in the 1960s and 1970s [1].

There are three major types of pattern recognition trends: unsupervised, semi-supervised and supervised learning. In the supervised category, also called classification or regression, each object of the data comes with a pre-assigned class label. In other hand, there is a teacher saying the true answer. The task is to train a classifier to perform the labeling, using the teacher. A procedure which tries to leverage the teacher's answer to generalize the problem and obtain his knowledge is learning algorithm. Most often this procedure cannot be described in a human understandable form, like Artificial Neural Networks classifiers. In these cases, the data and the teacher's labelings are supplied to the machine to run the procedure of learning over the data. Although the classification knowledge learned by the machine in this process might be obscure, the recognition accuracy of the classifier will be the judge of its quality of learning or its performance [1].

In some new classification systems, it is tried to investigate

the errors and propose a solution to compensate them [2-5]. There are many classification and clustering methods as well as the combinational approaches [6-8]. While the supervised learning tries to learn from the true labels or answers of the teacher, in semi-supervised the learner conversely uses teacher just to approve or not to approve the data in total. It means that in semi-supervised learning there is not really available teacher or supervisor. The procedure first starts with fully random manner, and when it reaches the state of final, it looks to the condition whether he won or lost. For example in the chess game, take it in consideration, that there may be none supervisor, but you gradually train to play better by trial-and-error process and looking at the end of the game to find you won or lost. K-Nearest Neighbor (KNN) classification is one of the most fundamental and simple classification methods. When there is little or no prior knowledge about the distribution of the data, the KNN method should be one of the first choices for classification. It is a powerful non-parametric classification system which bypasses the problem of probability densities completely [9]. The main idea is to classify an input query x into the most frequent tag in the set of its neighbor tags. The KNN rule classifies x by assigning it the label most frequently represented among the K nearest samples; this means that, a decision is made by examining the labels on the K -nearest neighbors and taking a vote. It is first introduced by Fix and Hodges in 1957 [10]. Later in 1967, KNN is looked at in theoretic perspective [11]. Once such consideration of KNN classification were established, a long line of investigation ensued including new rejection approaches [12], refinements with respect to Bayes error rate [13], distance weighted approaches [14, 15], soft computing [16] methods and fuzzy methods [17, 18].

ITQON et al. in [19] proposed a classifier, TFkNN, aiming at upgrading of distinction performance of KNN classifier and combining plural KNNs using testing characteristics. Their method not only upgrades distinction performance of the KNN but also brings an effect stabilizing variation of recognition ratio; and on recognition time, even when plural KNNs are performed in parallel, by devising its distance calculation it can be done not so as to extremely increase on comparison with that in single KNN. Some KNN advantages can be as follows: simplicity, robustness to noisy training data, and effectiveness in the adequate training data. It has some disadvantages such as: high computation cost in a test query, the large memory to implement, low accuracy rate in multidimensional data sets, parameter K , unclearness of distance type. Shall we use all attributes or certain attributes only [20]?

In this paper a new interesting algorithm is proposed which partially overcomes the low accuracy rate of KNN. Beforehand, it preprocesses the train set, computing the

validity of any train samples. Then the final classification is executed using weighted KNN which is employed the validity as the multiplication factor.

II. MODIFIED K-NEAREST NEIGHBOR

The main idea of the presented method is assigning the class label of the queried instance into K validated data training points. In other hand, first, the validity of all data samples in the train set is computed. Then, a weighted KNN is performed on any test samples. Fig 1 shows the pseudo code of the MKNN algorithm.

```

Output_label := MKNN ( train_set , test_sample )
Begin
For i := 1 to train_size
Validity(i) := Compute Validity of i-th sample;
End for;
Output_label:=Weighted_KNN(Validity,test_sample);
Return Output_label ;
End.
```

Fig.1. Pseudo-code of the MKNN Algorithm

III. VALIDITY OF THE TRAIN SAMPLES

In the MKNN algorithm, every training sample must be validated at the first step. The validity of each point is computed according to its neighbors. The validation process is performed for all train samples once. After assigning the validity of each train sample, it is used at the second step as impact or weight of the points in the ensembles of neighbors which the point is selected to attend.

To validate a sample point in the train set, the H nearest neighbors of the point is considered. Among the H nearest neighbors of a train sample x , $validity(x)$ counts the number of points with the same label to the label of x . Eq. 1 is the formula which is proposed to compute the validity of every points in train set.

$$Validity(x) = \frac{1}{H} \sum_{i=1}^H S(lbl(x), lbl(N_i(x))) \quad (1)$$

Where H is the number of considered neighbors and $lbl(x)$ returns the true class label of the sample x . also, $N_i(x)$ stands for the i th nearest neighbor of the point x . The function S takes into account the similarity between the point x and the i th nearest neighbor. Eq. 2 defines this function.

$$S(a,b) = \begin{cases} 1 & a = b \\ 0 & a \neq b \end{cases} \quad (2)$$

IV. APPLYING WEIGHTED KNN

Weighted KNN is one of the variations of KNN method which uses the K nearest neighbors, regardless of their classes, but then uses weighted votes from each sample rather than a simple majority or plurality voting rule. Each of the K samples is given a weighted vote that is usually equal to some decreasing function of its distance from the unknown sample. For example, the vote might set be equal to $1/(d_e+1)$, where d_e is Euclidian distance. These weighted

votes are then summed for each class, and the class with the largest total vote is chosen. This distance weighted KNN technique is very similar to the window technique for estimating density functions. For example, using a weighted of $1/(d_e+1)$ is equivalent to the window technique with a window function of $1/(d_e+1)$ if K is chosen equal to the total number of training samples [21].

In the MKNN method, first the weight of each neighbor is computed using the $1/(d_e+\alpha)$, where α is a smoothing regulator and here α is selected to 0.5. Then, the validity of that training sample is multiplied on its raw weight which is based on the Euclidian distance. In the MKNN method, the weight of each neighbor sample is derived according to Eq. 3.

$$W(i) = Validity(i) \times \frac{1}{d_e + \alpha} \quad (3).$$

Where $W(i)$ and $Validity(i)$ stand for the weight and the validity of the i th nearest sample in the train set. This technique has the effect of giving greater importance to the reference samples that have greater validity and closeness to the test sample. So, the decision is less affected by reference samples which are not very stable in the feature space in comparison with other samples. In other hand, the multiplication of the validity measure on distance based measure can overcome the weakness of any distance based weights which have many problems in the case of outliers. So, the proposed MKNN algorithm is significantly stronger than the traditional KNN method which is based just on distance.

V. EXPERIMENTAL RESULTS

This section discusses the experimental results and compares the MKNN method with original KNN algorithm.

1) Data sets

The proposed method is evaluated on nine standard data sets, namely Iris, Wine, Isodata, SAHeart, Balance-scale, Bupa and Monk's problems (including three problems). None of the databases had missing values, as well as they use continuous attributes. These standard data sets which are obtained from UCI repository [22] are described as follows. The iris database which is possibly one of the most frequently used benchmarks for evaluating classification and clustering algorithms is a well-defined problem with clear separating class boundaries. The data set contains 150 instances using three classes, where each class refers to a type of iris plant, namely Setosa, Versicolour and Virginica. This database uses four continuous attributes: sepal length, sepal width, petal length and petal width. The Wine data set is the result of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines. This data set has been used with many others for comparing various classifiers. In a classification context, this is a well-posed problem with well-behaved class structures. It has

three classes with 59, 71 and 48 instances. The more detail information about the wine data set is described in [23].

The Isodata set is the first test case in this study which is a two class data set and has 34 features as well as the 351 sample points. The SAHeart data set which is obtained from www-stat.stanford.edu/ElemStatLearn is a retrospective sample of males in a heart-disease high-risk region of the Western Cape, South Africa. There are roughly two controls per case of CHD. Many of the CHD positive men have undergone blood pressure reduction treatment and other programs to reduce their risk factors after their CHD event. In some cases the measurements were made after these treatments. This data set has nine continuous features and two classes with the number of 463 instances. These data are taken from a larger dataset, described in [24].

The Balance-scale data set was generated to model psychological experimental results. Each example is classified as having the balance scale tip to the right, tip to the left, or be balanced which means three classes. The attributes are the left weight, the left distance, the right weight, and the right distance. It means this data set has four attributes. It has total 625 samples which include 49 balanced, 288 left, 288 right. The Bupa data set is a two class data set for classification. It contains 345 data sample as well as six attributes. In these six data sets, the instances are divided into training and test sets by randomly choosing 90% and 10% of instances per each of them, respectively. Also, all above mentioned data sets are become normalized with the mean of 0 and variance of 1, $N(0,1)$ before applying the algorithms. The last experimented data set is Monk's problem which is the basis of a first international comparison of learning algorithms. There are three Monk's problems. The domains for all Monk's problems are the same. The second Monk's problem has added noise. For each problem, the domain has been partitioned into a train and test set. The number of Instances and attributes in all three problems are respectively, 432 and 6. These problems are two class problems. The train and test sets in all three Monk's problems are predetermined. The train sets in Monk 1, 2 and 3 are 124, 169 and 122, respectively.

2) Experiments

All experiments are evaluated over 500 independent runs and the average results of these examinations are reported. In all experiments, the number of considered neighbors (the value of parameter H in Eq. 1) is set to a fraction of the number of train data which is empirically set to 10% of the train size. Table 1 shows the results of the performance of classification using the presented method, MKNN, and traditional method, original version of KNN, comparatively. The experiments show that the MKNN method significantly outperforms the KNN method, with using different choices

of value K , over large variety of datasets. It is obvious that more information usually yields to more classification performance. Because of the MKNN classification is based on validated neighbors which have more information in comparison with simple class labels, it outperforms the KNN algorithm in performance. Fig 2 investigates the effect of parameter K on accuracy of algorithms KNN and MKNN comparatively in four different data sets: Iris, Balance-scale, Bupa and SAHeart. The value of K is the odd numbers in the range of [3-15]. Although usually the MKNN method initially overwhelms the KNN algorithm, the results of two algorithms gradually close to each other by growing the value of K . It can be because of the larger values of K result in invalidity of the validity of train samples. For some data sets the KNN even dominates the MKNN method with large K (see Fig 2b and 2c). In addition, since computing the validity measure is executed only once in training phase of the algorithm, computationally, the MKNN method can be applied with the nigh same burden in comparison with the weighted KNN algorithm.

VI. CONCLUSION

In this paper, a new algorithm for improving the performance of KNN classifier is proposed which is called Modified K-Nearest Neighbor, MKNN. The proposed method which considerably improves the performance of KNN method employs a kind of preprocessing on train data. It adds a new value named "Validity" to train samples which cause to more information about the situation of training data samples in the feature space. The validity takes into accounts the value of stability and robustness of the any train samples regarding with its neighbors. Applying the weighted KNN which employs validity as the multiplication factor yields to more robust classification rather than simple KNN method, efficiently. The method evaluated on nine different benchmark tasks: Wine, Isodata, Iris, Bupa, Inosphere and three Monk's problems. The results confirm authors' claim about its robustness and accurateness unanimously. So this method is better in noisy datasets and also in the case of outliers. Since the outliers usually gain low value of validity, it considerably yields to robustness of the MKNN method facing with outliers. The experiments on Monk 2 approve this claim.

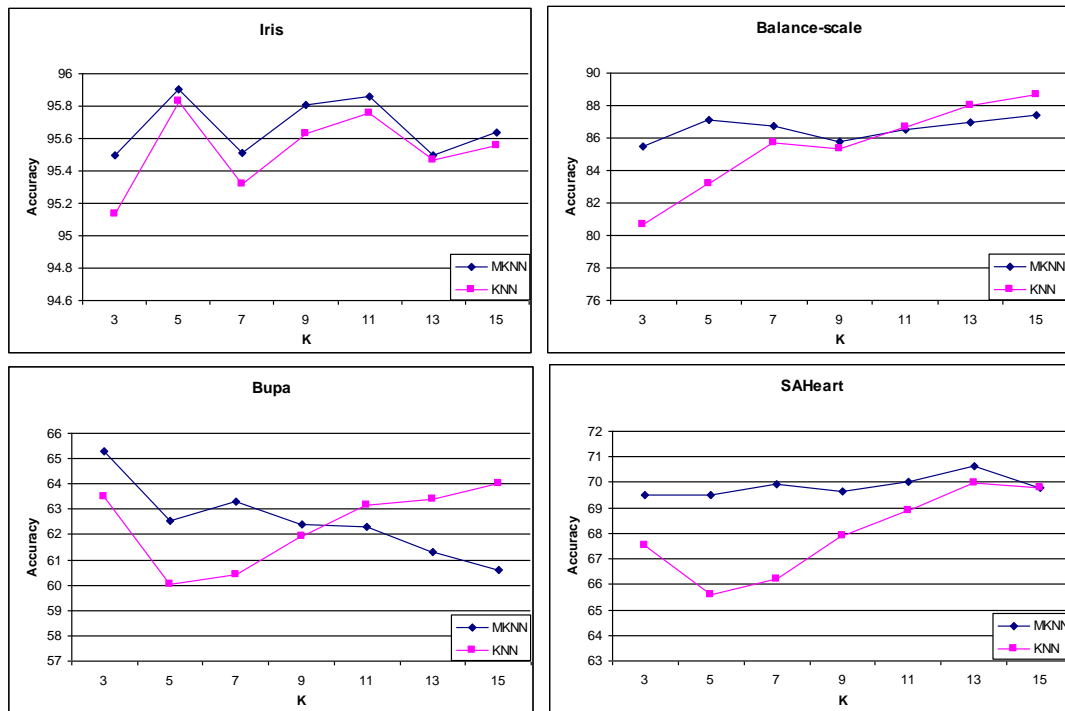


Fig.2. The effect of parameter K on accuracy of algorithms KNN and MKNN comparatively in four data sets (a) Iris (b) Balance scale (c) Bupa (d) SAHeart.

TABLE 1. Comparison of recognition rate between the MKNN and KNN algorithm (%).

	K=3		K=5		K=7	
	KNN	MKNN	KNN	MKNN	KNN	MKNN
Monk 1	84.49	87.81	84.26	87.81	79.86	86.65
Monk 2	69.21	77.66	69.91	78.01	65.74	77.16
Monk 3	89.12	90.58	89.35	90.66	88.66	91.28
Isodata	82.74	83.52	82.90	83.32	80.50	83.14
Wine	80.89	83.95	83.79	85.76	80.13	82.54
Iris	95.13	95.50	95.83	95.90	95.32	95.51
Balance-sc	80.69	85.49	83.22	87.10	85.74	86.77
Bupa	63.51	63.30	60.01	62.52	60.41	63.29
SAHeart	67.51	69.51	65.59	69.49	66.21	69.95

VII. REFERENCES

- 1) L. I. Kuncheva, Combining Pattern Classifiers, Methods and Algorithms, New York: Wiley, 2005.
- 2) H. Parvin, H. Alizadeh, B. Minaei-Bidgoli and M. Analoui, An Scalable Method for Improving the Performance of Classifiers in Multiclass Applications by Pairwise Classifiers and GA", In Proc. of the Int. Conf. on Networked Computing and advanced Information Management by IEEE CS, (NCM08), Sep. 2008.
- 3) H. Parvin, H. Alizadeh, M. Moshki, B. Minaei-Bidgoli and N. Mozayani, Divide & Conquer Classification and Optimization by Genetic Algorithm", In Proc. of the Int. Conf. on Convergence and hybrid Information Technology by IEEE CS, (ICCIT08), Nov. 11-13, 2008.
- 4) H. Parvin, H. Alizadeh, B. Minaei-Bidgoli and M. Analoui, CCHR: Combination of Classifiers using Heuristic Retraining", In Proc. of the Int. Conf. on Networked Computing and advanced Information Management by IEEE CS, (NCM 2008), Korea, Sep. 2008.
- 5) H. Parvin, H. Alizadeh and B. Minaei-Bidgoli, A New Approach to Improve the Vote-Based Classifier Selection", In Proc. of the Int. Conf. on Networked Computing and advanced Information Management by IEEE CS, (NCM 2008), Korea, Sep. 2008.

- 6) H. Alizadeh, M. Mohammadi and B. Minaei-Bidgoli, Neural Network Ensembles using Clustering Ensemble and Genetic Algorithm", In Proc. of the Int. Conf. on Convergence and hybrid Information Technology by IEEE CS, (ICCIT08), Nov. 11-13, 2008, Busan, Korea.
- 7) H. Parvin, H. Alizadeh and B. Minaei-Bidgoli, A New Method for Constructing Classifier Ensembles, International Journal of Digital Content: Technology and its Application, JDCTA, ISSN: 1975-9339, 2009 (in press).
- 8) H. Parvin, H. Alizadeh and B. Minaei-Bidgoli, Using Clustering for Generating Diversity in Classifier Ensemble, International Journal of Digital Content: Technology and its Application, JDCTA, ISSN: 1975-9339, 2009 (in press).
- 9) B.V. Darasay, Nearest Neighbor pattern classification techniques, Las Alamitos, LA: IEEE CS Press.
- 10) E. Fix, J.L. Hodges, Discriminatory analysis, nonparametric discrimination: Consistency properties. Technical Report 4, USAF School of Aviation Medicine, Randolph Field, Texas, 1951.
- 11) Cover, T.M., Hart, P.E. Nearest neighbor pattern classification. IEEE Trans. Inform. Theory, IT-13(1):21-27, 1967.
- 12) Hellman, M.E. The nearest neighbor classification rule with a reject option. IEEE Trans. Syst. Man Cybern., 3:179-185, 1970.
- 13) K. Fukunaga, L. Hostetler, k-nearest-neighbor bayes-risk estimation. IEEE Trans. Information Theory, 21(3), 285-293, 1975.
- 14) S.A. Dudani, The distance-weighted k-nearest-neighbor rule. IEEE Trans. Syst. Man Cybern., SMC-6:325-327, 1976.
- 15) T. Bailey, A. Jain, A note on distance-weighted k-nearest neighbor rules. IEEE Trans. Systems, Man, Cybernetics, Vol. 8, pp. 311-313, 1978.
- 16) S. Bermejo, J. Cabestany, Adaptive soft k-nearest-neighbour classifiers. Pattern Recognition, Vol. 33, pp. 1999-2005, 2000.
- 17) A. Jozwik, A learning scheme for a fuzzy k-nn rule. Pattern Recognition Letters, 1:287-289, 1983.
- 18) J.M. Keller, M.R. Gray, J.A. Givens, A fuzzy k-nn neighbor algorithm. IEEE Trans. Syst. Man Cybern., SMC-15(4):580-585, 1985.
- 19) K. ITQON, Shunichi and I. Satoru, Improving Performance of k-Nearest Neighbor Classifier by Test Features, Springer Transactions of the Institute of Electronics, Information and Communication Engineers 2001.
- 20) R. O. Duda, P. E. Hart and D. G. Stork, Pattern Classification, John Wiley & Sons, 2000.
- 21) E. Gose, R. Johnsonbaugh and S. Jost, Pattern Recognition and Image Analysis, Prentice Hall, Inc., Upper Saddle River, NJ 07458, 1996.
- 22) C.L. Blake, C.J. Merz, UCI Repository of machine learning databases:
<http://www.ics.uci.edu/~mllearn/MLRepository.html>, 1998.
- 23) S. Aeberhard, D. Coomans and O. de Vel, Comparison of Classifiers in High Dimensional Settings, Tech. Rep. no. 92-02, Dept. of Computer Science and Dept. of Mathematics and Statistics, James Cook University of North Queensland.
- 24) Rousseauw et al., South African Medical Journal, 1983.

IEEE WLANs Standards for Mobile Ad-hoc Networks (MANETs): Performance Analysis

Ajay Jangra¹, Nitin Goel², Priyanka³, Komal Kumar Bhatia⁴

GJCST Classification
C.2.5

Abstract- Mobile ad hoc networks(MANETs) catch the great attention of researchers because of its unique characteristics like independent of infrastructure, dynamic topology behavior, limited energy source, multihop routing, mobility of nodes etc. All these characteristics chase lots of challenges about security, reliability and performance issues. In this paper we present a comparative analysis a set of IEEE WLANs standards family used for MANETs (i.e IEEE 802.11, 802.15, 802.15.4, 802.16 and 802.20). We also present a close look to the performance of various IEEE standards and check their co-existence nature with zero interference occurrence. In this paper we give more emphasis how to choose one reliable technology for any wireless network on the bases of your required network attributes out of all wireless technologies.

Keywords- Mobile Ad Hoc Network (MANETs), Complimentary Code Keying(CCK), Direct-Sequence Spread Spectrum (DSSS), Multiple-Input/Multiple-Output (MIMO), Orthogonal Frequency-Division Multiplexing (OFDM), Binary Phase Shift Keying (BPSK), Offset Quadrature phase-shift keying (O-QPSK).

I. INTRODUCTION TO IEEE 802.15, PAN

Bluetooth is wireless LAN technology design to connect devices of different functions such as telephones, notebooks, computers (desktop and laptop), cameras, printers, coffee makers, ad so on. Bluetooth was originally started as a project by the Ericsson Company and later formalized by the Bluetooth Special Interest Group (SIG), includes Sony Ericsson, IBM, Intel, Toshiba and Nokia. In an ad hoc environment (infrastructure less) all Bluetooth (IEEE 802.15 PAN) devices (also called gadgets), find the existence of each other and finally make a network called piconet. We could also connect Bluetooth LAN to the internet by the means of gadgets capability. A Bluetooth network can't be large; if it is then there is chaos. Bluetooth use to communicate devices like PDA and peripherals. Bluetooth technology has several applications. With Bluetooth we could use wireless mouse and keyboard, monitoring devices with sensor devices, home security devices. [2, 3]

About^{1&2} - CSE department UIET, Kurukshetra University, Kurukshetra (INDIA) (+91-9466027922 er_jangra@yahoo.co.in)¹ (+91-9416407079 goelnitin0887@gmail.com)²

About³ - ECE department Kurukshetra Institute of Technology & Management (INDIA). (+91-9466751345 priyanka.jangra@gmail.com)³

About⁴ - CSE department YMCA University of Science & Technology, Faridabad (INDIA) (+91-9953637670 komal_bhatia1@yahoo.com)⁴

1) Characteristics & Architecture

A Bluetooth device has a built-in short range radio transmitter. The data rate of Bluetooth device is 1 Mbps with a 2.4-GHZ bandwidth. By this figure it might be possible it could interfere with IEEE.11b wireless LAN. Bluetooth follows two types of networks: piconet and scatternet. Where Piconet: A small network and can have up to eight nodes, among them one of which is called master; the rest are called slaves. In order to communicate between master and slave, they can be one-to-one or one-to-many. The maximum numbers of slaves in a piconet can have seven. Scatternet: Two or more piconet combined together to form a scatternet. A slave node in one piconet can be the master in another piconet. This node can receive messages from both the piconets and act as master/slave at the same time and they can communicate by means of multi hopping

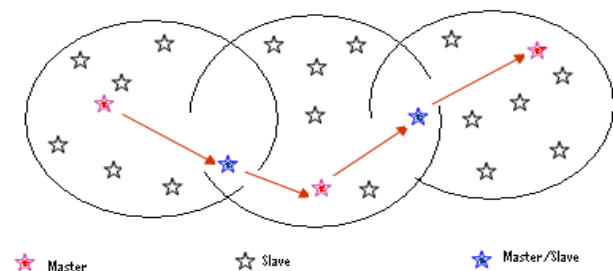


Fig 1:- Master/slave Configuration for Piconet/Scatternet
Bluetooth Characteristics

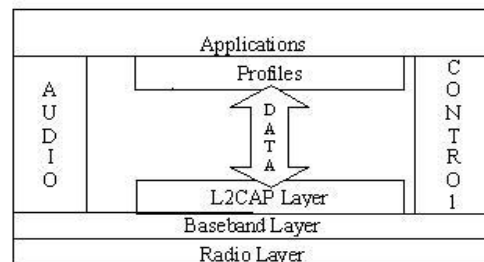


Fig. 2 Architecture

The radio layer is somewhat similar to the physical layer of the Internet model. Bluetooth devices are low power and have range of 10 m. It uses a 2.4 GHz ISM band which are further divided into 79 channels of 1 MHz each. In order to

avoid the interference from other devices and other network it uses frequency hopping spread spectrum method and it hops 1600 tps. The baseband layer works as the MAC layer in LANs. In order to access the given layer it uses the TDMA. All the master and slave communicate using time slot (625 μ s). In the case of master and slave configuration Bluetooth can create two types of physical links: SCO (synchronous connection-oriented) is used when first preference goes to avoid latency rather than error free delivery (integrity). ACL (asynchronous connectionless link) in the reverse case of SCO means when first preference goes to error free delivery (integrity). The logical Link Control and Adaptation Protocol (L2CAP) is similar to LLC sub-layer in LANs. ACL physical link follows the said protocol; SCO doesn't follow it. L2CAP has data packet and format as:

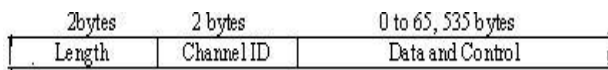


Fig. 3

Finally we could see that how Bluetooth technology works and can be best suited for ad hoc network.[2,3]

II. WIRELESS FIDILITY IEEE 802.11 (Wi-Fi)

Wireless Fidelity (Wi-Fi) standard given by the Wi-Fi Alliance to the IEEE 802.11 can also use by its extensions like 802.11a, 802.11b and 802.11g. There are several more extensions are going to be introduced by Wi-Fi (802.11) i.e. 802.11e, 802.11i, 802.11p and 802.11n.

Wi-Fi is the transmission of radio signals. In order to define data transmission and manages location independent network access using radio signals on the basis of that we can call it as a packet protocol. The structure of physical/link layer interface of Wi-Fi is similar to Ethernet. The layers above the physical and data link layers include TCP/IP. By the above introduction we can clearly see all programs and applications for TCP/IP that can run on an Ethernet can also be run on Wi-Fi interface.[2,3,8,7]

1) Architecture

All extensions of 802.11 like (a, b, g, etc) follows five layers TCP model and data link layer is divided into two sub-layers: Media Access Control (MAC): Logical Link Control (LLC). Five layer TCP Model.

TCP/IP	IEEE
Application Layer	Application Layer
Transport Layer	Transport Layer
Network layer	Network layer
Data link Layer (DDL)	Logical link Control (LLC)
	Media Access Control (MAC)
Physical Layer	Physical Layer

Fig. 4

The two lower layers are specified by IEEE 802 networks: 802.2 LLC, 802.3 PHY and MAC, Ethernet, 802.11 PHY and MAC, Wi-Fi. The primary task of physical layer is to perform encoding scheme, modulation and deals with the transmission of radio signals in actual through space. As we know that physical layer implementation work in particular specific bands means some specific frequency allocated for some specific applications.

Medium Access Control (MAC) layer is sub-layer of Data link Layer (DDL). The primary task of said layer is to control the transmission and sometimes it can be used to provide services like mobility management. The format of MAC address of Wi-Fi is similar to the MAC address of Ethernet (802.3) i.e. 6 octets. It uses the CSMA/CA (means to prevent collision before they happen) is similar to the mechanism used in Ethernet (802.3) i.e. CSMA/CD. Both act as a peer-to-peer protocol.

IEEE 802.11 defines lots of services but in this paper we only discuss the two main services i.e. Station services (SS): The services defined by: Authentication, De-authentication and privacy etc. and Distributed System Services (DSS): The services defined by: Association, Re-association, Disassociation, Distribution and Integration.[3,8]

2) Comparison of IEEE 802.11 a/b/g/n Standards

Wi-Fi Characteristics	IEEE 802.11 Protocols			
	802.11a	802.11b	802.11g	802.11n*
Operating Frequency	5.3 GHz and 5.8 GHz	2.4 GHz		2.4 GHz or 5 GHz
Average Signal Range**	~30 to 35 m			~60 to 70 m
Available Bandwidth per Channel	~20 to 22 MHz			20 or 40 MHz
Data Rate (Min.-Max)	6, 9, 12, 18, 24, 36, 48, 54 Mbps 6, 12, and 24 Mbps are mandatory	1, 2, 5.5, 11 Mbps	1, 2, 5.5, 11, 6, 9, 12, 18, 24, 33, 36, 48, 54 Mbps 1, 2, 5.5, 11, 6, 12 and 24 Mbps are mandatory	1, 2, 5.5, 6, 9, 11, 12, 18, 24, 36, 48, 54, 121.5, 130, 144.44, 270, 300 Mbps
Typical Throughput for Max Data Rate	18 to 22 Mbps	6 Mbps	18 to 22 Mbps	74 Mbps
Modulation Technique	OFDM	CCK or DSSS	OFDM	OFDM using MIMO and CB
Channels	36, 40, 44, 48, 52, 56, 60, 64, 149, 153, 157, 161	1-11		3 non-overlapping 2.4 GHz 12 non-overlapping UNII channels in 5 GHz
Special Considerations	HF signals have more trouble with physical obstruction	2.4 GHz subject to interference from: Bluetooth products, cordless phones, microwaves, radar, remote controls, ZigBee networks, etc.		

Fig. 5

*not finalized, **dependent on physical environment [3,8]

Fig 4. Shows IEEE 802.11 standard can be best suited for ad hoc network. In order to explore Wi-Fi we have some more extensions that are going to be implemented in the field of communication like:

802.11e – It follows quality of services, such as Streaming multimedia or VOIP.

802.11i – It is used for high security implementations.

802.11p – It is used to support exchange of data between high speed vehicles.

III. ZIGBEE IEEE 802.15.4

ZigBee technology (similar to Bluetooth) provides low data rate and low power connectivity for gadgets that follows low battery life as long as several months to several years. ZigBee has low cost and built to perform wireless networking protocol targeted towards automation and remote control application. In the started couple of work has been done for low data rate standards but later on the

ZigBee Alliance and the IEEE decided to join forces and ZigBee is the final name for the given technology.[3,4,5]

The main features of ZigBee are developed for application with relaxed throughput requirements which cannot handle the power consumption of heavy protocol stack, very low power consumption, low data rate in an ad hoc self-organizing network among inexpensive fixed, low cost, network flexibility moving and portable devices.

1) Characteristics

ZigBee also be implemented in highly meshed network but the same we can't pursue with Bluetooth. For a given application depending on the given radio frequency environment and power output consumption, so ZigBee could able to access wireless devices expected to transmit at 10-75 meters. It can operate in unlicensed radio frequency worldwide like 2.4 GHz global, 915 MHz Americas or 868 MHz Europe along with data rate 250 kbps at 2.4GHz, 40 kbps at 915 MHz and 20 kbps at 868 MHz.

2) Architectural Specification

On the specification of PHY and MAC, we can able to access it on different types of networks like as star, mesh and cluster tree. Communication redundancy elimination is a unique feature of Zigbee network layer. PHY layer works on order to include energy and link quality detection, add clear channel assessment for improved coexistence with other wireless networks. After all to ensure power conservation, and low latency through guaranteed time slots routing schemes are designed.[4,5]

PHY (MHz)	Frequency Band (MHz)	Spreading parameters		Data parameters		
		Chip rate (kchip/s)	Modulation	Bit rate (kb/s)	Symbol rate (ksymbols/s)	symbols
868/915	868-868.6	300	BPSK	20	20	Binary
	902-928	600	BPSK	40	40	Binary
2450	2400-2483.5	2000	O-QPSK	250	62.5	16-ary Orthogonal

Fig. 6 Information regarding frequency bands and data rates

Power required in	Bluetooth	Wi-Fi
Transmit (mA)	50-100	340-450
Receive (mA)	50-80	250-310
Idle/Sleep (mA)	1.5-2	10-32

Fig. 7: Typical Bluetooth and Wi-Fi power requirement

IV. COMPARISON OF IEEE 802.15/11/15.4

Wireless Parameter	Bluetooth	Wi-Fi	ZigBee
Frequency band	2.4 GHz	2.4 GHz	2.4 GHz
Physical/MAC layers	IEEE 802.15	IEEE 802.11	IEEE 802.15.4
Protocol stack size	250 KB	1 MB 32 KB	32 KB, 4 KB (for limited function end devices)
Minimum quiet	15 MHz	22 MHz	3 MHz
bandwidth required	(dynamic)	(static)	(static)
Number of channels	19	13	16
Maximum no. of nodes/nw	7	32 per access point	64 K
Raw data rate	1 Mbps	11 Mbps	250 Kbps
Range	9 m	75 to 90 m	Indoors : up to 30 m Outdoors : up to 100 m
Current consumption	60 mA (Tx mode)	400 mA (Tx mode) 20mA (Standby mode)	25-35 mA (Tx mode) 3 μ A (Standby mode)
Typical network join time	≥ 3 sec	variable, 1 sec typically	30 ms typically
Interference avoidance method	FHSS	DSSS	DSSS

Fig. 8 Comparison Table [3, 4, 6, 8, 9]

V. IEEE 802.16 Wi-MAX (WORLDWIDE INTEROPERABILITY FOR MICROWAVE ACCESS,):

As we know in the modern era of broadband wireless access, WiMAX (IEEE 802.16) is an outstanding, well suitable, useful connection oriented protocol to which access fixed and mobile with low cost, high reliability, very high data rate and better efficiency. WiMAX standard defines the formal specification for deployment of broadband wireless metropolitan area networks (wireless MANs) and with the help of WiMAX (802.16) we can access broadband anytime, on virtually any device and anywhere.

While moving at a speed of approximate 125 kmph, in that speed we can also be able to access broadband. WiMAX (802.16) has data rate up to 70 mbps and can be able to work in both license free and licensed band and have high efficiency. WiMAX can have coverage area approximately is up to 50 km.

People full interest on the communication when there is a better throughput, better reliability and better efficient, so in order to fulfill that in WiMAX couples of mechanisms are built into the physical and MAC layers of WiMAX Which also pursue our communication with full of Qos and high security through WiMAX (802.16).[3,7,8,9]

1) Features of IEEE 802.16

Property	Values
Range	50km(*LOS), 6-8 (*NLOS)
Data Rate	70 Mbps (shared)
Spectrum	2.3 – 2.7 Ghz, 3.4 – 3.6 Ghz, 5.8 Ghz (unlicensed)
Access Types	Fixed, nomadic/portable, mobile
Spectral efficiency	3.7 (bits/s)/Hz (for 802.16-2004)
Channel Size	1.5 MHz to 20 MHz (flexible)
Modes of Operation	Point-to-point, point-to-multipoint, mesh, multihop mode

Fig. 9

*In particular communication when there are several obstacles like trees and building then that type of communication is called line-of-sight (LOS) communication. Straight line communication is called non line-of-sight (NLOS) like communication in the house etc.

2) WiMAX (802.16) Architecture

WiMAX (802.16) follows the unidirectional flow of packet known as service flow and have some set of QoS. These kind of service flow is identified as a 32-bit service flow identifier. WiMAX (802.16) is a connection oriented protocol and its connection identified by a 16-bit connection identifier (CID). WiMAX (802.16) Physical Layer: IEEE 802.16 in 2001, microwave frequencies (10-66 GHz), high data rate, LOS and PTP mode and fixed subscriber stations only. IEEE 802.16-2004- in 2004, frequencies (2-11 GHz) and (10-66 GHz bands), medium data rates, PTP, LOS, NLOS and PMP modes, fixed subscriber stations only. IEEE 802.16-2005- in 2005, low to medium data rates, PTP and PMP modes, both fixed and mobile subscriber stations. [8]

3) Physical and MAC layers

MAC	Service Specific convergence sublayer (CS)	Management Entity
	MAC Common Part sublayer (MAC CPS)	
	Security Sublayer	
PHY	Physical Layer (PHY)	

Fig. 10

Service Specific convergence sub-layer (CS): converts higher layer packets into MAC level Service data Units (SDUs), interfaces with higher layers and finally maps higher-level transmission parameters to MAC level service flow and connection parameters. MAC Common Part sub-layer (MAC CPS): follows retransmissions, QoS, link initialization, transmission scheduling, error control, fragmentation, control channel access and link initialization. Security Sub-layer: it uses X.509 standard for certificate-based identification of nodes, encryption, key management, key distribution and authentication. [7]

Next on the agenda are:

- 802.16c/d, published in Jan 2003, c' protocol relates to protocols, test suite structures and test purposes while d' fixes errata and protocols not covered in c', and creates the system profiles.
- 802.16e, which adds mobility to the standard and really throws down the gauntlet to cellular

VI. IEEE 802.20 (MOBILE-FI)

Mobile-Fi (IEEE 802.20) is the youngest IEEE standard. In order to access fully mobile broadband, it is the first standard designed to carry native IP traffic with licensed airwave below 3.5 GHz and provides symmetrical wireless rates over long distance (~15km). Mobile-Fi, enables world wide deployment of affordable, ubiquitous, always-on and interoperable mobile broadband wireless access networks that meet the needs of business and residential

market” this the first theme behind IEEE 802.20 uses packet based air interference optimized for transport of IP services. If we compare all the factors with

Characteristics	WiMax	Wi-Fi
SECURITY	Licensed and license-exempt 128-bit Triple-DES and 1024-bit RSA security	License-exempt only WPA+WEP security, inadequate though 802.11i will improve
COVERAGE	Optimised for outdoor non-line of sight mesh networks, advanced smart antenna	Optimised for indoor, No mesh, Smart antenna
RANGE	50km Point to multipoint, Tolerant of greater multipath delay spread up to 10ms, PHY and MAC designed for multiple range Standard MAC	100 meters Point to point, PHY and MAC optimised for 100m range, Range can be extended but then MAC non standard
QoS	Grant request MAC Designed to support voice and video from the start Supports differentiated service levels e.g. TDD/FDD/HFDD -symmetric or asymmetric, Centrally enforced QoS	Contention-based MAC, no guaranteed QoS Standard cannot guarantee latency for voice or video, TDD only-asymmetric Proposed 802.11e QoS standard is prioritisation only
PERFORMANCE	Bandwidth 10, 20MHz; 1.75, 3.5, 7, 14Hz; 3, 6MHz, Maximum data rate 70Mbps Maximum 5.0 bps/Hz	Bandwidth 20MHz, Maximum data rate 54Mbps, Maximum 2.7 bps/Hz
SCALABILITY	Scalorization, Scalable independent of bandwidth with 15MHz to 20MHz width channels MAC supports thousands of users	MAC supports tens of users

Fig. 11 WiMax and Wi-Fi comparison table [2, 3]

VII. CONCLUSION

Computer networks specially Mobile Ad hoc Network is the emerging field and most popular for Research, Applications and Services. Different WLANs standards exhibit various characteristic. In this paper we have analyzed different WLANs IEEE standards like Bluetooth, Wi-Fi, WiMAX, ZigBee and Mobile-Fi and check their usability in ad hoc network with their characteristics, merits and demerits of each. On the bases of said characteristics of each, we could choose particular technology depends upon our network. We conclude all of the above said technologies are complimentary to each other rather than competing. We can also say that with some prerequisite arrangements the above said technologies can perform well in interference free co-existence environment.

VIII. REFERENCES

- 1) Ajay Jangra, Nitin Goel, Priyanka, Komal, Security Aspects in Mobile Ad Hoc Network (MANETs): A Big Picture”, International Journal of Electronics Engineering, 2(1), 2010, pp. 189-196.
- 2) C.Siva Ram Murthy & B.S Manoj, Mobile Ad Hoc Networks- Architectures & Protocols”, Pearson Education, New Delhi, 2004.
- 3) Behrouz A Forouzan, Data Communications and Networking”, Special Indian Forth Edition, 2006
- 4) Sinem Coleri ,Ergen, ZigBee IEEE 802.15.4” September 10, 2004
- 5) LAN-MAN Standards Committee of the IEEE Computer Society, Wireless LAN medium access control (MAC) and physical layer (PHY) specification, IEEE, New York, NY, USA, IEEE Std 802.11-1997 edition, 1997
- 6) N. Srinath [CS07M035] WiMAX - An Introduction”
- 7) N. Gupta and G. Kaur, WiMAX: Applications,” ser. The WiMAX Handbook, S. Ahson and M. Ilyas, Eds. CRC Press (Taylor and Francis Group), 2008, ch. 3: WiMAX Technology for Broadband Wireless Communication, pp. 35 – 54, ISBN 9781420045474.
- 8) An Introduction to Wi-Fi” 019-0170 • 090409-B USA 2007-2008
- 9) Caroline Gabriel, WiMax”, ARCchart Ltd., London EC2A 1LN
- 10) B R Sujatha, M V Satyanarayana, Improved Network Connectivity in MANETs”, International Journal of Computer Networks & Communications (IJCNC), Vol.1, No.3, October 2009
- 11) Ajay Jangra, Sunita Beniwal, Anil Garg, Co-existence behavior study of Bluetooth & Wi-Fi for 2.4 GHz ISM band” 2006

Analysis of Liver Disorder Using Data mining Algorithm

P.Rajeswari¹, G.Sophia Reena²

GJCST Classification
H.2.8,J.3

Abstract-there are many disorders of the liver that require clinical care by a physician or other healthcare professional. The study of liver development has significantly contributed to developmental concepts about morphogenesis and differentiation of other organs. Knowledge of the the understanding of human congenital diseases. Significantly, much of understanding of organ development has arisen from analyses of patients with liver deficiencies. In this paper the data classification is based on liver disorder the training data set is developed by collecting data from UCI repository consists of 345 instances with 7 different attributes. The instances in the dataset are pertaining to the two categories of blood tests which are thought to be sensitive to liver disorders that might arise from excessive alcohol consumpt mechanisms that regulate hepatic epithelial cell differentiation has been essential in creating efficient cell culture protocols for programmed differentiation of stem cells to hepatocytes as well as developing cell transplantation therapies. Such knowledge also provides a basis for ion, labeled as Low (L), and (H) to represent the profit as 0 and 1 which result in accuracy and time taken to build the algorithm. WEAK tool is used to classify the data and the data is evaluated using 10-fold cross validation and the results are compared.

Keywords-Naive Bayes, KStar, FT Tree, WEKA tool.

I. INTRODUCTION

The liver is the largest glandular organ of the body. It weighs about 3 lb (1.36 kg). It is reddish brown in colour and is divided into four lobes of unequal size and shape. The liver lies on the right side of the abdominal cavity beneath the diaphragm. Blood is carried to the liver via two large vessels called the hepatic artery and the portal vein. The hepatic artery carries oxygen-rich blood from the aorta (a major vessel in the heart). The portal vein carries blood containing digested food from the small intestine. These blood vessels subdivide in the liver repeatedly, terminating in very small capillaries. Each capillary leads to a lobule. Liver tissue is composed of thousands of lobules, and each lobule is made up of hepatic cells, the basic metabolic cells of the liver. This paper describes Excessive consumption of alcohol can cause an acute or chronic inflammation of the liver and may even harm other organs in the body, alcohol induced liver disease remains a major problem. This paper also describes the blood test taken when a person is affected to liver disorder such as alkaline phosphatase, alamine aminotransferase, aspartate aminotransferase, gamma-glutamyl transpeptidase.

About¹M.phil Research Scholar P.S.G.R.Krishnammal College for Women
About²-HOD (BCA) Dept, P.S.G.R.Krishnammal College for Women
#pa.raji.87@gmail.com

II. CAUSES OF LIVER DISEASE

When the liver becomes diseased, it may have many serious consequences. Liver disease (also called hepatic disease) is a broad term describing any single number of diseases affecting the liver. Many are accompanied by jaundice caused by increased levels of bilirubin in the system. The bilirubin results from the breakup of the hemoglobin of dead red blood cells; normally, the liver removes bilirubin from the blood and excretes it through bile.

1) Disease of Liver

Several diseases states can affect the liver. Some of the diseases are Wilson's disease, hepatitis (an inflammation of the liver), liver cancer, and cirrhosis (a chronic inflammation that progresses ultimately to organ failure). Alcohol alters the metabolism of the liver, which can have overall detrimental effects if alcohol is taken over long periods of time. Hemochromatosis can cause liver problems.

2) Common Liver Disorder

Fatty liver (also known as steatorrhoeic hepatitis or steatosis hepatitis) is a reversible condition where large vacuoles of triglyceride fat accumulate in liver cells via the process of steatosis. It can occur in people with a high level of alcohol consumption as well as in people who never had alcohol.

Hepatitis (usually caused by a virus spread by sewage contamination or direct contact with infected body fluids).

Cirrhosis of the liver is one of the most serious liver diseases. It is a condition used to denote all forms of diseases of the liver characterized by the significant loss of cells. The liver gradually contracts in size and becomes leathery and hard. The regenerative activity continues under liver cirrhosis but the progressive loss of liver cells exceeds cell replacement.

Liver cancer. The risk of liver cancer is higher in those who have cirrhosis or who have had certain types of viral hepatitis; but more often, the liver is the site of secondary (metastatic) cancers spread from other organs.

3) Symptoms of Liver Disorder

The external signs include a coated tongue, itchy skin, excessive sweating, offensive body odor, dark circles under the eyes, red swollen and itchy eyes, acne rosacea, brownish spots and blemishes on the skin, flushed facial appearance or excessive facial blood vessels.^[1] Other symptoms include jaundice, dark urine, pale stool, bone loss, easy bleeding, itching, small, spider-like blood vessels visible in the skin, enlarged spleen, and fluid in the abdominal cavity, chills,

pain from the biliary tract or pancreas, and an enlarged gallbladder.^[2] The symptoms related to liver dysfunction include both physical signs and a variety of symptoms related to digestive problems, blood sugar problems, immune disorders, abnormal absorption of fats, and metabolism problems.^[1] Nervous system disorders include depression, mood changes, especially anger and irritability, poor concentration and "foggy brain", overheating of the body, especially the face and torso, and recurrent headaches (including migraine) associated with nausea. The blood sugar problems include a craving for sugar, hypoglycaemia and unstable blood sugar levels, and the onset of type 2 diabetes. Abnormalities in the level of fats in the blood stream, whether too high or too low levels of lipids in the organism. Hypercholesterolemia: elevated LDL cholesterol, reduced HDL cholesterol, elevated triglycerides, clogged arteries leading to high blood pressure heart attacks and strokes, build up of fat in other body organs (fatty degeneration of organs), lumps of fat in the skin (lipomas and other fatty tumors), excessive weight gain (which may lead to obesity), inability to lose weight even while dieting, sluggish metabolism, protuberant abdomen (pot belly), cellulite, fatty liver, and a roll of fat around the upper abdomen (liver roll) etc.^[1] Or too low levels of lipids: hypocholesterolemia: low total cholesterol, low LDL, VLDL cholesterol and low triglycerides.

Cirrhosis of the liver



Fig.1 Snapshot of Cirrhosis of the Liver

Symptoms may include:

- 1) Jaundice
- 2) Tendency to bruise or bleed easily
- 3) ascites
- 4) Impaired brain function
- 5) General failing health

III. ANALYSIS OF LIVER DISORDER

1) Risk Factors

Hepatitis is an inflammation of the liver that can be caused by a virus, inherited disorders, and sometimes by certain medications or toxins such as alcohol and drugs. Scientists have identified four main types of viral hepatitis: hepatitis A, hepatitis B, and hepatitis C, and hepatitis D. A fifth type, hepatitis E, is generally not found in North America.

Hepatitis A is waterborne and spread mainly via sewage and contaminated food and water. Hepatitis B is transmitted by contact with infected semen, blood, vaginal secretions and from mother to newborn. Hepatitis B is most commonly spread by unprotected sex and by sharing of infected needles (including those used for tattooing, acupuncture, and ear piercing). Hepatitis C spreads via direct blood-to-blood contact. Hepatitis D is spread by infected needles and blood transfusions.

Improved screening of donated blood has greatly reduced the risk of catching hepatitis B or C from blood transfusions. Both hepatitis B and C can be spread through sharing of razors, toothbrushes, and nail clippers.

The main cause of cirrhosis is chronic infection with the hepatitis C virus. Other causes include:

- 1) Long-term, excessive alcohol consumption
- 2) Chronic infection with hepatitis B virus
- 3) Inherited disorders of iron and copper metabolism
- 4) Severe reactions to certain medications
- 5) Fatty liver caused by obesity
- 6) Infections from bacteria and parasites usually found in the tropics
- 7) Repeated episodes of heart failure with liver congestion and bile-duct obstruction

With cirrhosis, the liver tissue is irreversibly and progressively destroyed as a result of infection, poison or some other disease. Normal liver tissue is replaced by scars and areas of regenerating liver cells.

IV. EXTRACTION OF LIVER DISEASE DATAWAREHOUSE

The liver disorder data warehouse contains the screening the data of liver disorder patients. Initially, the data warehouse is pre-processed to make the mining process more efficient. In this paper WEKA tool is used to compare the performance accuracy of data mining algorithms for diagnosis liver disease dataset. The pre-processed data warehouse is then classified using WEKA tool. The feature selection in the tool describes the attribute status of the data present in the liver disease. Using supervised machine learning algorithm such as Naive Bayes, FT Tree and Kstar and the result are compared. WEKA is a collection of machine learning algorithms for data mining tasks. The algorithms can be applied directly to a dataset. WEKA contains tools for data classification, Associate, clustering and visualization. It is also well suited for developing new machine learning schemes. This paper concentrates on functional algorithms like Naive Bayes, FT Tree and Kstar.

1) Classification

The basic classification is based on supervised algorithms. Algorithms are applicable for the input data. Classification is done to know the exactly how data is being classified. The Classify Tab is also supported which shows the list of machine learning algorithms. These algorithms in general operate on a classification algorithm and run it multiple times manipulating algorithm parameters or input data weight to increase the accuracy of the classifier. Two learning performance evaluators are included with WEKA.

The first simply splits a dataset into training and test data, while the second performs cross-validation using folds. Evaluation is usually described by the accuracy. The run information is also displayed, for quick inspection of how well a classifier works.

2) *Manifold machine learning algorithm*

The main motivation for different supervised machine learning algorithms is accuracy improvement. Different algorithms use different rule for generalizing different representations of the knowledge. Therefore, they tend to error on different parts of the instance space. The combined use of different algorithms could lead to the correction of the individual uncorrelated errors. As a result the error rate and time taken to develop the algorithm is compared with different algorithm.

3) *Algorithm selection*

Algorithm is selected by evaluating each supervised machine learning algorithms by using supervised learning assessment (10-fold cross-validation) on the training set and selects the best one for application on the test set. Although this method is simple, it has been found to be highly effective and comparable to other methods. Several methods are proposed for machine learning domain. The overall cross validation performance of each algorithm is evaluated. The selection of algorithms is based on their performance, but not around the test dataset itself, and also comprising the predictions of the classification models on the test instance. Training data are produced by recording the predictions of each algorithm, using the full training data both for training and for testing. Performance is determined by running 10-fold cross-validations and averaging the evaluations for each training dataset. Several approaches have been proposed for the characterization of learning domain. The performance of each algorithm on the data attribute is recorded. The algorithms are ranked according to their performance of the error rate.

4) *Manuscript details*

This paper deals with Naive Bayes, KStar, and FT Tree algorithm. Experimental setup is discussed using 345 data

and the results are compared. The performance analysis is done among these algorithms based on the accuracy and time taken to build the model.

V. EXPERIMENTAL SETUP

The data mining method used to build the model is classification. The data analysis is processed using WEKA data mining tool for exploratory data analysis, machine learning and statistical learning algorithms. The training data set consists of 345 instances with 7 different attributes. The instances in the dataset are representing the results of different types of testing to predict the accuracy of liver disease. The performance of the classifiers is evaluated and their results are analysed. The results of comparison are based on 10 ten-fold cross-validations. According to the attributes the dataset is divided into two parts that is 70% of the data are used for training and 30% are used for testing.

1) *Learning Algorithms*

This paper consists of three different supervised machine algorithms learning algorithms derived from the WEKA data mining tool. Which include:

- Naive Bayes,
- KStar
- FT Tree

The above algorithms were used to predict the accuracy of liver disease.

2) *Performance study of Algorithms*

The table I consists of secondary values of different classifications. According to these values the accuracy is calculated and analysed. It has 7 attributes for classification. Each one has a distinct value. Performance can be determined based on the evaluation time of calculation and the error rates. Comparison is made among these classification algorithms out of which the FT Tree algorithm is considered as the better performance algorithm. Because it takes only some time to calculate the accuracy than other algorithms.

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

- weka
 - classifiers
 - bayes
 - AODE
 - AODEsr
 - BayesianLogisticRegression
 - BayesNet
 - ComplementNaiveBayes
 - DMNBtext
 - HNB
 - NaiveBayes**
 - NaiveBayesMultinomial
 - NaiveBayesMultinomialUpdateable
 - NaiveBayesSimple
 - NaiveBayesUpdateable
 - WAODE

to build model: 0.17 seconds

classified cross-validation ==
Y ==

Classified Instances 333 96.5217 %
Y Classified Instances 12 3.4783 %

istic 0.9283
ute error 0.0396
quared error 0.1867
bsolute error 8.1489 %
ive squared error 37.8942 %
er of Instances 345

ed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.958	0.03	0.958	0.958	0.958	0.949	H
	0.97	0.042	0.97	0.97	0.97	0.949	L
Weighted Avg.	0.965	0.037	0.965	0.965	0.965	0.949	

=== Confusion Matrix ===

```

a  b  <-- classified as
137  6 | a = H
 6 196 | b = L
  
```

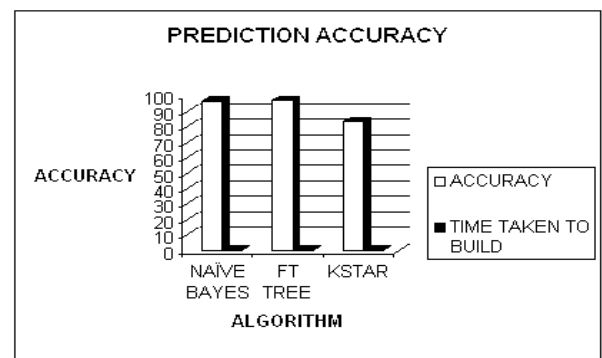
Status: OK

Log x 0

Fig .2 Selection of Naïve Bayes from the Classify Tab

TABLE 11
PERFORMANCE STUDY OF ALGORITHMTABLE I
PERFORMANCE STUDY OF ALGORITHM

Algorithm Used	Accuracy	Time Taken
Naive Bayes	96.52%	0 sec
FT Tree	97.10%	0.2 sec
KStar	83.47%	0 sec



VI. CONCLUSION

Data mining in health care management is unlike the other fields owing to the fact that the data present are heterogeneous and that certain ethical, legal, and social constraints apply to private medical information. Health care related data are voluminous in nature and they arrive from diverse sources all of them not entirely appropriate in structure or quality. These days, the exploitation of knowledge and experience of numerous specialists and clinical screening data of patients gathered in a database during the diagnosis procedure, has been widely recognized. This paper deals with the results in the field of data classification obtained with Naive Bayes algorithm, FT Tree algorithm and KStar algorithm, and on the whole performance made known FT Tree Algorithm when tested on liver disease datasets, time taken to run the data for result is fast when compared to other algorithms. It shows the enhanced performance according to its attribute. Attributes are fully classified by this algorithm and it gives 97.10% of accurate result. Based on the experimental results the classification accuracy is found to be better using FT Tree algorithm compare to other algorithms. From the above results FT Tree algorithm plays a key role in shaping improved classification accuracy of a dataset.

VII. REFERENCES

- 1) Duggan JM, Duggan AE. Systematic review: The liver in celiac disease. *Aliment Pharmacol Ther* 2005; 21: 515-518.
- 2) Cassagnou M, Boruchowicz A, Guillemot F *et al.* Hepatic steatosis revealing celiac disease: a case complicated by transitory liver failure. *Am J Gastroenterol* 1996; 91 : 1291-1292.
- 3) Austin A, Campbell E, Lane P, Elias E. Nodular regenerative hyperplasia of the liver and coeliac disease: potential role of IgA anti-cardiolipin antibody. *Gut* 2004; 53 : 1032-1034.
- 4) Hagander B, Brandt L, Sjolund K, Berg NO, Nordén Å Stenstam M. Hepatic injury in adult coeliac disease. *Lancet* 1977; 2 : 270-272.
- 5) Bonamico M, Pitzalis G, Culasso F *et al.* A. Il danno epatico nella malattia celiaca del bambino. *Minerva Pediatr* 1986; 38: 959-962.
- 6) Leonardi S, Bottaro G, Patané R, Musumeci S. Hypetransaminasemia as the first symptom in infant celiac disease. *J Pediatr Gastroenterol Nutr* 1990; 11 : 404-406.
- 7) Maggiore G, De Giacomo C, Scotta MS, Sessa F. Celiac disease presenting as chronic hepatitis in girl. *J Pediatr Gastroenterol Nutr* 1986; 5 : 501-503.
- 8) Vajro P, Fontanella A, Mayer M *et al.* Elevated Serum minotransferases activity as an early manifestation of glutensensitive enteropathy. *J Pediatr* 1993; 122 : 416-419.
- 9) Volta U, De Franceschi L, Lari F, Molinaro N, Zoli M, Bianchi FB. Coeliac disease hidden by

cryptogenic hypertransaminasemia. *Lancet* 1998; 352 : 26-29.

Performance Analysis of QoS in PMP Mode Wimax Networks

Harwinder singh¹, Maninder Singh²

GJCST Classification
C.2.1, B.8.2

Abstract -IEEE 802.16 standard supports two different topologies: point to multipoint (PMP) and Mesh. In this paper, a QoS mechanism for point to multipoint of IEEE 802.16 and BS scheduler for PMP Mode is proposed. This paper also describes quality of service over WiMAX networks. Average WiMAX delay, Average WiMAX load and Average WiMAX throughput at base station is analyzed and compared by applying different scheduler at Base station and at fixed nodes.

I. INTRODUCTION

IEEE 802.16 is a set of telecommunications technology standards aimed at providing wireless access over long distances in variety of ways- from point to point to full mobile cell type access. IEEE 802.16 standard is developed to serve fixed subscriber stations (SSs) through a central base station (BS) using a PMP topology. In PMP mode, every subscriber stations are directly communicate with central base station. PMP mode(in WiMAX) easily provide different type of services than wired networks at lower cost of arrangement. IEEE 802.16 is developed with QoS in mind. In PMP mode, five different service classes are introduced for different application and packets from different service classes are handled based on their QoS constraints. In this paper, QoS mechanism using WFQ queue compare with DWRR queue in PMP mode(in WiMAX).

II. PMP MODE OF IEEE 802.16

IEEE 802.16 -2004 defined in 2004, operates in 2-11 GHz as well as the original 10-66 GHz band ,provides medium data rates and supports PTP and PMP operation modes for fixed subscribers only. Only LOS and NLOS communication are supported. Where communication made possible between transmitter and receiver(s) are placed on high rise towers so as to avoid all physical obstacles between them, is called **line-of-sight(LOS)** and when LOS communication is not possible(e.g. when transmitter/receivers are devices inside a home), signals transmitted from the receiver undergo attenuation and multipath distortion (after bouncing off trees and building). This type of communication is called **non line-of-sight(NLOS)** communication. In PMP mode, Physical and Medium Access Control Layer plays important role in communication between base station and subscriber stations. WiMAX defines the concept of *service flow*. A *service flow*

is unidirectional flow of packets with a particular set of *quality of service (QoS)* parameters. A service flow is identified by a 32-bit service flow identifier (SFID). In [1], N. Srinath describes WiMAX is connection-oriented protocol. This connection-oriented scheme provides a means for handling bandwidth requests and allocation traffic and QoS parameter with service flow etc. A connection is identified by a 16-bit connection identifier (CID). MAC layer divide into three sub layers, first *service specific convergence sublayer (CS)* define interface with higher layers, converts higher layer packets into MAC level service flow and parameters. Second, *MAC Common Part sublayer (MAC CPS)* implements common MAC functionalities like link initialization, admission control, controlling channel access, transmission scheduling, quality of service, fragmentation, error control and retransmission. Third, *Security Sublayer* provides security through authentication, key management and encryption.

1) IEEE 802.16 Mac Protocol

PMP architecture, which consists of one BS managing multiple SSs. Transmissions between the BS and SSs are realized in fixed-sized frames by means of time division multiple access (TDMA) / time division duplexing (TDD) mode of operation . According to Alexey Vinel[2], the frame structure consists of downlink sub-frame for transmission from the BS to SSs and an uplink sub-frame for transmission in the reverse direction as shown in fig 1. The Tx/Rx transition gap (TTG) and Rx/Tx transition gap(RTG) shall be inserted between the sub-frames to allow terminals to turn around from reception to transmission and vice versa. In the downlink sub-frame the Downlink Map (DL-MAP) and Uplink Map (UL-MAP) message are transmitted by the BS, which comprise the bandwidth allocation for data transmission in both downlink and uplink direction, respectively.

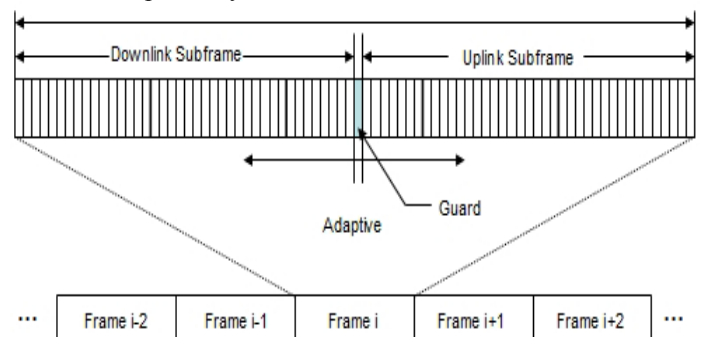


Figure1. TDMA-Frame structure

About¹- Department of Computer Science, Punjabi University, Patiala
harwinder002@gmail.com

About²- Department of Computer Science, Punjabi University, Patiala
singhmaninder25@yahoo.com

2) IEEE802.16 QoS Classes and Scheduling

IEEE 802.16 standard can support multiple communication services (data, voice and video) with different QoS requirements. The MAC layer defines QoS signaling mechanisms and functions that can control BS and SS data transmissions. On the downlink, the transmission is relatively simple, because the BS is the only one that transmit during a downlink sub-frame. Data packets broadcast to all SSs and an SS only listens in on the the packets destined for it. On the uplink, the BS determines the number of time slots for which each SS will be allowed to transmit in an uplink sub-frame. This information is broadcast by the BS through the uplink map message (UL-MAP) at the beginning of each frame. The UL-MAP contains an information element (IE) per SS, which includes the transmission opportunities for each SS, i.e., the time slots in which a SS can transmit during the uplink sub-frame. The BS uplink-scheduling module determines the IEs by using the bandwidth request message sent from the SSs to the BS.

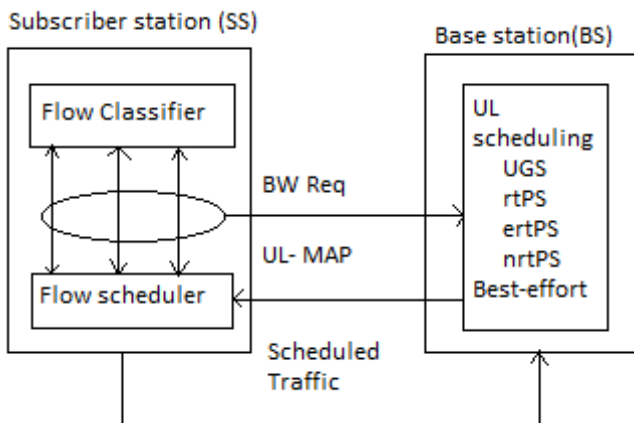


Figure 2. QoS Architecture of IEEE802.16

In the IEEE 802.16 standard, bandwidth-requests are normally transmitted in two modes according to Hemant kumar rath in [3]: a *contention mode* and *contention-free mode*(polling). In the contention mode, the SSs send bandwidth-requests during a contention periods, and contention is resolved by the BS using exponential back-off strategy. In the contention-free mode, the BS polls each SS, and an SS in reply sends its BW-request. There are five types of basic services described in the standard. Namely, Unsolicited Grant Service (UGS); Real-Time Polling Service (rtPS); Non-Real-Time Polling Service (nrtPS); Extended-Real-Time Polling Service (ertPS); Best-Effort (BE) service. Variable bandwidth assignment is possible in rtPS, nrtPS, ertPS and BE services. Whereas UGS service needs fixed and dedicated bandwidth assignment. Figure 2 shows the QoS architecture of IEEE 802.16 based services.UGS is designed for constant bit-rate (CBR) like flows such as VoIP which require constant bandwidth allocation. rtPS service is designed for variable bit-rate (VBR) flows such as MPEG video, which have specific bandwidth requirements as well as the latency. ertPS builds

on the efficiency of both UGS and rtPS and designed to support real-time service flows that generate variable-size data packets on periodic basis, such as voice over IP services with silence suppression. According to Yanqun Le,Yi Wu[4], the nrtPS and BE are for VBR non-real time applications(e.g. bandwidth intensive file transfer) and best-effort applications(e.g. HTTP), respectively.

In [5],Aun Haider and Richard j. Harris describes packets schedulers can be classified into the following two types: work conserving and nonwork conserving . Examples of work conserving scheduling algorithms include Generalized Processor Sharing (GPS), Weighted Round Robin (WRR), Deficit Weighted Round Robin (DWRR),Weighted Fair Queueing (WFQ), Self Clocked Fair Queueing (SCFQ); whereas Hierarchical Round Robin (HRR), Stop-and-Go, and Jitter-Earliest-Due-Date are some examples of nonwork conserving schedulers. In our Proposed QoS mechanism, we have used Deficit Weighted Round Robin (DWRR) and Weighted Fair Queueing (WFQ) schedulers. In DWRR is modified weight round robin scheduling discipline. It can handle packets of variable size without knowing their mean size. A maximum packet size number is subtracted from the packet length and packets that exceed that number are held back until the next visit of the scheduler. WRR serves every non empty queue whereas DWRR serves packet at head of every non-empty queue whose deficit counter is greater than the packet's size at Head of Queue (HoQ) if the deficit counter is lower, then the queue is skipped (HoQ packet is not served) and its credit increased by some given value called quantum. The increased value is used to calculate the deficit counter the next time around when the scheduler examines this queue for serving its head-of-line. If the queue is served, then the credit is determined by the size of packet being served. In [5], Aun Haider describes that DWRR is simple $O(1)$. It can be employed for scheduling at the BS of a WiMAX network. Weighted Fair Queueing (WFQ) is data packet scheduling technique allowing priorities statistically multiplexed data flows WFQ is a generalization of fair queueing (FQ). Both in WFQ and FQ, each data flow has a separate FIFO queue. In FQ, with a link data rate of R , at any given time the N active data flows (the ones with non-empty queues) are serviced simultaneously, each at an average data rate of R / N . Since each data flow has its own queue, an ill-behaved flow (who has sent larger packets or more packets per second than the others since it became active) will only punish itself and not other sessions. Contrary to FQ, WFQ allows different sessions to have different service shares. If N data flows currently are active, with weights w_1, w_2, \dots, w_N , data flow number i will achieve an average data rate of

$$\frac{Rw_i}{(w_1+w_2+w_3+\dots+w_N)}$$

Network with WFQ switches and a data flow that is leaky bucket constrained, an end-to-end delay bound can be guaranteed. By regulating the WFQ weights dynamically, WFQ can be utilized for controlling the quality of service.

PROPOSED QoS MECHANISM

The network topology of simulation scenarios is illustrated in Figure 3. There is one BS, five fixed nodes. We have applied DWRR and WFQ scheduler at each fixed node and at the base station but one can also use DWRR scheduler at the BS (Base station) for scheduling in WiMAX network and WFQ Scheduler at fixed stations for scheduling the traffic belonging to the nrtPS class. But in our QoS mechanism, we have used DWRR and WFQ scheduler for five different traffic classes like UGS, rtPS, nrtPS, Best-Effort and ertPS at one BS and five fixed SSs in network topology. First we have assigned different interface having different IP addresses to the BS and five fixed nodes also called subscriber stations (SS) then applied DWRR and WFQ scheduler at BS and SSs and used best-effort type of service (TOS) respectively.

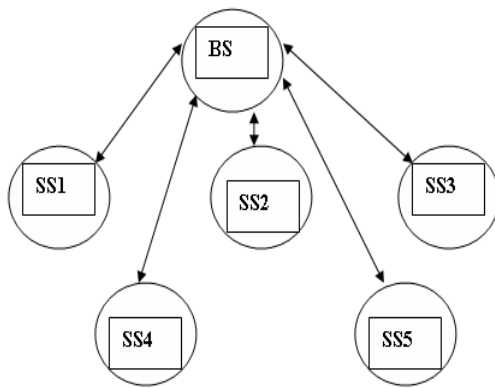


Figure 3. Network Topology

III. SIMULATION SCENARIO

In the simulation, we have used a topology that consists of one base station (BS) and five fixed node (SSs). SS1 sends

ftp traffic to SS2, SS2 sends video traffic to SS3, SS3 sends http traffic to SS4, SS4 sends VoIP with silence suppression and SS4 sends voice traffic to SS1 fixed node. We have assumed error free link conditions. Wireless OFDMA PHY layer of IEEE 802.16 standard is used with a channel bandwidth of 20MHz. The frame duration is 12.5 ms is used. ARQ and packing mechanisms are not used. Other simulation parameters are provided in Table 1.

Simulation parameter	Value
Channel Bandwidth	20MHz
Frame Duration	12.5 ms
TTG	106 ms
RTG	60 ms
Modulation scheme	64 QAM,16 QAM
Coding rate	3/4
Duplexing Technique	TDD

Table 1. Simulation Parameters

IV. SIMULATION RESULTS

To present the results of simulations we have compared Average WiMAX delay at base station (BS) and at each fixed node (SSs) using DWRR and WFQ scheduler with different type-of-service (TOS) respectively. In following simulation result figures, nnn-scenario1-DES-1 is referred to simulation run with WFQ scheduler and rrr-scenario1-DES-1 is referred to simulation run with DWRR scheduler.

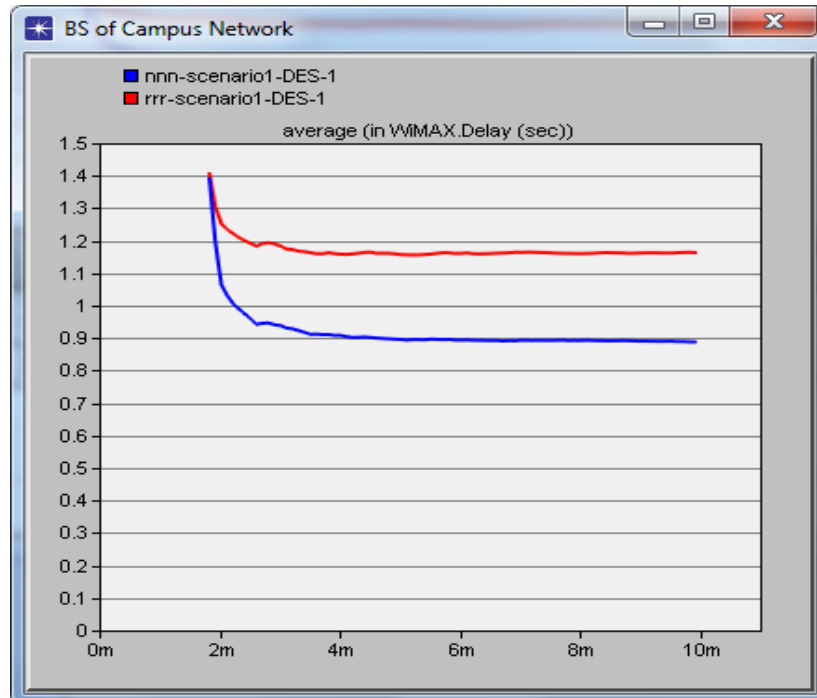


Figure 4 Average delay (in sec) in WiMAX .

Average WiMAX delay (in sec) using WFQ scheduler at base station is less as compared to average WiMAX delay using DWRR scheduler as shown in the fig 4. Average

WiMax throughput (bits/sec) using WFQ scheduler at base station is higher than average WiMAX throughput (bits/sec) using DWRR scheduler at base station, shown in fig. 5.

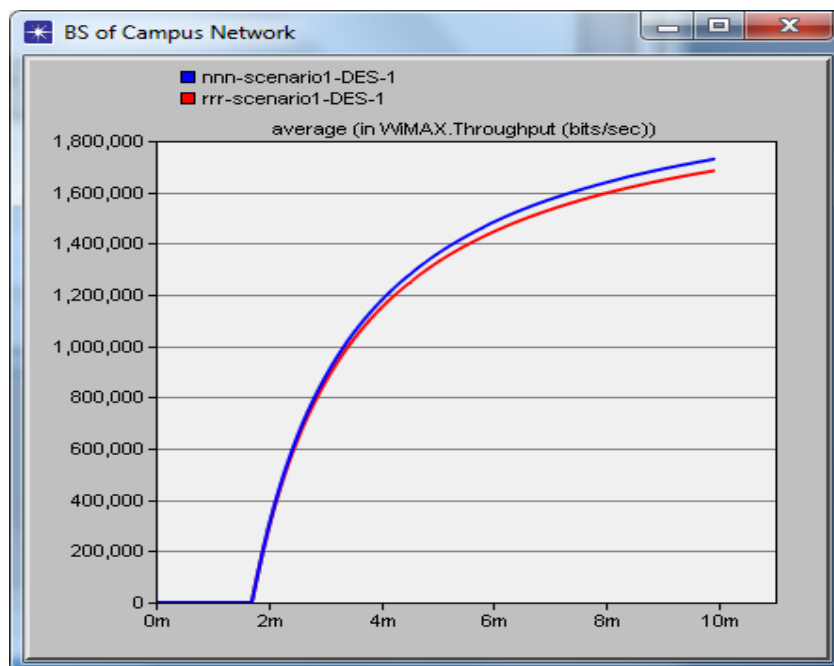


Figure 5. Average Throughput (bits/sec) in WiMAX .

But Average Load in WiMAX (bits/sec) is totally converted to Average delay in WiMAX (in sec) at base station. i.e. Average Load in WiMAX (bits/sec) at base station using WFQ scheduler is higher than Average load in WiMAX

(bits/sec) at base station using DWRR scheduler which is shown in fig 6.

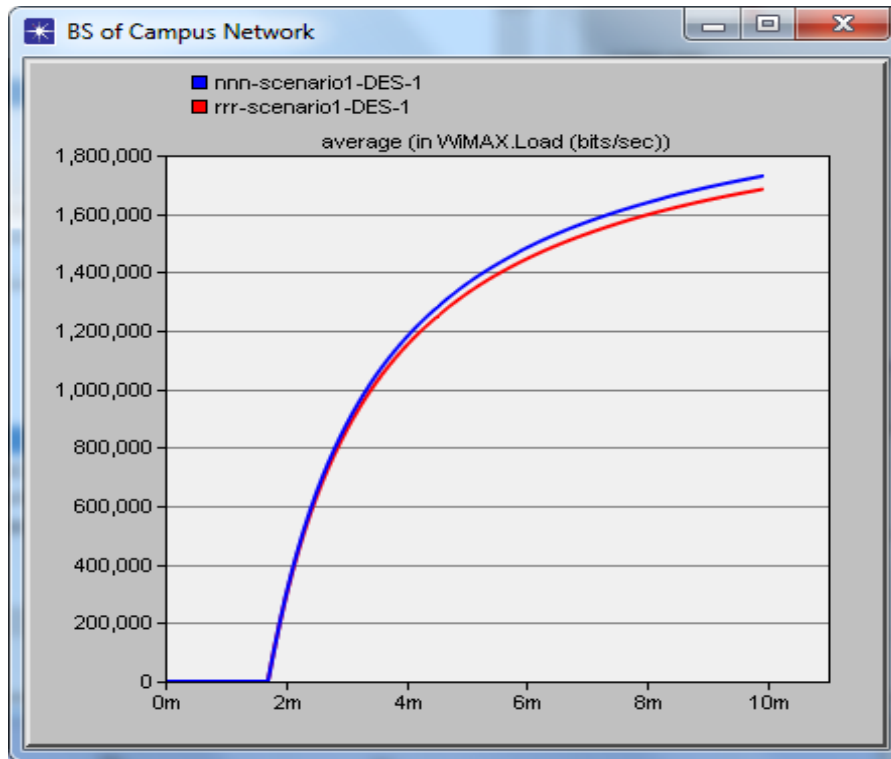


Figure 6. Average Load (in bits/sec) in WiMAX

Next, we have shown the performance of interfaces which is used at base station using DWRR and WFQ scheduler.

Traffic received/sent is higher at the base station if we are using Weighted Fair Queue (WFQ) scheduler as compared to the DWRR scheduler, as shown in fig. 7 & 8.

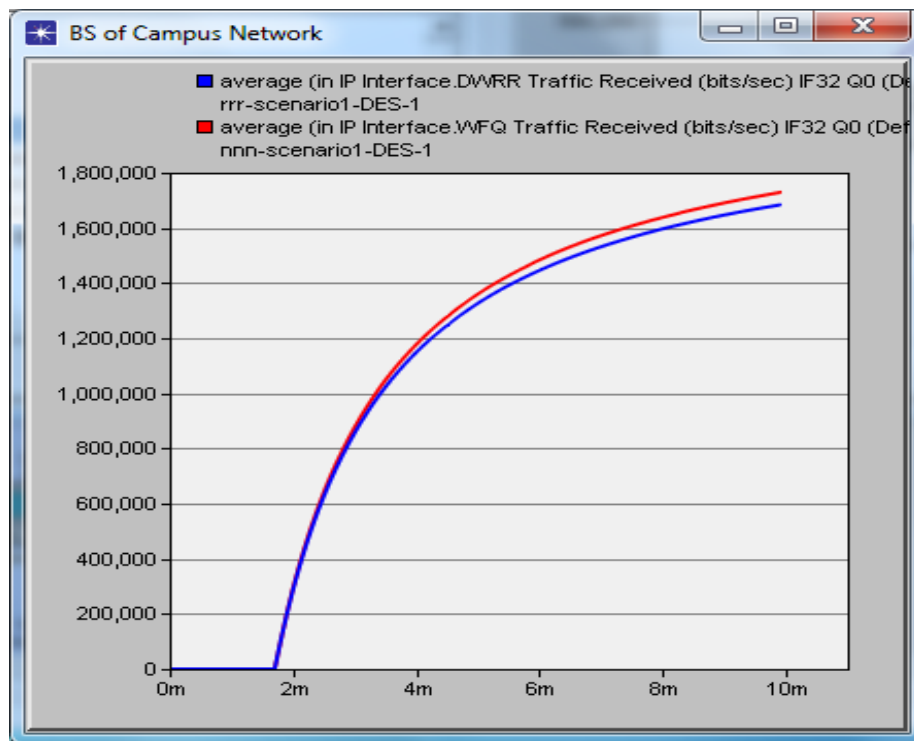


Figure 7. Traffic Received (bits/sec) through IP interface.

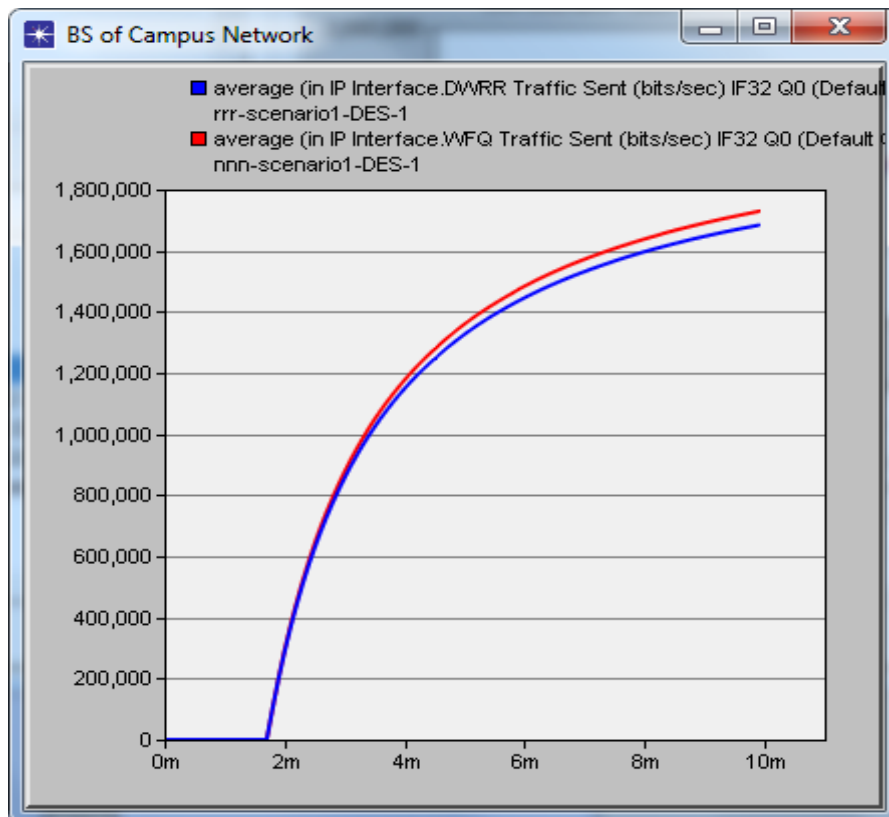


Figure 8. Traffic sent (bits/sec) through IP interface

V. CONCLUSION

In this paper, we have proposed a QoS mechanism for WiMAX delay in PMP mode of IEEE 802.16. We have used simple scheduling for the base station and fixed nodes. In which WFQ performs better than DWRR scheduler. The results of the comparison have shown that IP interface gives better output for received and sent the traffic(bits/sec) and Delay in DWRR is more compared to using WFQ. The data transfer rate of DWRR is also less than WFQ.

VI. REFERENCES

- 1) N. Srinath(CS07M035), "WiMAX-An Introduction" Department of Computer science and Engineering, Indian Institute of Technology, Madras, 2008.
- 2) Alexey Vinel, Ying Zhang, Qiang Ni, Andrey Lyakhov, "Efficient Request Mechanism Usage in IEEE 802.16", Information Systems Department, State University of Aerospace Instrumentation, Saint-Petersburg, Russia, 2006.
- 3) Hemant Kumar Rath, Abhijeet Bhorkar, Vishal Sharma, "An Opportunistic DRR(O-DRR) Uplink Scheduling Scheme For IEEE802.16-based Broadband Wireless Networks, Department of Electrical Engineering, Indian Institute of technology Bombay, Mumbai, 2006.
- 4) Yanqun Le, Yi Wu, "An Improved Scheduling Algorithm for rtPS Services in IEEE 802.16", Nokia Siemens Networks, Beijing, China, 2009.
- 5) Aun Haider, Richard j. Harris, "WiMAX Network Planning and Optimization", Taylor & Francis Group, LLC, 2008.

The Extraction of the Tooth Contour for Biometric Identification Device

{ GJCST Classification
1.4.8, 1.4.7 }

Mr. M. Moorthi¹, Dr. M. Arthanari², Mr. M. Sivakumar³

Abstract-The contours of the teeth are important features in dental biometrics. They can be utilized as cues for identifying individuals. The extraction of other features (e.g., the dental work) also needs the contours of teeth as reference. However, without the proper guidance of tooth shape knowledge, these methods are subject to errors when the tooth contours are fuzzy and only partially visible. To extract the contours from these images, we propose a method based on edge detection, and an extraction method by using binary template and knowledge based model. The method is developed using Matlab 7.5. The accuracy and robustness of this method allow its usage in biometric identification system. Matching results shows the contours extracted with this method perform better than the traditional method

Keywords-Face recognition, Facial features extraction, tooth detection

I. INTRODUCTION

In different organizations like financial services, e-commerce, telecommunication, government, traffic, health care the security issues are more and more important. It is important to verify that people are allowed to pass some points or use some resources. The security issues are arisen quickly after some crude abuses. For these reason, organizations are interested in taking automated identity authentication systems, which will improve customer satisfaction and operating efficiency. The authentication systems will also save costs and be more accurate than a human being. Basically there are three different methods for verifying identity: (i) possessions, like cards, badges, keys; (ii) knowledge, like userid, password, Personal Identification Number (PIN); (iii) biometrics like fingerprint, face, ear. Biometrics is the science of identifying or verifying the identity of a person based on physiological or behavioral characteristics. Biometrics offer much higher accuracy than the more traditional ones. Possession can be lost, forgot or replicated easily. Knowledge can be forgotten. Both possessions and knowledge can be stolen or shared with other people. In biometrics these drawbacks do exist only in small scale. We all can recognize people from faces, but we hardly can recognize anyone from tooth. This paper is organized as follows. Literature surveys are given in section 2. In section 3 we will devote ourselves to discussing the knowledge-based teeth detection method in detail.

About¹- Assistant Professor, Kongu Arts and Science College, Erode – 638 107 Email: moorthi.bmka@gmail.com

About²- Director, Bharathidasan School of Computer Applications, Ellispettai – 638 116, Email: arthanarimsvs@gmail.com

About³- Doctoral Research Scholar, Anna University, Coimbatore, Email: sivala@gmail.com

Experimental results are reported in section 4. Conclusions will be drawn in section 5.

II. LITERATURE SURVEY

The detection of facial features has been approached by many researchers and a variety of methods exist. Nevertheless, due to the complexity of the problem, robustness and preprocessing steps of these approaches are still a problem. Most commonly, natural face feature templates taken from real persons are used for a template matching algorithm [1] [2]. These templates have to satisfy a set of requirements like orientation, size, and illumination. Therefore a preprocessing step is necessary for at least aligning, and size changes. A wavelet based approach is described in [3]. Face images and face features from a database have to be aligned in orientation and size in a preprocessing step. Both previous described methods are limited by the used template and face database. In [4], an approach with synthetic templates is presented for tracking eye corners as well as mouth points within a video sequence. A training sequence is necessary to determine needed parameter values for the specific face. Deformable templates [5] also belong to the class of artificially constructed templates. This approach is using more detailed templates for eye and mouth outline matching and needs initial parameters. We are going to describe and evaluate detection algorithms for facial features like iris-pupil centers, mouth corners, and mouth inner lip line. The detection algorithms are based on templates and knowledge based method for each face feature. The accuracy and robustness of this method allow its usage in a biometric identification system

III. PROPOSED METHOD

Dental biometrics about teeth, including tooth contours, relative positions of neighboring teeth, and shapes of the dental work (e.g., crowns, fillings, and bridges). The proposed system has two main stages: feature extraction and matching. The feature extraction stage uses linear filter to enhance the images and knowledge based method to segment the dental work and store the features in the database for identification. The matching stage has three sequential steps: tooth-level matching, computation of image distances, and subject identification. In this paper we do only features extraction.

1) Face Area Detection

Before starting the search for face features the face area has to be found in which the face features are located. In [6] an

algorithm is described which is able to mark possible eye positions with high robustness and acceptable accuracy. The desired face area is defined by applying head and face relations which are described in [7] to the possible eye positions. The half distance of the possible eye positions can be interpreted as eye width or length of the eye fissure. Furthermore the eye width can be used as a unit for face feature relations.

2) Face Detection

Feature points like iris-pupil centers, eye corners, nose tip, chin outline, cheek outline, eyebrows and mouth corners are used to adjust a generic head model to an image of that person and to extract a texture map for the model. For a practical system, this feature detection has to be done automatically without any manual interaction. The following sections describing algorithms for iris-pupil centers, iris outlines, and mouth corners detection.

3) Eyes

We present a novel approach for iris-pupil centers detection within the face area. This approach is based on the knowledge of the iris and pupil. The iris is a circle which can have partly occlusions of the upper and lower lid. Mostly the upper lid is the reason for such occlusions. The iris and pupil are both dark compared to the white of the eye ball and to the luminance values of skin color. There are mostly two eyes found within a face with the same diameter for the iris. The locations of the eyes are described in relations to the other face features like mouth, and nose, but also to head weight and height. These relations are called anthropometric information or data and could be found for example in [7]. First the iris-pupil area and afterwards the iris-pupil center detection are explained.

Eye area Search. This part of the iris-pupil centers search is used to mark a position within the iris-pupil area. Key elements are a iris-pupil template and the inverse version of this template as well as the 90 degree rotated vertical versions of the described templates. The iris-pupil template is a filled circle surrounded by a box. The filled circle represents the iris and pupil as one part. The horizontal version of the template has left and right areas, which represent parts of eyeball area. The vertical version has upper and lower areas, which represent parts of upper and lower eye lid. The purpose of these templates is to extract desired values of the luminance, and first order gradient data, but also for calculating the correlation of the shape after extracting the data. Such templates have only the diameter of the filled circle (iris-pupil area) as a free parameter which is adjusted at the beginning of the search within the face area. The diameter is determined from the eye width as unit for facial feature relations. The eye width to eye height relation described in [7] can be expressed as 1: 1/3 of the determined unit eye width. The eye height is interpreted as iris diameter and is therefore used to create the iris-pupil templates. The width of the left and right eyeball areas as well as the upper and lower lid areas are adjusted to the iris diameter. Because of the dependency of

the possible eye position found by [6] and anthropometric relations described by [7] the desired templates are generated according to the found size of the face area within the scene. Therefore these templates can be scaled automatically depend on the size of the face area. The algorithm consists of three steps: First, a search using the templates, second combination of both search results. Third step is applying anthropometric information to the combined search results.

Step one is done for each pixel (horizontal, vertical) within the face area. Cost functions used together with the corresponding templates consist of the following elements:

- The luminance values of the iris-pupil filled circle and the surrounding areas are estimated for each search position in order to minimize the difference of template and search image,
- correlation of synthesized iris-pupil template parts with the search image,
- horizontal gradient values along the horizontal iris-pupil template outline and vertical gradient values along the vertical iris-pupil template,
- luminance value variation inside the iris-pupil area,
- average of luminance values inside the pupil-iris area, and
- lowest luminance value inside the iris-pupil area.

Step one generates two 2D error images. Step two starts with excluding regions of the error images found only by one search (horizontal or vertical template). Both error images are combined by simple addition after normalization. Prospective iris-pupil centers are extracted by analyzing the combined result for high density regions. The third step arranges these regions using anthropometric information. Prospective iris-pupil area combinations for left and right iris-pupil centers are the output of this part.

Eye area Center Search. In order to refine the iris-pupil center positions and to obtain the iris outline in a second step, deformable iris outline templates are introduced. These templates can be changed in diameter and geometry. Taken into account that in most cases the iris is partly covered with the upper lid makes it necessary to change the geometry of the iris outline template. The iris outline template is not a filled area like the iris-pupil template because of the possible light reflection(s) within the iris-pupil area. A cost function which is defined by property descriptions of the iris is used with the described iris outline template. The algorithm consists of two steps. First, the iris outline template is only changed in iris diameter and position. The iris diameter can be changed in both directions; this means a bigger or smaller iris is possible. The position found by the part one of the iris-pupil center search can be shifted within the initial iris diameter. Second, the geometry changes are applied using the previous determined iris diameter. The used cost function consists of the following elements:

- Absolute value of the horizontal and vertical gradient values along the specified iris outline,
- size of the iris outline (circumference), and
- the average of the luminance values inside the iris outline.

4) Mouth

Before extracting specific mouth features like mouth corners and mouth inner lip line, the mouth self has to be located within the face area. For this purpose eye location and a mouth template as well as the inverse version similar to the iris-pupil template are used. Note this template has the same size as the vertical version of the iris-pupil template, because of the given anthropometric information in [7]. Therefore this template does not represent a whole mouth. The mouth templates are used to determine possible mouth locations. Analyzing the possible regions with the anthropometric location of the mouth marks the position of a mouth region. Within the found mouth region highest second order gradient values of the luminance values are connected to the initial mouth inner lip line. The mouth corners and inner lip line search region is framed by applying vertical the upper and lower lip height along the initial line, and by extending the left and right search region boundaries to the maximum possible distance given by the biocular line. The lowest luminance values are connected in this procedure to the possible mouth inner lip line. In order to specify the mouth corners a color plane in the RGB color space is used to separate mouth from skin color pixels of the mouth line. Samples of the mouth and the skin color have to be extracted from the color image for positioning of the separation plane. Skin color pixels are found by excluding regions with high gradient values of the face area. Morphological operations like dilatation are applied as well. Mouth color pixels are extracted from the inner lip line. The positioning of the separation plane consists of the following steps:

- Separation vector centered by mean values of skin and mouth pixels,
- normal on the separation vector towards mean skin position, and
- rotation around the separation vector for maximum skin color data enclosure.

In order to separate skin from mouth pixels the normal of the plane is used for color space transformation. Positive results are pointing to the skin color space and negative results to the mouth color space. The usage of this method for separating skin and mouth color pixels along the found inner lip line marks the mouth corners.

5) Tooth Extraction

A tooth includes two parts: the upper part is called the crown, which is above the gum line, and the lower part is called the root, which sits in the bone below the gum. Because of the bone and the soft tissue, the root of the tooth is not very visible in the image. Since the crown has a higher contrast with the background than the root, we extract the shape of the crown only.



Fig 1: tooth

To simplify the shape extraction, we manually select a rectangular region R and a point c in the input image, I . R is a region containing the tooth and the point c is inside the crown of the tooth, which is called the Crown Center. First, we compute the gradient image, $|\Delta I|$, of the input image, I , as

$$|\Delta I(x, y)| = \text{Sqrt}((I(x, y) - I(x, y-1))^2 + (I(x, y) - I(x-1, y))^2) \quad (1)$$

Due to the proximity of the neighboring teeth, there is some interference between the edges of a tooth and the edges of adjacent teeth. To solve this problem, we define an auxiliary image, M , by

$$M(x, y) = B(x, y) |\Delta I(x, y)| \quad (2)$$

Where

$$B(x, y) = \begin{cases} 0 & \text{if } \Delta I(x, y) \cdot E(x, y) < 0 \\ 1 & \text{otherwise,} \end{cases}$$

and $\Delta I(x, y)$ is the gradient direction vector at (x, y) , $E(x, y)$ is a vector from Crown Center c to (x, y) , and \cdot denotes the dot product. Note that when $E(x, y)$ and $\Delta I(x, y)$ form an acute angle, $B(x, y)$ equals 1; otherwise, $B(x, y)$ equals 0. As a result, the effect of the edges of the neighboring teeth in computing the gradient is greatly decreased. From the crown center, we make a radial scan of the crown. The edge of the crown should be somewhere along this radial line. We sort all the points on the radial line in terms of $M(x, y)$, and record the three largest ones. If we draw a sufficiently large number of radial lines, then a list of edge candidates is produced. Let M be the mean of all the recorded $M(x, y)$ values. We define the reliability, $R(x, y)$, of each edge candidate (x, y) by

$$R(x, y) = \begin{cases} e^{-\alpha(M(x, y) - M)^2} & \text{if } M(x, y) < M, \\ 1 & \text{if } M(x, y) \geq M, \end{cases} \quad (3)$$

Where α is a constant to prevent $R(x, y)$ from decaying too fast. Along each radial line, the point with the largest reliability is selected to be an edge point.

IV. EXPERIMENTAL RESULTS

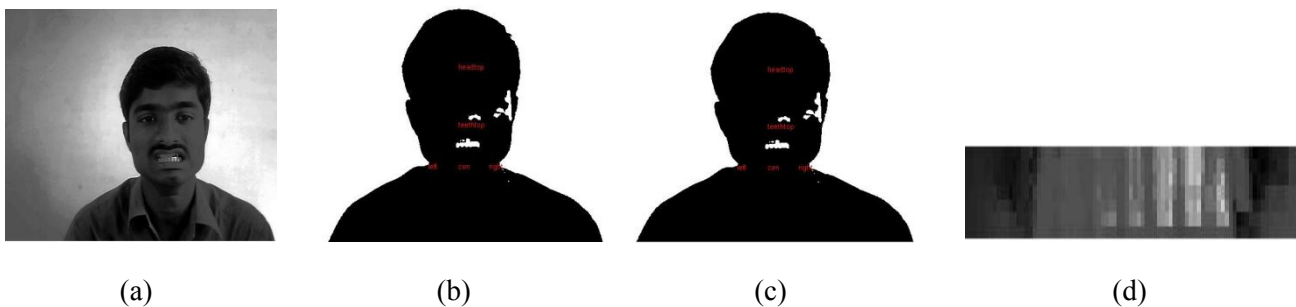
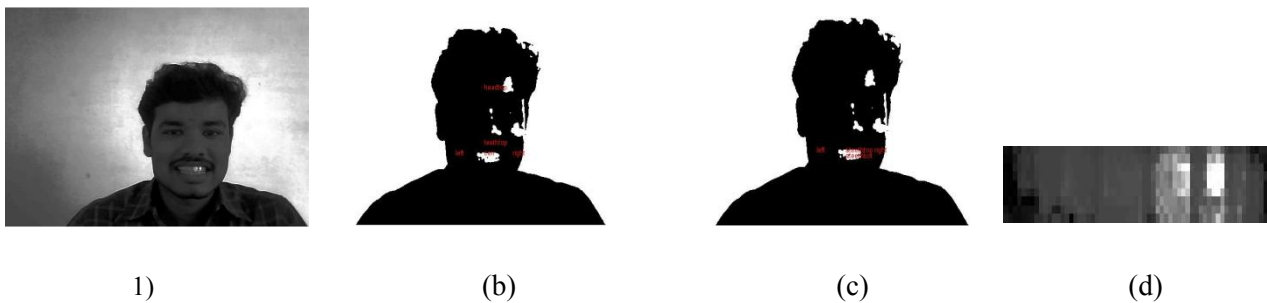
Computers of the future will interact with us more like humans. The key element of that interaction will be their ability to recognize human beings and even understand their expressions. Proposed system serves an integral part in establishing smart environments, where computers are employed everywhere and it is suitable for any application where people require access to a technical system: computer networks, Internet commerce, banking systems, and ATMs.

In addition, this system secures access to rooms and buildings. Depending on application, the proposed system authorizes people either through identification or verification. In identification mode, the system identifies a person exclusively through biometric traits. In verification

mode a person gives his name or a number, which the system then verifies by means of biometric traits. The correct extraction rate is given in Fig 2. and the results shows in Fig 3 and Fig 4.

Algorithm	Test Scenario	Sample Number	Correct extraction	%	Mean false extraction
DSM and GVF	Day	125	120	96	5%
	Night	125	118	94.4	
Proposed Method	Day	125	123	98.4	2%
	Night	125	122	97.6	

Fig 2: Right teeth extraction



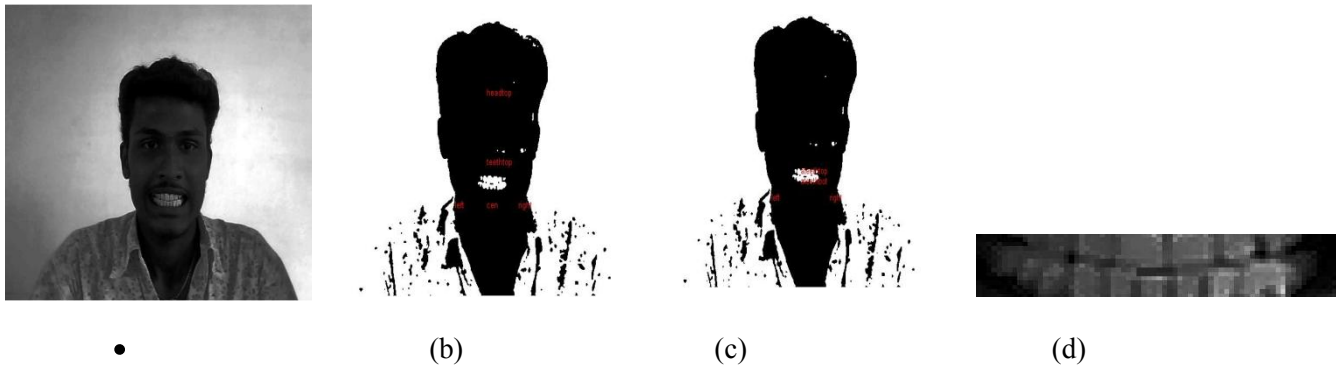


Fig. 3 (a) captured image (b) binary image (c) feature identification (d) teeth

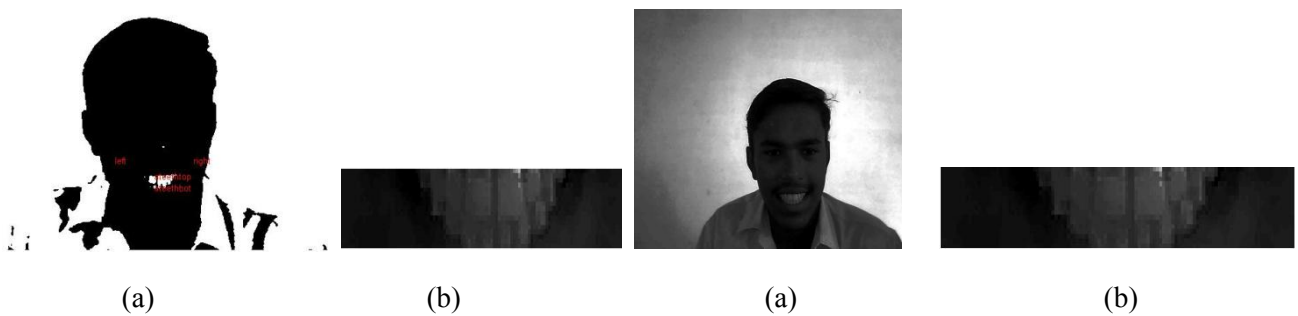


Fig. 4 (a) captured image (b) teeth

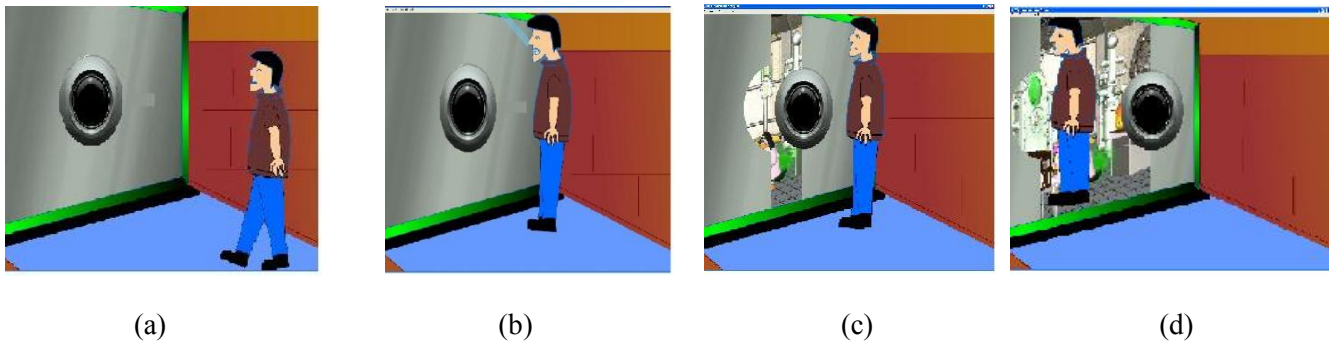


Fig. 5 Teeth authentication (a) person walking (b) show teeth (c) teeth identified (d) door opened

In a test involving 125 persons for three months, the proposed system reduced the false-acceptance rate significantly below 2 percent, depending on the security level. The higher the security level, the higher the false-rejection rate. Thus system administrators must find an acceptable false rejection rate without letting the false-acceptance rate increase too much. The security level depends on the purpose of the biometric system.

V. CONCLUSION

As the fraud in our society grows, the pressure to deliver more and more authentication services also grows. The

proposed system guarantees a high degree of security from falsification and unauthorized access. It also protects the privacy rights of system users. To catapult biometric technology into the main-stream identification market, it is important to encourage its evaluation in realistic contexts, to facilitate its integration into end-to-end solutions, and to foster innovations of inexpensive and user-friendly implementations. We hope that a pervasive, accountable use of biometrics technology will help to establish a more open and fair society.

VI. REFERENCES

- 1) Kampmann, M.: Automatic 3-D Face Model Adaptation for Model-Based Coding of Videophone Sequences. IEEE Transaction on Circuits and Systems for Video Technology 12 (2002)
- 2) Kampmann, M., Ostermann, J.: Automatic adaptation of a face model in a layered coder with an object-based analysis-synthesis layer and a knowledge-based layer. Signal Processing: Image Communication 9 (1997) 201-220
- 3) Feris, R.S., Gemmell, J., Toyama, K., Krueger, V. Hierarchical wavelet networks for facial feature localization. In: Proceedings of the 5th International Conference on Automatic Face and Gesture Recognition, Washington D.C., USA (2002)
- 4) Malciu, M., Preteux, F.: Tracking facial features in video sequences using a deformable model-based approach. In: Proceedings of the SPIE. Volume 4121. (2000) 51-62
- 5) Yuille, A.L.: Deformable Templates for Face Recognition. Journal of Cognitive Neuroscience 3 (1991) 59-79
- 6) Froba, B., Ernst, A., Kublbeck, C.: Real-time face detection. In: Proceedings of the 4th IASTED International Conference on Signal and Image Processing (SIP 2002), Kauai (2002) 479-502
- 7) Farkas, L.G.: Anthropometry of the Head and Face. 2nd edn. Raven Press (1995)

A Proposed Novel Framework for Early Effort Estimation using Fuzzy Logic Techniques

Roheet Bhatnagar¹, Vandana Bhattacharjee², Mrinal Kanti Ghose³

GJCST Classification
1.5.1

Abstract-Estimating software development effort is an important task in the management of software projects. The task of effort estimation is challenging and is an important area of research in the field of Software Project Management. A number of estimation models exist for effort prediction. However, many newer models are still being proposed and researched upon to obtain more accurate estimations. The development of software has always been characterized by parameters that possess certain level of fuzziness. This requires that some degree of uncertainty be introduced in the models, in order to make the models realistic. Fuzzy logic techniques come to play in such a situation to tackle the uncertainty issues. Besides, fuzzy logic had been combined with algorithmic, non-algorithmic effort estimation models as well as a combination of them to deal with the inherent uncertainty issues. The primary purpose of this paper is to review the studies carried over the years for estimating the software development effort using Fuzzy Logic Techniques in order to improve prediction accuracies. In this paper we have also tried to propose a novel fuzzy logic based framework to estimate the effort at the early stages of software project development. The paper discusses the theoretical foundations and concepts for the framework and in subsequent publications we will be discussing the results obtained from our ongoing experimental work.

Keywords- software development effort, effort estimation, fuzzy logic techniques, estimation models, effort estimation framework.

I. INTRODUCTION

Estimating software development effort remains a complex problem that attracts considerable research attention. Improving the estimation techniques that are currently available to project managers would facilitate increased control of time and money in software development. Furthermore, any improvement in the accuracy of predicting the development effort can significantly reduce the costs from errors, such as estimating inaccurately, misleading tendering bids, and disabling the monitoring progress [1]. Software development effort estimates are the basis for project bidding and planning (both are critical practices in the software industry). The consequences of poor budgets and plans can be dramatic: if they are too pessimistic, business opportunities can be lost, while over optimism may be followed by

significant losses [2]. Software effort estimation has even been identified as one of the three great challenges in Computer Science [3]. During the development process, the cost and time estimates are useful for the initial rough validation and monitoring of the project's after completion, these estimates may be useful for project productivity assessment for example. Software effort estimation models are divided into two main categories: algorithmic and non-algorithmic. The most popular algorithmic estimation models include Boehm's COCOMO [4], Putnam's SLIM [5] and Albrecht's Function Point [6]. These models require as inputs, accurate estimate of certain attributes such as line of code (LOC), complexity and so on which are difficult to obtain during the early stage of a software development project. The models also have difficulty in modeling the inherent complex relationships between the contributing factors, are unable to handle categorical data as well as lack of reasoning capabilities [7]. The limitations of algorithmic models led to the exploration of the non-algorithmic techniques which are soft computing based. New paradigms offer alternatives to estimate the software development effort, in particular the Computational Intelligence (CI) that exploits mechanisms of interaction between humans and processes domain knowledge with the intention of building intelligent systems (IS) [8]. Amongst IS, fuzzy logic may be a convenient tool for software development effort estimation. Fuzzy logic-based cost estimation models are more appropriate when vague and imprecise information is to be accounted for. Fuzzy systems try to behave just like the processes of the brain with a rule base. The basic concept is inspired by the human processes where the decisional criteria are not clear cut, but blurred and it is difficult to find objective to make the decisions more precise and clear. Use of fuzzy sets in logical expression is known as fuzzy logic. Fuzzy decision systems are based on fuzzy logic that tries to reproduce the fuzzy human reasoning. Fuzzy logic is extended to take care of the partially truth values. Fuzzy logic as the name describes, the mode of reasoning to be approximate rather than exact. In fuzzy approach there are four stages: Fuzzification, Extending complexity matrix, determining the productivity rate and the effort, Defuzzification.

Why Fuzzy Logic?

Fuzzy logic effort prediction brings numerous benefits. Undeniably the development of software is characterized by parameters that possess certain level of fuzziness. This requires that some degree of uncertainty be introduced in the models, in order to make the models realistic.

About¹- Associate Professor in the Department of Computer Science & Engineering at Sikkim Manipal Institute of Technology, Sikkim, India.

(telephone: +91 9475078498 email: roheetbhatnagar@yahoo.com) About²- Associate Professor in the Department of Computer Science & Engineering at BITEC, Lalpur, Birla Institute of Technology, Mesra, Ranchi, India. (email: vbhattacharya@bitmesra.ac.in)

About³ - Professor & Head, Department of Computer Science & Engineering at Sikkim Manipal Institute of Technology, Sikkim, India. (email: mkghose2000@yahoo.com)

Moore and Edwards discovered that the accepted level of estimation accuracy was typically $\pm 20\%$ [9]. This shows that software project managers are aware of the irrefutable uncertainty in estimation and therefore expect some inaccuracy in the prediction output. Besides, the use of formal models did not improve the accuracy of estimation [10]. Fuzzy logic enables linguistic representation of the input and output of a model to tolerate imprecision. It is particularly suitable for software estimation as many software attributes are measured on nominal or ordinal scale type which is a particular case of linguistic values [11]. The use of fuzzy set satisfies the first criterion of soft computing which is the tolerance of imprecision, as defined by Zadeh [12]. More importantly, the model should support the fact that human reasons in fuzziness. The use of fuzzy set supports continuous belongingness (membership) of elements to a given concept (such as small software project) thus alleviating a dichotomy problem (yes/no) [13] that caused similar projects having different estimated efforts. Apart from that, fuzzy logic approach is less dependent on historical data [14] [15] [16]. Fuzzy logic models can be constructed without any data or with little data [14][15]. This makes fuzzy logic superior over data-driven model building approaches such as neural network, regression and case-based reasoning. In addition, fuzzy logic models can adapt to new environment when data become available [7]. Another advantage of fuzzy logic model is that it has the

ability to represent different levels of uncertainty for the inputs and outputs whilst inferring based on the same model (rules and membership functions). Consequently it is able to cater for the needs of different level of precision for the different stages of development life cycle [14][15].

II. FUZZY LOGIC IN SOFTWARE EFFORT ESTIMATION

A fuzzy model is a modeling construct featuring two main properties [17]: (1) It operates at a level of linguistic terms (fuzzy sets), and (2) it represents and processes uncertainty. Fuzzy logic offers a particularly convenient way to generate a keen mapping between input and output spaces thanks to the natural expression of fuzzy rules. In software development effort estimation, two considerations justify the decision of implementing a fuzzy model: first, it is impossible to develop a precise mathematical model of the domain [18]; second, metrics only produce estimations of the real complexity. Thus, according to the previous assertions, formulating a tiny set of natural rules describing underlying interactions between the software metrics and the effort estimation could effortlessly reveal their intrinsic and wider correlations. In Table 1 below we present a tabular view of a group of papers regarding aspects related to previous research on software development effort estimation based on Fuzzy Logic techniques and concepts.

Table I
Research On Software Development Effort Estimation Based On Fuzzy Logic Techniques And Concepts

Authors	Year	Related Work Done	Results Reported
Fei and Lui [19]	1992	Introduced the f-COCOMO model which applied Fuzzy Logic to the COCOMO model for software effort estimation.	Since there was no comparison of results between the f-COCOMO and other effort estimation models in their study the estimation capability of their model is unknown.
Kumar et al [20]	1994	Had applied fuzzy logic in Putnam's manpower buildup index (MBI) estimation model. MBI selection process was based upon 64 different fuzzy associative memory (FAM) rules.	The work showed how fuzzy FAM's can be effectively applied to the domain of software project management and control for the estimation of the MBI.
Gray and MacDonell [21] [15]	1997	Compared Function Point Analysis, Regression techniques, feedforward neural network and fuzzy logic in software effort estimation.	Their results showed that fuzzy logic model achieved good performance, being outperformed in terms of accuracy only by neural network model with considerably more input variables.
	1999	Developed FULSOME (Fuzzy Logic for Software Metrics) which is a set of tools that helps in creating fuzzy model.	The automatically generated fuzzy model performs acceptably when compared to regression-based models.
Ryder [22]	1998	Researched on the application of fuzzy logic to COCOMO and Function Points models.	Result showed Fuzzy Logic is good at making effort estimations.
Musflek et al [13]	2000	Worked on fuzzifying basic COCOMO model without considering the adjustment factor. In their simple f-COCOMO model, the size input into the COCOMO model is represented by a fuzzy set, while a and b coefficients are crisp values. Besides the size, augmented f-COCOMO also fuzzified both the coefficients	They concluded that (a) fuzzy sets help articulate the estimates and their essence (by exploiting fuzzy numbers described by asymmetric membership functions) and (b) they generate a feedback as to the given uncertainty (granularity) of the results.

		related to the development mode. Triangular membership functions are used in this study.	
Idri et al [23] [11]	2000	Proposed fuzzy intermediate COCOMO'81. The FLM is based upon trapezoidal membership functions. The dataset is randomly generated and compared with actual data of COCOMO 81. The effort multiplier for each cost driver is obtained from fuzzy set, enabling its gradual transition from one interval to a contiguous interval (such as from high to very high).	Validation results showed that the fuzzy intermediate COCOMO'81 can tolerate imprecision in its input (cost drivers) and generate more gradual outputs. Thus fuzzy intermediate COCOMO'81 is less sensitive to the changes in the inputs as compared to intermediate COCOMO'81.
	2002	Proposed an approach based on fuzzy logic named Fuzzy Analogy. Its dataset is that of COCOMO 81.	Taking into account their results, they suggested the following ranking of the four techniques in terms of accuracy and adequacy to deal with linguistic values: 1. Fuzzy Logic, 2. Fuzzy intermediate COCOMO'81, 3. Classical intermediate COCOMO'81, and 4. Classical Analogy.
Huang et al. [24]	2003	Proposed a model combining fuzzy logic and neural networks. The dataset was obtained from the original COCOMO (1981).	The results of the fuzzy logic model were better than those of the COCOMO equations. The FLM was based upon triangular membership functions. The main benefit of this model is its good interpretability by using the fuzzy rules.
Saliu et al. [7]	2004	They fuzzyfied the two different portions of the intermediate COCOMO model i.e. nominal effort estimation and the adjustment factor. They proposed a fuzzy logic framework for effort prediction by integrating the fuzzified nominal effort and the fuzzified effort multipliers of the intermediate COCOMO model.	This approach is able to deal with uncertainty, provides transparency on prediction rationale through rules, incorporate experts knowledge in the definition of membership functions and rules, as well as adaptable to new data by changing the parameters of membership functions.
Ahmed et al. [25]	2004	Presented a FLM based upon triangular membership functions. The dataset for validating the FLM was (a) generated randomly and (b) that of COCOMO 81 was used.	Results showed that the FLM was slightly better than COCOMO equations. In addition, they reported promising experimental summary results in spite of the little background knowledge of the rule base and training data.
Crespo et al. [26]	2004	Explored fuzzy regression techniques based upon fuzzification of input values. Project database of COCOMO-81 are used.	fuzzy regression is able to obtain estimation models with similar predictive properties than existing basic estimation models.
Braz et al. [27]	2004	Applied Fuzzy Logic for effort estimation of object-oriented software. FUSP (Fuzzy use case size points) metric allows gradual classifications of use case size points in the effort estimation by using fuzzy numbers.	Results showed that FUSP fares better than USP.
Xu and Khoshgoftaar [28]	2004	Presented a fuzzy identification cost estimation modeling technique to deal with linguistic data, and automatically generate fuzzy membership functions and rules. A case of study based on the COCOMO'81 database compared the proposed model with all three COCOMO'81 models (basic, intermediate and detailed).	It was observed that the fuzzy identification model provided significantly better cost estimations than the three COCOMO'81 models.
Cuauhtemoc et	2006	Carried out a study to compare personal Fuzzy	Results show that a FLS can be used as

al. [29]		Logic Systems (FLS) with linear regression using evaluation criteria which is based upon ANOVA of MRE and MER, as well as MMRE, MMER and pred(25)	an alternative for estimating the development effort at personal level.
Moon Ting Su et al. [30]	2007	Proposed an enhanced fuzzy logic model for the estimation of software development effort. The model <i>Fuzzy Logic Model for Software Development Effort and Cost Estimation</i> (FLECE) possesses similar capabilities as the previous fuzzy logic model. In addition to that, the enhancements done in FLECE improved the empirical accuracy of the previous model in terms of MMRE (Mean Magnitude of Relative Error) and threshold-oriented prediction measure or prediction quality (pred).	The analysis of the results shows that FLECE is able to obtain more accurate results in the estimation of software development effort when compared to the previous fuzzy logic model. Hence, the enhancements to FLECE are truly useful and had given better performance to the model.
Iman Attarzadeh and Siew Hock Ow [31]	2009	Proposed an enhanced Fuzzy Logic approach for the estimation of software development effort.	Results showed that the value of MMRE applying their Fuzzy Logic model was substantially lower than MMRE values as calculated by applying other Fuzzy Logic models.

III. PROPOSED FRAMEWORK FOR EARLY STAGE EFFORT ESTIMATION USING FUZZY LOGIC

Our proposed Early Stage Software Development Effort Estimation (ESSDEE) Fuzzy Logic Model (as shown in Figure 1 below) imbibes all the benefits of Fuzzy Logic applications in developing estimation models as discussed in the previous section. This model is a novel proposal and the work to validate the model based on project data of (B.Tech. Final Year) students is under progress and it is planned to establish the model with input data from industry sources as well at a later stage. In this model we have considered Entity Relationship Diagrams, Use-Case Diagrams and Class Diagrams as the conceptual modeling tool/technique used by students to capture the project requirements.

Students were monitored during their major projects and timesheet was maintained which gave us their actual development time (Effort in months/hours). As the prediction accuracy of effort estimation at the early stage of development is difficult due to the uncertainties involved we will consider the linguistic values for rule based inference component of our model. The model consists of three important components namely, Fuzzification block, Rule based inference and Defuzzification block. The inputs to the fuzzification block will be from various conceptual modeling parameters taken from the student's major project data. Linguistic values are supplied to the Rule Based Inference component and we can obtain a Fuzzy Effort as another linguistic value or a Crisp Effort after Defuzzification.

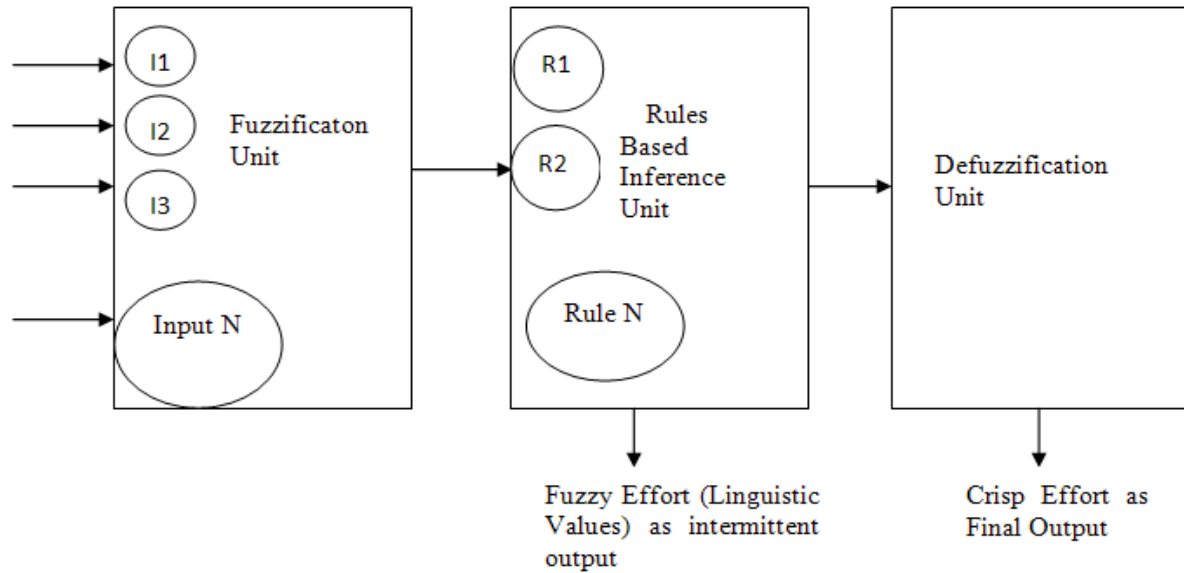


Figure 1: Proposed Fuzzy Logic based Framework to estimate the Effort of Software Project development

Following Table 2 describes the Data Capture Format from student's data.

Table 2
Student Data Collection Format

Conceptual Modeling Technique		
ER Diagrams	Use Case Diagrams	Class Diagrams
Total Number of Entity Sets	Total Number of Use-Cases	Total number of Classes
Total Number of Attributes	Total Number of Actors	Total count of attributes of a class
Relationship Types and their total count		
1:1	Type and Number of Communications <<extend>> <<include>> relations	Class Relationships
1:M		
M:1		
M:M		
Number of Weak Entity sets	Project Duration in Hours	Project Duration in Hours
	Project Category	Project Category
	CGPA of student	CGPA of student
Project Category		
CGPA of student		
Project Duration in Hours		

Based on the above data collection format student's data is collected and linguistic values (*High, Medium, Low*) will be established for each of the parameters for different conceptual modeling techniques. The proposed model is

diagrammatically shown in Figure 1 above which will make use of the above data from Table 2 for making out the effort

estimation in project development at the early stages of development.

IV. CONCLUSION

In this paper we have presented a review on the Fuzzy Logic applications in Software development effort estimation models development. We also discussed the various advantages of Fuzzy Logic for developing prediction models. Finally we propose a novel method for carrying out

the effort estimation at the early stages of software development namely the Requirements Gathering/Analysis and Design phases of Software Development Life Cycle. We plan to validate our model based on the data collected from student's project data over a span of past 3 years and will fine tune our model with large input sets. Finally we intend to establish the model based on the actual data from the software development industry.

V. REFERENCES

- 1) Luiz Fernando Capretz and Venus Marza, Improving Effort Estimation by Voting Software Estimation Models Hindawi Publishing Corporation Advances in Software Engineering Volume 2009
- 2) Grimstad, S., Jorgensen, M., Molokken-Ostfold, K., Software Effort Estimation Terminology: The Tower of Babel. Information and Software Technology. Elsevier, 2005.
- 3) Brooks Jr., F.P., Three great challenges for half-century-old computer science. Journal of the ACM, 2003.
- 4) B. W. Boehm, Software Engineering Economics, Englewood Cliffs, NJ, Prentice Hall, 1981.
- 5) L. H. Putnam, A general empirical solution to the macrosoftware sizing and estimating problem". IEEE Transactions on Software Engineering, SE-4(4), 1978, pp 345-361.
- 6) Attar Software. Fuzzy Logic in Knowledge Builder", White Paper, <http://www.intellicrafters.com/fuzzy.htm>, 2002.
- 7) M.O. Saliu, M. Ahmed and J. AlGhamdi. Towards adaptive soft computing based software effort prediction" Fuzzy Information, 2004. Processing NAFIPS '04. IEEE Annual Meeting of the North American Fuzzy Information Processing Society, 27-30 June 2004, Vol.1, pp. 16-21.
- 8) Witold Pedrycz. Computational Intelligence as an Emerging paradigm of Software Engineering. SEKE 2002, Ischia, Italy
- 9) T.T. Moores, and J.S. Edwards, Could Large UK Corporations and Computing Companies Use Software Cost Estimating Tools? - A Survey". European Journal of Information Systems, Vol.1, No.5, 1992, pp.311-319.
- 10) F.J. Heemstra, R.J. Kusters. "Controlling Software Development Costs: A Field Study". International Conference on Organisation and Information Systems, Bled, Yugoslavia, 1989.
- 11) A. Idri, A. Abran,, T.M. Khoshgoftaar. Estimating software project effort by analogy based on linguistic values" in.Proceedings of the Eighth IEEE Symposium on Software Metrics, 4-7 June 2002, pp. 21 – 30.
- 12) L.A. Zadeh, Fuzzy Logic, Neural Networks, and Soft Computing", Communication of ACM, Vol. 37, No.3, March 1994, pp.77-84.
- 13) P. Musflek, W. Pedrycz, G. Succi, M. Reformat, "Software Cost Estimation with Fuzzy Models", Applied Computing Review, Vol. 8, No.2, 2000, pp. 24-29.
- 14) A. R. Gray, S. G. MacDonell, Fuzzy Logic for Software Metric Models throughout the Development Life-Cycle" in Proceedings of the 18th International Conference of the North American Fuzzy Information Processing Society – NAFIPS, 10 – 12 June 1999, New York, USA, pp. 258 – 262.
- 15) S. G. MacDonell, A. R. Gray, J. M. Calvert. FULSOME: Fuzzy Logic for Software Metric Practitioners and Researchers" in Proceedings of the 6th International Conference on Neural Information Processing ICONIP'99, ANZIIS'99, ANNES'99 and ACNN'99, Perth, Western Australia, IEEE Computer Society Press, 1999, pp.308-313.
- 16) J. Ryder, Fuzzy Modeling of Software Effort Prediction" in Proceeding. of IEEE Information Technology Conference, Syracuse, NY, 1-3 Sept 1998, pp: 53-56.
- 17) Pedrycz, W., Gomide, F., 1998. An Introduction to Fuzzy Sets. The MIT Press.
- 18) Lewis, J.P., 2001. Large limits to software estimation. ACM Software Engineering Notes 26 (4), 54–59.
- 19) [Fei Z, Liu X, f-COCOMO Fuzzy Constructive Cost Model in Software Engineering in IEEE international conference on Fuzzy systems, 1992, pp. 331-337.
- 20) S. Kumar, B.A. Krishna, and P.S. Satsangi, "Fuzzy systems and neural networks in software engineering project management". Journal of Applied Intelligence, Vol. 4, 1994, pp. 31-52.
- 21) A. R. Gray, S. G. MacDonell, Applications of Fuzzy Logic to Software Metric Models for Development Effort Estimation". Fuzzy Information Processing Society 1997 NAFIPS' 97, Annual Meeting of the North American, 21 – 24 September 1997, pp. 394 – 399.
- 22) J. Ryder, Fuzzy Modeling of Software Effort Prediction" in Proceeding. of IEEE Information Technology Conference, Syracuse, NY, 1-3 Sept 1998, pp: 53-56.
- 23) A. Idri, A. Abran, L. Kjiri. „COCOMO cost model using Fuzzy Logic", 7th International Conference on Fuzzy Theory & Techniques, 27 Feb – 3 March 2000, Atlantic City, New Jersey.
- 24) Huang, X., Capretz, L.F., Ren, J., Ho, D.A., Neuro-Fuzzy Model for Software Cost Estimation. In: Proceedings of the Third International Conference on Quality Software, 2003.
- 25) Ahmed, M.A., Saliu, M.O., AlGhamdi, J., Adaptive Fuzzy Logic-Based Framework for Software

Development Effort Prediction, Information and Software Technology. Elsevier, 2004.

- 26) Crespo, F.J., Sicicila, M.A., Cuadrado, J.J., On the use of fuzzy regression in parametric software estimation models: integrating imprecision in COCOMO cost drivers. WSEAS Transactions on Systems, 2004.
- 27) M.R. Braz, S.R. Vergilio, Using fuzzy theory for effort estimation of object-oriented software". 16th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2004, 15-17 Nov 2004, pp. 196 – 201.
- 28) Xu, Z., Khoshgoftaar, T.M., 2004. Identification of fuzzy models of software cost estimation. Elsevier Fuzzy Sets and Systems 145 (1), 141–163.
- 29) L.M. Cuauhtemoc, Y.M. Cornelio, G.T. Agustin. Fuzzy Logic Systems for Software Development Effort Estimation Based Upon Clustering of Programs Segmented by Personal Practices", Electronics, Robotics and Automotive Mechanics Conference (CERMA'06), 2006, pp. 367-372.
- 30) Moon Ting Su, Teck Chaw Ling, Keat Keong Phang, Chee Sun Liew, Peck Yen Man, ENHANCED SOFTWARE DEVELOPMENT EFFORT AND COST ESTIMATION
- 31) USING FUZZY LOGIC MODEL, Malaysian Journal of Computer Science, Vol. 20(2), 2007
- 32) Iman Attarzadeh and Siew Hock Ow, Software Development Effort Estimation based on a New Fuzzy Logic Model, International Journal of Computer Theory and Engineering, Vol 1. No. 4 October 2009.

Class Break Point Determination Using CK Metrics Thresholds

Dr. E. Chandra¹, P. Edith Linda²

{ GJCST Classification
D.2.8 }

Abstract-The design and development of any project has got a well-defined project development cycle. But once the project or the product has been developed, it is subject to change due a lot a policy changes on the part of the organization or the government. These changes are implemented on the code but most of the time these changes are not reflected on the design document. This leads to inconsistencies in terms of design and code thereby causing depreciation in terms of quality. In this work we propose to use the object oriented metrics which uses the parameters mentioned in the CK metrics suite to assess the quality of the software at the class level. The proposed tool namely "Class Break Point" which could be used to determine the class design validity. This tool can be used to check if the class is adhering to the OO design specifications. The tool is useful in predicting the decomposition point of the class.

Keywords- OO metrics, Weighted Method Per Class, Depth of Inheritance Tree, LCOM, Response for a class, Number of Children, design refinement.

I. INTRODUCTION

The quality of software can be evaluated using metrics. Software quality has been a major challenge in various software projects. Quality has been a major issue in software development but it lacks in standards measuring quality. Metrics are the continuous application of measurement-based techniques to the software development process and its products to supply meaningful and timely management information together with the use of techniques to improve the process and its products [2]. The use of Metrics can help us *understand* more about our software products, processes and services. Metrics can be used to *evaluate* our software products, processes and services against established standards and goals. Metrics can provide the information we need to *control* resources and processes used to produce our software. Metrics can be used to *predict* attributes of software entities in the future [4]. A metrics program that is based on the goals of the organizations will help communicate those goals. People will work to accomplish what they believe to be important. Well-designed metrics with documented objectives can help our organization obtain the information it needs to continue to improve its software products, processes and services while maintaining a focus on what is important to that organization.

About¹-Director, Department of MCA, D J. Academy of Managerial Excellence, Coimbatore -32, India. Telephone:+919894255832. Email: crcspeech@gmail.com

About²-Assistant Professor, School of IT and Science, Dr. G. R Damodaran College of Science, Coimbatore-14, India. Telephone:+919894876043. Email:p.lindavinod@gmail.com

II. INTRODUCTION TO THE TOOL

When Object Oriented design was introduced, it guaranteed robust, maintainable and reusable systems. But these claims given by the object oriented designs are not fulfilled in large. Simply by using object oriented language or design does not assure a robust and reusable system. It mainly depends on the pattern or interdependencies between the subsystems and the communication between them. So the design of the object oriented system has to be validated to develop high quality applications which could be easily maintained and can be reused. The Design quality metrics is used to check the conformance of the object oriented design in the particular application. These design quality metrics provide information to the designers regarding the ability to survive to the changes [7].

The Code metrics is a set of software measures that provide developers better insight into the code they are developing. By taking advantage of code metrics, developers can understand which types and/or methods should be reworked or more thoroughly tested. Development teams can identify potential risks, understand the current state of a project, and track progress during software development [6].

Here in this paper we propose a tool called "Class Break Point" to evaluate the code quality of the software. Most of the works mentioned in the literature are covering the metrics that can be computed at the design level. But here in this paper we demonstrate the same set of metrics to evaluate the code quality. This tool could be used in the maintenance phase or even the developers could use this tool to find the adherence of code to the object oriented designing principles. Hence here the input to the tool would be the source code of existing software. Another type of input to the tool can be the code developed by the programmer to check the design quality of the code. Here in the second case the tool can be used as a self-evaluation tool to the programmer. In the first case the tool can be used to evaluate the breakthrough point of the class. In continuation with our previous work as mentioned in [1], this tool in the initial phase extracts only the parameters of the CK metrics tool. The main objective of the class break point tool is to check the quality of the source code and the adherence of the source code to the design documents. This tool finds its usage in the maintenance phase of the software. Here in this

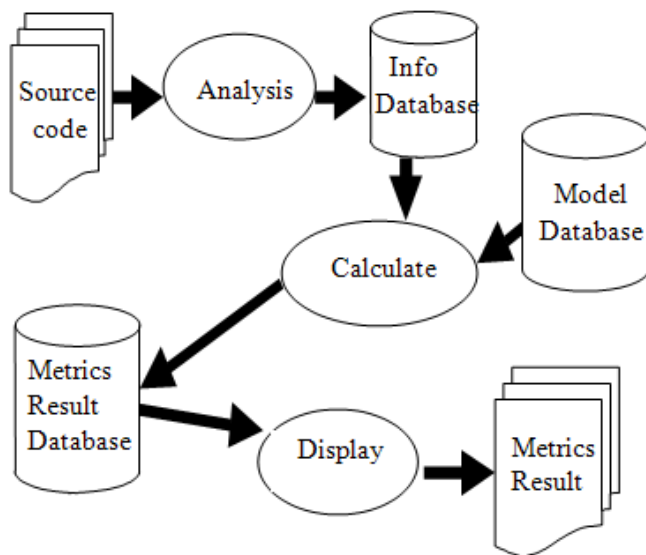


Fig1: Architecture of the Class Break Point Tool

phase after a period of year or two the software would have undergone lots of changes in code due to changing requirements. The changing requirements may be due to the change in the government policies or due the change in the company's policies. Hence the code is changed to meet the new set of requirements. When the code is changed, the developers usually change the code making the software ready to meet the new set of requirements. But most of the time this change is not reflected in the design document. Due to this factor it makes the next developer in jeopardy as the developer is not aware if the design document is up to date with the latest changes made.

- 1) Explanation of the parameters to be extracted by the code analyzer tool:

Parameters Identified to check the design validity of the class are same as the CK metrics suite[9, 10].

S.No	PARAMETERS OF CK METRIC	THRESHOL D
1	WMC	0-15
2	DIT	0-6
3	NOC	0-6
4	CBO	0-8
5	RFC	0-35
6	LCOM	0-1

Table 1: The threshold values for the CK metric suite parameters.

Metric 1: Weighted Methods per class (WMC):

The Weighted Method Per Class (WMC) – It is a count of sum of complexities of all methods in a class. To calculate the complexity of a class, the specific complexity metric that is chosen (e.g., cyclomatic complexity) should be

normalized so that nominal complexity for a method takes on value 1.0. Consider a class KI , with methods $M1, \dots, Mn$ that are defined in the class. Let $C1, \dots, Cn$ be the complexity of the methods[4].

$$WMC = \sum ci..cn$$

For the sake of simplicity we assume that the complexity of all the class is the same. Hence WMC is the sum of all the methods in the class. To compute WMC use the method `getDeclaredMethods()` to compute the number of methods in the class. If the number of methods in the class is high then the class is considered to be very complex. If the number of methods in the class is low the complexity of the class is less. If the number of methods in a class is less than or equal to 15, then the class can be considered to have normal complexity. The threshold limit is set to 15 per class.

```

public class MetricsWMC {
    public int getMethods(String className) {
        try {
            Class classObj = Class.forName(className);
            return classObj.getDeclaredMethods().length;
        }
        catch (Exception e) {
            e.printStackTrace();
        }
        return 0;
    }
}

```

Metric2:

Depth of Inheritance Tree (DIT)

Assess how deep, a class is in hierarchy structure. This metric assesses the potential reuse of a class and its probable ease of maintenance. A class with a small DIT has much potential for reuse it tends to be a general abstract class. On the other side, as a class gets deeper into a class hierarchy, it becomes more difficult to maintain due to the increased mental burden needed to capture it functionally. There are certain viewpoints regarding DIT. The deeper a class is in the hierarchy, the greater the number of methods it is likely to inherit, making it more complex” hence higher the value of DIT it is bad. Another viewpoint is Its useful to have a measure of how deep a particular class is in the hierarchy so that the class can be designed with reuse of inherited methods” ” hence higher the value of DIT it is good.

DIT = maximum inheritance path from the class to the root class.

In cases involving multiple inheritance, the DIT will be the maximum length from the node to the root of the tree. The theoretical Basis of the DIT metric is a measure of how many ancestor classes can potentially affect this class. The deeper a class is in the hierarchy, the more methods and variables it is likely to inherit, making it more complex. Deeper the tree the greater is the design complexity. Inheritance has to manage the complexity and to increase the reusability of the class and not to create issues with the design. A high DIT has been found to increase faults. Since the fault-prone classes may be at the middle of the tree, it

may contribute to the fault to the rest of the inheriting classes. This work suggests that lower DIT has great potential of reuse; hence a threshold value of 6 levels is set for DIT. To compute the value of DIT `getSuperclass()` is used to find the super class of an existing class. `getName()` is used to return the name of the super class. These two methods can be used in combination to get the list level of the inheritance of the particular class.

```
public int getDITValue(String className)
{
    int ditValue = 1;
    try {
        Class classObj = Class.forName(className);
        String superClass = classObj.getSuperclass().getName();
        while( ! "java.lang.Object".equals(superClass))
        {
            classObj = Class.forName(superClass);
            superClass = classObj.getSuperclass().getName();
            System.out.println(superClass);
            ditValue++;
        }
        catch (Exception e)
        { e.printStackTrace(); }
    }
    return ditValue;
}
```

Metric 3: Number of Children (NOC):

It is a simple measure of the number of classes associated with a given class using an inheritance relationship. It could be used to assess the potential influence that a class has on the overall design. NOC measures how many classes inherit directly methods or fields from a super-class. The greater the number of children in the inheritance hierarchy the greater the reuse. Then again a large number of children of a class might indicate improper abstraction for a parent class. High DIT value and low NOC means better reusability but the issue of maintainability is at stake. It also has a negative impact on understandability and is more difficult to modify. Since there are no empirical or theoretical boundary values, then we should find the proper threshold value for the system under development. Here in the system we have set the value of NOC to be 6 as same as DIT.

NOC = number of immediate sub-classes of a class

To determine the value of NOC the reflection classes, such as `Method`, are found in `java.lang.reflect`. There are three steps that must be followed to use these classes. The first step is to obtain a `java.lang.Class` object for the class that you want to manipulate. `java.lang.Class` is used to represent classes and interfaces in a running Java program. To obtain the class object use `Class c = Class.forName("java.lang.String");`. Then call a method such as `getDeclaredMethods()`, to get a list of all the methods declared by the class. Using both the information on hand tackle the specific application using reflection

```
public int getDITValue(String className)
{
    int ditValue = 0;
    try{
```

```
String packageName[] =
{"com.test.MWCTest2","com.test.MWCTest3","com.test.MWCTest4"};
Class classObj = Class.forName(className);
for (String string : packageName)
{
    Object obj = Class.forName(string).newInstance();
    if(obj instanceof MWCTest)
    {
        ditValue++;
    }
    catch (Exception e)
    {
        e.printStackTrace();
    }
    return ditValue;
}
```

Metric 4: Lack of Cohesion in Methods (LCOM)

It is the difference between the number of methods whose similarity is zero and not zero. The similarity of two methods is the numbers of attributes used were common. LCOM can judge the cohesiveness among the class methods. Low LCOM indicates high cohesiveness and vice versa. High LCOM indicates that a class shall be considered for splitting into two or more classes. However, a LCOM measure of zero is not strong evidence that a class enjoys cohesiveness. The single responsibility principle states that a class should not have more than one reason to change. Such a class is said to be cohesive. A high LCOM value generally pinpoints a poorly cohesive class. There are several LCOM metrics. The LCOM takes its values in the range 0 to 1. The computation of LCOM is as follows:

$$LCOM = 1 - \sum MF / (M * F)$$

Where:

- M is the number of methods in class (both static and instance methods are counted, it includes also constructors, properties getters/setters, events add/remove methods).
- F is the number of instance fields in the class.
- MF is the number of methods of the class accessing a particular instance field.
- Sum(MF) is the sum of MF over all instance fields of the class.

To find the value for LCOM find the number of methods in a class(M) using the method `getDeclaredMethods()`. Find the number of classes in the package level using the `getClass()`. Then find the number of instance fields in the class(F) using the `InstanceFieldAccess` class and its associated objects. The `InstanceFieldAccess` class defines an instance field `s`. The main method creates an object, sets the instance field, and then calls the native method `InstanceFieldAccess.accessField`. This native method prints out the existing value of the instance field and then sets the field to a new value. To know more about the class objects emphasis has to be put to the `instanceof` operator. `Class.isInstance` method can be used to simulate the

instanceof operator. `field.get(objectInstance)` where can be used to find the number of methods of the class accessing a particular instance field. After determining the value of the above mentioned parameter the formula mentioned above could be used for the computation of LCOM. At this juncture there are certain issues to be addressed with the LCOM metric. The reusability parameter is to be negatively influenced by LCOM. Higher the values for LCOM lower the scope for reusability. Also the maintainability for a class containing higher LCOM values is higher as it directly affects other classes also. Higher cohesion also decreases the changeability, stability and the portability of the classes as it triggers changes in the other classes with are closely coupled. Hence a good solution to this issue could be to keep the value of LCOM to a minimum value.

The Lack of Cohesion in Methods metric can be computed using the following three formats:

LCOM1: Take each pair of methods in the class and determine the set of fields they each access. If they have disjointed sets of field accesses, the count P increases by one. If they share at least one field access, Q increases by one. After considering each pair of methods:

$$\text{RESULT} = (P > Q) ? (P - Q) : 0$$

A low value indicates high coupling between methods. This also indicates potentially high reusability and good class design. Chidamber and Kemerer provided the definition of this metric in 1993.

LCOM2: This is an improved version of LCOM1. Say you define the following items in a class:

m: number of methods in a class

a: number of attributes in a class.

mA: number of methods that access the attribute a.

sum(mA): sum of all mA over all the attributes in the class.

$$\text{LCOM2} = 1 - \text{sum(mA)} / (m * a)$$

If the number of methods or variables in a class is zero (0), LCOM2 is undefined as displayed as zero (0).

LCOM3: This is another improvement on LCOM1 and LCOM2 and is proposed by Henderson-Sellers. It is defined as follows:

$$\text{LCOM3} = (m - \text{sum(mA)} / a) / (m - 1)$$

where m, a, mA, sum(mA) are as defined in LCOM2.

The LCOM3 value varies between 0 and 2. $\text{LCOM3} > 1$ indicates lack of cohesion and is considered a kind of alarm. If there is only one method in a class, LCOM 3 is undefined and also if there are no attributes in a class LCOM3 is also undefined and displayed as zero (0). Each of these different measures of LCOM has a unique way to calculate the value of LCOM. An extreme lack of cohesion such as $\text{LCOM3} > 1$ indicates that the particular class should be split into two or more classes. If all the member attributes of a class are only accessed outside of the class and never accessed within the class, LCOM3 will show a high-value. A slightly high value of LCOM means that you can improve the design by either splitting the classes or re-arranging certain methods within a set of classes [3].

Metric 5: Coupling between objects (CBO).

When one object interacts with another object that is a coupling. Strong coupling means that one object is strongly

coupled with the implementation details of another object. Strong coupling is discouraged because it results in less flexible, less scalable application. However, coupling can be used so that it enables objects to talk to each other while also preserving the scalability and flexibility. OO metrics can help you to measure the right level of coupling. CBO is defined as the number of non-inherited classes associated with the target class. It is counted as the number of types that are used in attributes, parameters, return types, throws clauses, etc. Primitive types and system types (e.g. `java.lang.*`) are not counted. Method Invocation Coupling (MIC) is defined as the relative number of classes that receive messages from a particular class.

$$\text{MIC} = \text{nMIC} / (N - 1)$$

Where N = total number of classes defined within the project.

nMIC = total number of classes that receive a message from the target class

to find the value of N (total number of classes in the system) use the following the methods. The fully qualified class name (including package name) is obtained using the `getName()` method. The class name without the package name can be obtained using the `getSimpleName()` method. You can access the modifiers of a class via the Class object. The class modifiers are the keywords "public", "private", "static" etc. The modifiers are packed into an int where each modifier is a flag bit that is either set or cleared. You can check the modifiers using these methods in the class `java.lang.reflect.Modifiers`. The method used to get the modifier is `getModifiers()`. The `Method.invoke(Object target, Object ... parameters)` method takes an optional amount of parameters, but you must supply exactly one parameter per argument in the method you are invoking. Using the above methods to extract the values and apply in the formulae the value of CBO can be obtained.

Metric 6: Response for Class (RFC)

It is defined as a count of the set of methods that can be potentially executed in response to a message received by an instance of the class. Response set of an class = { set of all methods that can be invoked in response to a message to the object }

$$\text{RFC} = |\text{RS}| \text{ where RS is the response set for the class}$$

The RFC is defined as the total number of methods that can be executed in response to a message to a class. This count includes all the methods available in the whole class hierarchy. If a class is capable of producing a vast number of outcomes in response to a message, it makes testing more difficult for all the possible outcomes. Response For a Class (RFC) is the sum of the number of its methods and the total of all other methods that they directly invoke. If the number of methods invoked in response to a message received by an object is large, the maintenance and testing are more demanding. Large number of method invocation means more testing and debugging. Larger the method invocation greater is the complexity. The response set of a class is a set of methods that can potentially be executed in response to a message received by an object of that class. RFC is simply the number of methods in the set.

$RFC = M + R$ (First-step measure)

$RFC' = M + R'$ (Full measure) M = number of methods in the class

R = number of remote methods directly called by methods of the class

R' = number of remote methods called, recursively through the entire call tree

A given method is counted only once in R (and R') even if it is executed by several methods M . Since RFC specifically includes methods called from outside the class, it is also a measure of the potential communication between the class and other classes. A large RFC has been found to indicate more faults. Classes with a high RFC are more complex and harder to understand. Testing and debugging is complicated. It counts only the first level of calls outside of the class. RFC' measures the full response set, including methods called by the callers, recursively, until no new remote methods can be found. If the called method is polymorphic, all the possible remote methods executed are included in R and R' . The use of RFC' should be preferred over RFC . RFC was originally defined as a first-level metric because it was not practical to consider the full call tree in manual calculation. With an automated code analysis tool, getting RFC' values is not longer problematic. As RFC' considers the entire call tree and not just one first level of it, it provides a more thorough measurement of the code executed.

```
private void incRFC(String className, String methodName,
Type[] arguments) {
String argumentList = Arrays.asList(arguments).toString();
String args = argumentList.substring(1,
argumentList.length() - 1);
String signature = className + "." + methodName + "(" +
args + ")";
responseSet.add(signature);
}
```

III. RESULTS AND DISCUSSIONS

As per the description given in the previous sections the parameters identified are: WMC, DIT, NOC, LCOM, CBO, RFC. Each of these parameters are extracted from the source code. A threshold is set for each of the parameters and the conformance of the extracted values are expected with the set threshold values i.e the parameters values are expected to lie within the range of the thresholds. If the value lies within the range then the class satisfies the OO design paradigms. If the parameters lie outside the threshold then design refinement suggestion has to be provided to the developer or user. The results for the first three parameters in three projects are depicted below in the table:

Class Name	WMC	DIT	NOC
com.test.MWCTest2	6	1	2
com.test.MWCTest3	5	3	2
com.test.MWCTest	8	2	7
com.test.MWCTest4	5	4	2

Table2. The extracted parameter values for the first three parameters of the CK metrics suite.

Hence in the above classes the extracted parameter values are within the threshold so no design refinements regarding the reusability are suggested. But in classes where the extracted value does not fall into the threshold then decision has to be made to split the class to meet the OO design specifications.

IV. REFERENCES

- 1) Dr. E. Chandra and P. Edith Linda, Assessment of of Software Quality through Object Oriented Metrics, CIIT International Journal of Software Engineering, Vol2, Iss:2 Feb 2010
- 2) Goodman, Paul. 1993. Practical Implementation of Software Metrics. London: McGraw Hill.
- 3) Heung Seok Chae, Yong Rae Kwon, and Doo Hwan Bae, Improving Cohesion Metrics for Classes by Considering Dependent Instance Variables, IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, VOL. 30, NO. 11, NOVEMBER 2004
- 4) Humphrey, Watts S. 1989. *Managing the Software Process*. Reading: Addison-Wesley.
- 5) <http://javaboutique.internet.com/tutorials/coupcoh/index-2.html>
- 6) <http://msdn.microsoft.com/enus/library/bb385914.aspx>
- 7) <http://www.objectmentor.com/resources/articles/oodmetric.pdf>
- 8) Ramanath Subramanyam and M.S. Krishnan, Empirical Analysis of CK Metrics for Object-Oriented Design Complexity: Implications for Software Defects, IEEE Transactions on Software Engineering, Vol. 29, No. 4, April 2003
- 9) Shyam R. Chidamber and Chris F. Kemerer, A Metrics Suite for Object Oriented Design, IEEE Transactions on Software Engineering. Vol. 20, No. 6, June 1994
- 10) Yuming Zhou and Hareton Leung, Empirical Analysis of Object-Oriented Design Metrics for Predicting High and Low Severity Faults, IEEE Transactions On Software Engineering, Vol. 32, No. 10, October 2006

Fuzzy Logic Based Digital Image Edge Detection

Aborisade, D.O

GJCST Classification
I.4.6

Abstract -In this paper fuzzy based edge detection algorithm is developed. In the proposed algorithm, edginess at each pixel of a digital image is calculated using three (3) 3×3 linear spatial filters i.e. low-pass, high-pass and edge enhancement (Sobel) filters through spatial convolution process. Edge strength values derived from the three masks form three set of edges, utilized as fuzzy system inputs. Based on the Gaussian membership functions and fuzzy rules the fuzzy system decide whether a pixel in focus belong to an edge or non-edge. The final pixel classification of a given image is produced by generating output pixel value using Mamdani defuzzifier method. The robustness of the proposed method results for different captured images are compared to those obtained with Sobel and Prewitt operators. Experimental results show the ability and high performance of proposed algorithm.

Keywords- Fuzzy Logic, Fuzzy inference system, Edge strength, Edge detection

I. INTRODUCTION

An edge is defined as an abrupt variation in pixel intensity within an image while the process of detecting outlines of an object and boundaries between objects and the background in the image is known as edge detection. Edge detection is a very important tool widely used in many computer vision and image processing applications. Earlier edge detection methods, such as Sobel, Prewitt, Kirsch and Robert are based on the calculation of the intensity gradient magnitude at each image pixel. In these algorithms, the gradient value is compared to the threshold value and a pixel location is classified as an edge if the value of the gradient is higher than a threshold. Gradient-based edge detectors have a major drawback of being very sensitive to noise. In order to counter noise problems Canny proposed an approach to edge detection in which the image is convolved with the first order derivatives of Gaussian filter for smoothing in the local gradient direction followed by edge detection and thresholding [Canny 1986]. Extensions of Canny edge detector can be found in [Laligant *et al.* 1994].

Edge detection represents an extremely important step facilitating higher-level image analysis and therefore remains an area of active research, with new approaches continually being developed. Recent examples include edge detectors using fuzzy logic, neural networks, or wavelets [Sun and Scialabassi 1995, Law *et al.* 1996, Bezdek *et al.* 1996, Wang *et al.* 2005]. Comparison of edge detection approaches and an assessment of their performance may be found in [Demigny *et al.* 1995, Ramesh and Haralick 1994]. In this paper, fuzzy logic based approach to edge

detection in digital images is proposed. Firstly, for each pixel in the input image edginess' measure is calculated using three 3×3 linear filters after which three fuzzy sets characterized by three (3) Gaussian membership functions associated to linguistic variable "Low", "Medium" and "High" were created to represent each of the edge strengths. The second phase involves application of fuzzy inference rule to the three fuzzy sets to modify the membership values in such a way that the fuzzy system output ("edge") is high only for those pixels belonging to edges in the input image. The last step is final pixel classification as edge or non-edge using Mamdani defuzzification method.

II. APPLICATION OF FUZZY LOGIC BASED EDGE DETECTION

Fuzzy logic represents a powerful approach to decision making [Zadeh 1965, Kaufmann 1975, Bezdek 1981]. Since the concept of fuzzy logic was formulated in 1965 by Zadeh, many researches have been carried out on its application in the various areas of digital image processing such as image quality assessment, edge detection, image segmentation, etc. Many techniques have been suggested by researchers in the past for fuzzy logic-based edge detection [Cheung and Chan 1995, Kuo, *et al.* 1997, El-Khamy *et al.* 2000]. In [Zhao, *et al.* 2001], Zhao, *et al.* proposed an edge detection technique based on probability partition of the image into 3-fuzzy partitions (regions) and the principle of maximum entropy for finding the parameters value that result in the best compact edge representation of images. In their proposed technique the necessary condition for the entropy function to reach its maximum is derived. Based on this condition an effective algorithm for three-level thresholding is obtained. Several approaches on fuzzy logic based edge detection have been reported based on fuzzy If-Then rules [Tao, *et al.* 1993, Li 1997]. In most of these methods, adjacent points of pixels are assumed in some classes and then fuzzy system inference are implemented using appropriate membership function, defined for each class [Mahani, *et al.* 2008]. In Liang, *et al.* [Liang, *et al.* 2003], adjacent points are assumed as 3×3 sets around the concerned point. By predefining membership function to detect edges. In these rules discontinuity in the color of different 3×3 sets, edges are extracted. It uses 5 fuzzy rules and predefined membership function to detect edges. In these rules discontinuity of adjacent point around the concerned point are investigated. If this difference is similar to one of predefined sets, the pixel is assumed as edge. A similar work is proposed by Mansoori, *et al.* [Mansoori, *et al.* 2006], wherein adjacent points of each pixel are grouped in six different set. Then by using of appropriate bell shape membership function, the value from zero to one is determined for each group. Based on the membership

values, and fuzzy rules, decision about existing/not existing and direction of edge pixels are obtained.

III. PROPOSED ALGORITHM

In this paper, at first an input image is pre-process to accentuate or remove a band of spatial frequencies and to locate in an image where there is a sudden variation in the grey level of pixels. For each pixel in the image edge strength value is calculated with three (3) 3×3 linear spatial filters i.e. low-pass, high-pass and edge enhancement filters (Sobel) through spatial convolution process. In carrying out a 3×3 kernel convolution, nine convolution coefficients called the convolution mask are defined and labeled as seen below:

a	b	c
d	e	f
g	h	i

Every pixel in the input image is evaluated with its eight neighbors, using each of the three masks shown in Figure 1 to produce edge strength value. The equation used for the calculation of edginess values between the center pixel and the neighborhood pixels of the three (3) masks using spatial convolution process is given by:

$$\begin{aligned} O(x, y) = & aI(x-1, y-1) + bI(x-1, y) + cI(x-1, y+1) \\ & + dI(x, y-1) + eI(x, y) + fI(x, y+1) \\ & + gI(x+1, y-1) + hI(x+1, y) + iI(x+1, y+1) \end{aligned} \quad (1)$$

However, the result of convolution of the two Sobel kernels is combine thus, the approximate absolute gradient

magnitude (edge strength) at each point is computed as:

$$O_g = |O_x| + |O_y| \quad (2)$$

The normalized edge strength is then defined as:

$$NO(x, y) = \text{round}[(O(x, y) / \max(O)) \times 100] \quad (3)$$

where $x = \{0, 1, \dots, M-1\}$ and $y = \{0, 1, \dots, N-1\}$ for an M-by-N image.

The edge strength values derived from the three (3) masks served as the inputs used in the construction of the fuzzy inference system based on which decision on pixel as belonging to an edge or not are made. Membership functions are defined for fuzzy system inputs. Many membership functions have been introduced in the literature.

In the proposed edge detection Gaussian membership functions are used. To apply these functions, each of the

edge strength values of O_g , O_{Hp} , and O_{Lp}

are mapped into fuzzy domain between 0 and 1, relative to the normalized gray levels between 0 and 100, using Gaussian membership functions given as

$$\mu_{mn} = G(x_{mn}) = e^{[-(x_{\max} - x_{mn})^2 / 2\sigma^2]} \quad (4)$$

where $G(x_{mn})$ is a Gaussian function, x_{\max} , x_{mn} are the maximum and (m, n) th gray values respectively and σ is the standard deviation associated with the input variable.

$$h_{LP} = \begin{bmatrix} \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \end{bmatrix}, \quad h_{HP} = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

$$h_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}, \quad h_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

Fig. 1: 3×3 kernels used for edge detection

Each of the mapped values are partition into three fuzzy regions Low", Medium", and High". The defined regions and membership functions are shown in Fig. 2. Fuzzy inference rules are applied to assign the three fuzzy sets characterized by membership functions μ_{Low} , μ_{Medium} , and μ_{High}

to the output set. The rules, tabulated in Table 1 are defined in such a way that in the fuzzy inference system, output set E_L , E_M , and E_H correspond to pixels with low, medium and high probability value respectively. The output of the system P_{Final} representing the probability used for final pixel classification as edge or non-edge was computed using a singleton fuzzifier, Mamdani defuzzifier method given by;

$$p_{Final} = \frac{\sum_{\ell=1}^M \bar{y}^{\ell} (\prod_{i=1}^n \mu_{k_i^{\ell}}(\alpha_i))}{\sum_{\ell=1}^M (\prod_{i=1}^n \mu_{k_i^{\ell}}(\alpha_i))} \quad (5)$$

where α_i are the fuzzy sets associated with the antecedent part of the fuzzy rule base, \bar{y}^{ℓ} is the output class center and M is the number of fuzzy rules being considered.

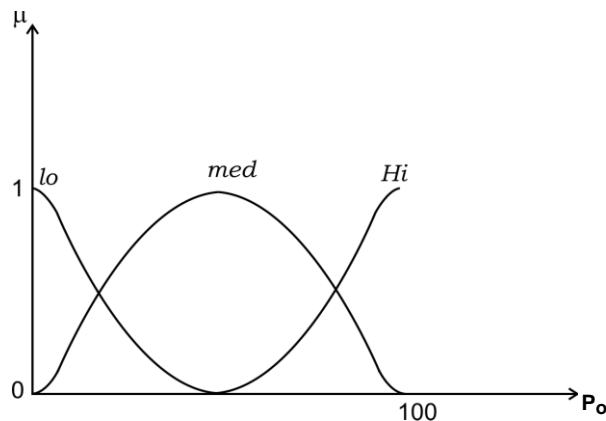


Fig. 2: Gaussian membership functions Low, Medium and High

Table 1

If $edginess_{Lp}$ is LO and $edginess_{So}$ is LO and $edginess_{Hp}$ is LO then p_{edge} is E_L
If $edginess_{Lp}$ is LO and $edginess_{So}$ is LO and $edginess_{Hp}$ is MD then p_{edge} is E_L
If $edginess_{Lp}$ is LO and $edginess_{So}$ is LO and $edginess_{Hp}$ is HI then p_{edge} is E_L
If $edginess_{Lp}$ is LO and $edginess_{So}$ is MD and $edginess_{Hp}$ is LO then p_{edge} is E_L
If $edginess_{Lp}$ is LO and $edginess_{So}$ is MD and $edginess_{Hp}$ is MD then p_{edge} is E_L
If $edginess_{Lp}$ is LO and $edginess_{So}$ is MD and $edginess_{Hp}$ is HI then p_{edge} is E_M
If $edginess_{Lp}$ is LO and $edginess_{So}$ is HI and $edginess_{Hp}$ is LO then p_{edge} is E_L
If $edginess_{Lp}$ is LO and $edginess_{So}$ is HI and $edginess_{Hp}$ is MD then p_{edge} is E_H
If $edginess_{Lp}$ is LO and $edginess_{So}$ is HI and $edginess_{Hp}$ is HI then p_{edge} is E_H
If $edginess_{Lp}$ is MD and $edginess_{So}$ is LO and $edginess_{Hp}$ is LO then p_{edge} is E_L
If $edginess_{Lp}$ is MD and $edginess_{So}$ is MD and $edginess_{Hp}$ is LO then p_{edge} is E_L
.
.
.
If $edginess_{Lp}$ is HI and $edginess_{So}$ is LO and $edginess_{Hp}$ is HI then p_{edge} is E_H
If $edginess_{Lp}$ is HI and $edginess_{So}$ is MD and $edginess_{Hp}$ is HI then p_{edge} is E_H
If $edginess_{Lp}$ is HI and $edginess_{So}$ is HI and $edginess_{Hp}$ is HI then p_{edge} is E_H

IV. EXPERIMENTAL RESULTS

The proposed fuzzy edge detection method was simulated using MATLAB on different images, its performance are compared to that of the Sobel and Kirsch operators. Samples for a set of four test images are shown in Fig. 3(a). The edge detection based on Sobel and Kirsch operators using the image processing toolbox in MATLAB with threshold automatically estimated from image's binary value is illustrated in Fig. 3(b) and 3(c). The sample output of the proposed fuzzy technique is shown in Fig. 3(d). The resulting images generated by the fuzzy method seem to be much smoother with less noise and has an exhaustive set of fuzzy conditions which helps to provide an efficient edge representation for images with a very high efficiency than the conventional gradient-based methods (Sobel and Kirsch methods).

V. CONCLUSION

In this paper, effective fuzzy logic based edge detection has been presented. This technique uses the edge strength information derived using three (3) masks to avoid detection

of spurious edges corresponding to noise, which is often the case with conventional gradient-based techniques. The three edge strength values used as fuzzy system inputs were fuzzified using Gaussian membership functions. Fuzzy if-then rules are applied to modify the membership to one of low, medium, or high classes. Finally, Mamdani defuzzifier method is applied to produce the final edge image. Through the simulation results, it is shown that the proposed algorithm is far less computationally expensive, its application on the image improve the quality of edges as much as possible compared to the Sobel and Kirsch methods. This algorithm is suitable for applications in various areas of digital image processing such as face recognition, fingerprint identification, remote sensing and medical imaging where boundaries of specific regions need to be determined for further image analysis.

VI. ACKNOWLEDGEMENT

The author is grateful to Engineer I.A Isaiah at Ladoke Akintola University of Technology, Ogbomosho, Nigeria for his helpful advice.

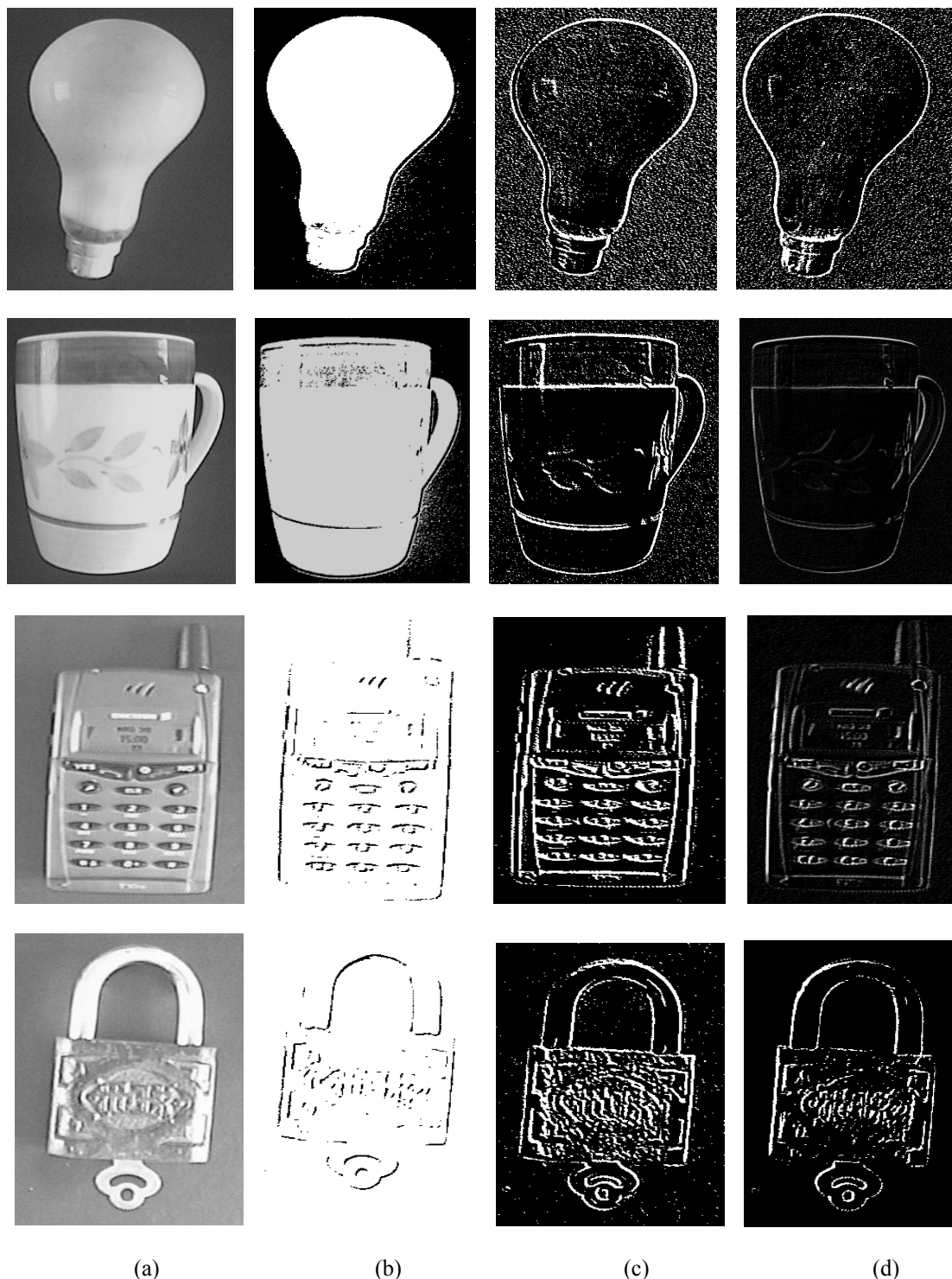


Fig. 3 (a) Original images, (b) Sobel operator results, (c) Kirsch operator results, (d) Proposed fuzzy edge detection algorithm results

VII. REFERENCES

- 1) Canny, J.F., "A computational approach to edge detection", IEEE Trans. on Pattern Analysis and Machine Intelligence, 8(6), 1986, 679-698.
- 2) . Laligant, F. Truchete, and J. Miteran , "Edge detection by multiscale merging," in Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis, pp. 237-240, IEEE, Los Alamitos, CA, 1994.
- 3) M. Sun and R. J. Sclabassi, "Sym metric wavelet edge detector of the minimum length," In International Conference on Image Processing, pp. 177-180. IEEE, Los Alamitos, CA, 1995.

- 4) T. Law, H. Itoh, and H. Seki, Image filtering, edge detection, and edge tracing using fuzzy reasoning," IEEE Transaction on Pattern Analysis and Machine Intelligence, 18: 481-491, 1996.
- 5) J. C. Bezdek, R. Chandrasekhar, and Y. Attikiouzel, A new fuzzy model for edge detection ," In Applications of Fuzzy Logic Technology III, Proc. SPIE Vol. 2761, pp. 11-28, SPIE, Bellingham, WA, 1996.
- 6) R. Wang, L. Gao, S. Yang, and Y. Liu, "An Edge detection method by combining fuzzy logic and neural networks," Machine Learning and Cybernetics, 7(2005) 4539-4543, 18-21, Aug 2005.
- 7) D. Demigny, F.G. Lorca, and L. Kessal, Evaluation of edge detectors performances with a discrete expression of Canny's criteria," In International Conference on Image Processing, pp. 169-172, IEEE, Los Alamitos, CA, 1995.
- 8) V. Ramesh and R. M. Haralick, An integrated gradient edge detector. Theory and performance evaluation," In ARPA Image Understanding Workshop, Monterey, CA, pp. 689-702, ARPA, Los Altos, CA, 1994.
- 9) L. A. Zadeh, Fuzzy sets," Information and Control, 8: 338-353, 1965.
- 10) A. Kaufmann, Introduction to the Theory of Fuzzy Subsets Fundamentals Theoretical Elements, Vol. 1. Academic Press, New York, 1975.
- 11) L.C. Bezdek, Pattern Recognition with fuzzy Objective Function Algorithm," Plenum Press, New York, 1981.
- 12) K. Cheung and W. Chan, "Fuzzy One Mean Algorithm for Edge Detection," IEEE Inter. Conf. On Fuzzy Systems, pp. 2039-2044, 1995.
- 13) Y. Kuo, C. Lee, and C. Liu, "A New Fuzzy Edge Detection Method for image Enhancement," IEEE Inter. Conf. on Fuzzy Systems, pp. 1069-1074, 1997.
- 14) S. El-Khamy, N. El-Yamany, and M. Lotfy, "A Modified Fuzzy Sobel Edge Detector," Seventeenth National Radio Science Conference (NRSC'2000), February 22-24, Minufia, Egypt, 2000.
- 15) M. Zhao, A. M. N. Fu, and H. Yan, A Technique of Three-Level Thresholding Based on Probability Partition a Fuzzy 3-Partition". IEEE Trans. on Fuzzy Systems, vol.9, no.3, pp. 469- 479, June 2001.
- 16) Tao, C. W. et al(1993), A Fuzzy if-then approach to edge detection, Proc. of 2nd IEEE intl.conf. on fuzzy systems, 1356-1361.
- 17) Li, W. (1997), Recognizing white line markings for vision-guided vehicle navigation by fuzzy reasoning, Pattern Recognition Letters, 18: 771- 780.
- 18) M. N. Mahani, M. K. Moqadam, H. N. pour, Technology, Kish, 2008, (In Persian).
- 19) L. Liang and C. Looney, Competitive Fuzzy Edge Detection," Applied Soft Computing, (3) 123-137, 2003.
- 20) G. Mansoori and H. Eghbali, Heuristic edge detection using fuzzy rule-based classifier," Journal of Intelligent and Fuzzy Systems, Pages 457 - 469, Volume 17, Number 5 / 2006.

Secure Mining of Association Rules over Horizontally Partitioned Data in Data Mining

M.Mohan Rao ,Asst.professor¹, M.Laxmaiah²

GJCST Classification
H.2.8

Abstract-Data mining can extract important knowledge from large data collection. Sometimes these large collections are split among various parties, which can share the data. For example, insurance companies share the data from medical hospitals. Privacy concerns may prevent the parties from directly sharing the data. Here this project addresses secure mining of association rules over horizontally partitioned data. The existing work provides security by using some techniques like Randomization, Secure Multi Party computation, etc., The drawbacks in the existing work are inaccuracy, inefficiency and lacking of security. To overcome the drawbacks of existing system we proposed a new method by using Commutative Encryption tool, Randomization and Secure multi party computation. This work also provides security for multiple numbers of sites. Examples include knowledge discovery among intelligence services of different countries and collaboration among corporations without revealing trade secrets.

Keywords-Data Mining, Security, Privacy.

I. INTRODUCTION

Data mining technology has emerged as a means of identifying patterns and trends from large quantities of data. Data mining and data warehousing go hand-in-hand: Most tools operate by gathering all data into a central site, then running an algorithm against that data. However, privacy concerns can prevent building a centralized warehouse data may be distributed among several custodians, none of which are allowed to transfer their data to another site. This paper addresses the problem of computing association rules. The goal is to produce association rules that hold globally while limiting the information shared about each site. We can compute the global support and confidence of an association rule $AB \Rightarrow C$ knowing only the local supports of AB and ABC and size of each database.

$$\text{support}_{AB \Rightarrow C} = \frac{\sum_{i=1}^{\text{sites}} \text{support count}_{ABC}(i)}{\sum_{i=1}^{\text{sites}} \text{database_size}(i)}$$

$$\text{support}_{AB} = \frac{\sum_{i=1}^{\text{sites}} \text{support count}_{AB}(i)}{\sum_{i=1}^{\text{sites}} \text{database_size}(i)}$$

$$\text{confidence}_{AB \Rightarrow C} = \text{support}_{AB \Rightarrow C} / \text{support}_{AB}$$

1) Private Association Rule Mining Overview:

Our method follows the two-phase approach described above, but combining locally generated rules and support counts is done by passing encrypted values between sites. The two phases are discovering candidate itemsets (those

that are frequent on one or more sites) and determining which of the candidate itemsets meet the global support/confidence thresholds. The first phase (Fig. 1) uses commutative encryption. Each party encrypts its own frequent itemsets (e.g., Site 1 encrypts itemset C). The encrypted itemsets are then passed to other parties until all parties have encrypted all itemsets. These are passed to a common party to eliminate duplicates and to begin decryption. (In the figure, the full set of itemsets are shown to the left of Site 1, after Site 1 decrypts.) This set is then passed to each party and each party decrypts each itemset. The final result is the common itemsets (C and D in the figure). In the second phase (Fig. 2), each of the locally supported itemsets is tested to see if it is supported globally. In the figure, the itemset ABC is known to be supported at one or more sites and each computes their local support. The first site chooses a random value R and adds to R the amount by which its support for ABC exceeds the minimum support threshold. This value is passed to site 2, which adds the amount by which its support exceeds the threshold (note that this may be negative, as shown in the figure.) This is passed to site 3, which again adds its excess support. The resulting value (18) is tested using a secure comparison to see if it exceeds the Random value (17). If so, itemset ABC is supported globally.

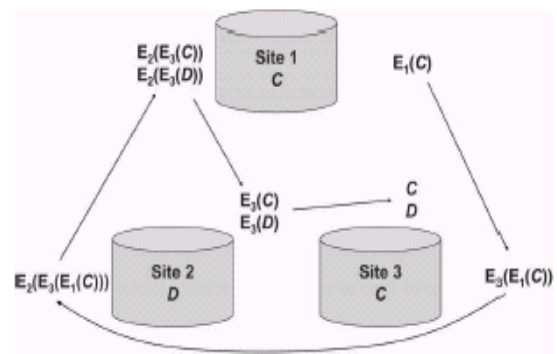


Fig. 1. Determining global candidate itemsets

About¹- Tirumala engg college, Mohanrao19@yahoo.com

About²- Associate Professor & Hod, Tirumala engg College ,
laxman_mettu@yahoo.com.

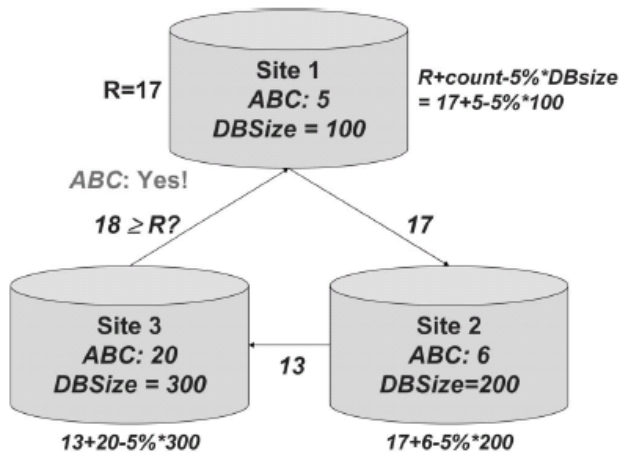


Fig. 2. Determining if item set support exceeds 5 percent threshold

II. BACKGROUND AND RELATED WORK

There are several fields where related work is occurring. We first describe other work in privacy-preserving data mining, then go into detail on specific background work on which this paper builds. Previous work in privacy-preserving data mining has addressed two issues. In one, the aim is preserving customer privacy by distorting the data values [2]. The idea is that the distorted data does not reveal private information and thus is safe to use for mining. More recently, the data distortion approach has been applied to Boolean association rules. Again, the idea is to modify data values such that reconstruction of the values for any individual transaction is difficult, but the rules learned on the distorted data are still valid. One interesting feature of this work is a flexible definition of privacy, e.g., the ability to correctly guess a value of “1” from the distorted data can be considered a greater threat to privacy than correctly learning a 0. The other approach uses cryptographic tools to build decision trees. In this work, the goal is to securely build an ID3 decision tree where the training set is distributed between two parties. The basic idea is that finding the attribute that maximizes information gain is equivalent to finding the attribute that minimizes the conditional entropy. The conditional entropy for an attribute for two parties can be written as a sum of the expression of the form $(v_1 + v_2) \log(v_1 + v_2)$. The authors give a way to securely calculate the expression $(v_1 + v_2) \log(v_1 + v_2)$ and show how to use this function for building the ID3 securely.

1) Mining of Association Rules

The association rules mining problem can be defined as follows [1]: Let $I = \{i_1, i_2, \dots, i_n\}$ be a set of items. Let DB be a set of transactions where each transaction T is an itemset such that $T \subseteq I$. Given an itemset $X \subseteq I$, a transaction T contains X if and only if $X \subseteq T$. An association rule is an implication of the form $X \Rightarrow Y$, where $X \subseteq I$, $Y \subseteq I$, and $X \cap Y = \emptyset$. The rule $X \Rightarrow Y$ has support s in the transaction database DB if $s\%$ of transactions in DB contains $X \cup Y$.

The association rule holds in the transaction database DB with confidence c if $c\%$ of transactions in DB that contain X also contains Y . An itemset X with k items is called k -itemset. The problem of mining association rules is to find all rules whose support and confidence are higher than certain user-specified minimum support and confidence. In this simplified definition of the association rules, missing items, negatives, and quantities are not considered. In this respect, transaction database DB can be seen as 0/1 matrix where each column is an item and each row is a transaction. In this paper, we use this view of association rules.

2) Distributed Mining of Association Rules:

The above problem of mining association rules can be extended to distributed environments. Let us assume that a transaction database DB is horizontally partitioned among n sites (namely, S_1, S_2, \dots, S_n) where $DB = DB_1 \cup DB_2 \cup \dots \cup DB_n$ and DB_i resides at site S_i ($1 \leq i \leq n$). The itemset X has local support count of $X.\text{sup}_i$ at site S_i if $X.\text{sup}_i$ of the transactions contains X . The global support count of X is given as $X.\text{sup} = \sum_{i=1}^n X.\text{sup}_i$. An itemset X is globally supported if $X.\text{sup} \geq s * \sum_{i=1}^n |DB_i|$. Global confidence of a rule $X \Rightarrow Y$ can be given as $\{X \cup Y\}.\text{sup} / X.\text{sup}$. A fast algorithm for distributed association rule mining is given in Cheung et al. [1]. Their procedure for fast distributed mining of association rules (FDM) is summarized below:

1) Candidate Sets Generation: Generate candidate sets $CG_{l(k)}$ based on $GL_{l(k-1)}$, itemsets that are supported by the S_i at the $(k-1)$ th iteration, using the classic a priori candidate generation algorithm. Each site generates candidates based on the intersection of globally large $(k-1)$ itemsets and locally large $(k-1)$ itemsets.

2) Local Pruning: For each $X \in CG_{l(k)}$, scan the database DB_i at S_i to compute $X.\text{sup}_i$. If X is locally large S_i , it is included in the LL_i set. It is clear that if X is supported globally, it will be supported in one site.

3) Support Count Exchange: $LL_{i(k)}$ are broadcast and each site computes the local support for the items in $ULL_{i(k)}$.

4) Broadcast Mining Results: Each site broadcasts the local support for itemsets in $ULL_{i(k)}$. From this, each site is able to compute $L_{(k)}$.

3) Secure Multiparty Computation:

Substantial work has been done on secure multiparty computation. The key result is that a wide class of computations can be computed securely under reasonable assumptions. We give a brief overview of this work, concentrating on material that is used later in the paper. The definitions given here are from Goldreich. For simplicity, we concentrate on the two-party case. Extending the definitions to the multiparty case is straightforward.

1) Security in Semihonest Model:

A semihonest party follows the rules of the protocol using its correct input, but is free to later use what it sees during execution of the protocol to compromise security. This is somewhat realistic in the real world because parties who want to mine data for their mutual benefit will follow the protocol to get correct results. Also, a protocol that is buried

in large, complex software cannot be easily altered. A formal definition of private two-party computation in the semihonest model is defined. Computing a function privately is equivalent to computing it securely.

2) Yao's General Two-Party Secure Function Evaluation

Yao's general secure two-party evaluation is based on expressing the function $f(x,y)$ as a circuit and encrypting the gates for secure evaluation [3]. With this protocol, any two-party function can be evaluated securely in the semihonest model. To be efficiently evaluated, however, the functions must have a small circuit representation. We will not give details of this generic method; however, we do use this generic result for securely finding whether $a \geq b$ (Yao's millionaire problem). For comparing any two integers securely, Yao's generic method is one of the most efficient methods known, although other asymptotically equivalent but

practically more efficient algorithms could be used as well.

3) Commutative Encryption

Commutative encryption is an important tool that can be used in many privacy-preserving protocols. An encryption algorithm is commutative if the following two equations hold for any given feasible encryption keys $k_1, k_2, \dots, k_n \in K$, any message M , and any permutations of i, j :

$$E_{k_{i1}}(\dots E_{k_{in}}(M) \dots) = E_{j_{i1}}(\dots E_{j_{in}}(M) \dots)$$

$\forall M_1, M_2 \in M$ such that $M_1 \neq M_2$ and for given $k, \epsilon < 1/2^k$

$$\Pr(E_{k_{i1}}(\dots E_{k_{in}}(M) \dots) = E_{j_{i1}}(\dots E_{j_{in}}(M) \dots)) < \epsilon$$

These properties of commutative encryption can be used to check whether two items are equal without revealing them. For example, assume that party A has item i_A and party B has item i_B . To check if the items are equal, each party encrypts its item and sends it to the other party: In addition to meeting the above requirements, we require that the encryption be secure. Specifically, the encrypted values of a set of items should reveal no information about the items themselves.

III. SECURE ASSOCIATION RULE MINING

We will now use the tools described above to construct a distributed association rule mining algorithm that preserves the privacy of individual site results. The algorithm given is for three or more parties.

4) Problem Definition

Let $i \geq 3$ be the number of sites. Each site has a private transaction database DB_i . We are given support threshold s and confidence c as percentages. The goal is to discover all association rules satisfying the thresholds, as defined in Section 2.1.1. We further desire that disclosure be limited: No site should be able to learn contents of a transaction at any other site, what rules are supported by any other site, or the specific value of support/confidence for any rule at any other site unless that information is revealed by knowledge of one's own data and the final result. For example, if a rule is supported globally but not at one's own site, we can

deduce that at least one other site supports the rule. Here, we assume no collusion.

5) Method

Our method follows the general approach of the FDM algorithm, with special protocols replacing the broadcasts of $LL_{i(k)}$ and the support count of items $LL_{i(k)}$. We first give a method for finding the union of locally supported itemsets without revealing the originator of the particular itemset. We then provide a method for securely testing if the support count exceeds the threshold.

6) Secure Union of Locally Large Itemsets

In the FDM algorithm (Section 2.1.1), Step 3 reveals the large item sets supported by each site. To accomplish this without revealing what each site supports, we instead exchange locally large itemsets in a way that obscures the source of each itemset. We assume a secure commutative encryption algorithm with negligible collision probability. The main idea is that each site encrypts the locally supported itemsets, along with enough "fake" itemsets to hide the actual number supported. Each site then encrypts the itemsets from other sites. In Phases 2 and 3, the sets of encrypted itemsets are merged. Since (3) holds, duplicates in the locally supported itemsets will be duplicates in the encrypted itemsets and can be deleted. The reason this occurs in two phases is that if a site knows which fully encrypted itemsets come from which sites, it can compute the size of the intersection between any set of sites. While generally innocuous, if it has this information for itself, it can guess at the itemsets supported by other sites. Permuting the order after encryption in Phase 1 prevents knowing exactly which itemsets match; however, separately merging itemsets from odd and even sites in Phase 2 prevents any site from knowing the fully encrypted values of its own itemsets. Phase 4 decrypts the merged frequent itemsets. Commutativity of encryption allows us to decrypt all itemsets in the same order regardless of the order they were encrypted in, preventing sites from tracking the source of each itemset. The detailed algorithm is given in Protocol 1 (see Fig. 3).

IV. CONCLUSIONS

Naturally, ever increasing data collection, along with the influx of analysis tools capable of handling huge volumes of information, has led to privacy concerns. The cryptographic community has formally defined much stronger notions of privacy. Also cryptography tools can enable data mining that would otherwise be prevented due to security concerns. We have given procedures to mine distributed association rules on horizontally partitioned data. We have shown that distributed association rule mining can be done efficiently under reasonable security assumptions. We believe the need for mining of data where access is restricted by privacy concerns will increase. Examples include knowledge discovery among intelligence services of different countries and collaboration among corporations without revealing trade secrets. Even within a single multinational company,

privacy laws in different jurisdictions may prevent sharing individual data. Allowing error in the results may enable more efficient algorithms that maintain the desired level of security. Continued research will expand the scope of privacy-preserving data mining, enabling most or all data mining methods applied in situations where privacy concerns would appear to restrict such mining.

Protocol 1 Finding secure union of large itemsets of size k
Require: $N = 3$ sites numbered $1, \dots, N$, set F of nonitemsets

Phase 0: Encryption of all the rules by all sites

for each site i do

generate $LL_i(k)$ as in steps 1 and 2 of the FDM algorithm

$LL_i(k) = \emptyset$;

for each $X \in LL_i(k)$ do

$LL_i(k) = LL_i(k) \cup \{E_i(X)\}$

end for

for $j = |LL_i(k)| + 1$ to $|CG(k)|$ do

$LL_i(k) = LL_i(k) \cup \{E_i(\text{random selection from } F)\}$

end for

end for

Phase 1: Encryption by all sites

for Round $j = 0$ to $N - 1$ do

if Round $j = 0$ then

Each site i sends permuted $LL_i(k)$ to site $(i+1) \bmod N$

else

Each site i encrypts all items in $LL_{(i-j \bmod N)}(k)$

with E_i , permutes, and sends it to site $(i+1) \bmod N$

end if

end for {At the end of Phase 1, site i has the itemsets of site $(i+1) \bmod N$ encrypted by every site}

Phase 2: Merge odd/even itemsets

Each site i sends $LL_{i+1 \bmod N}$ to site $i \bmod N$

Site 0 sets $RuleSet1 = [d(N-1)/2e]$

$j=1$ $LL_{(2j-1)}(k)$

Site 1 sets $RuleSet0 = [b(N-1)/2c]$

$j=0$ $LL_{(2j)}(k)$

Phase 3: Merge all itemsets

Site 1 sends permuted $RuleSet1$ to site 0

Site 0 sets $RuleSet = RuleSet0 \cup RuleSet1$

Phase 4: Decryption

for $i = 0$ to $N - 1$ do

Site i decrypts items in $RuleSet$ using D_i

Site i sends permuted $RuleSet$ to site $i+1 \bmod N$

end for

Site $N-1$ decrypts items in $RuleSet$ using D_{N-1}

$RuleSet(k) = RuleSet$

Site $N-1$ broadcasts $RuleSet(k)$ to sites $0..N-2$

Fig. 3. Protocol 1: Finding secure union of large itemsets of size k .

3.3 Algorithm to Calculate Global Support

Step 1 Calculate s for the data item whose support count is to be calculated

Step 2 Calculate d

Step 3 choose r such that $0 < r < d$

Step 4 calculate $a = d/k$

Step 5 Calculate $t = r + a$

Step 6 if $s > t$ then the data item is public

Else the data item is private

Thus the global support is calculated and the secret and the non secret data are thus separated.

V. REFERENCES

- 1) Achlioptas, D. (2004). Random matrices in data analysis. Proceedings of the 15th European Conference on Machine Learning, pp. 1-8, Pisa, Italy.
- 2) J. Vaidya and C. Clifton, Privacy Preserving Association Rule Mining in Vertically Partitioned Data," Proc. Eighth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 639-644, 2002, available: <http://doi.acm.org/10.1145/775047.775142>.
- 3) Wang, J., Zhang, J., Zhong, W. J., & Xu, S. (2007). A novel data distortion approach via selective SSVD for privacy protection. *International Journal of Information and Computer Security*, to appear.
- 4) R. Agrawal and R. Srikant, Privacy-Pre-serving Data Mining," Proc. 2000 ACM SIGMOD Conf. Management of Data, pp. 439-450, 2000, available: <http://doi.acm.org/10.1145/342009.335438>.
- 5) A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, Privacy Preserving Mining of Association Rules," Proc. Eighth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 217-228, 2002, available: <http://doi.acm.org/10.1145/775047.775080>.
- 6) R.L. Rivest, A. Shamir, and L. Adleman, A Method for Obtaining Digital Signatures and Public-Key Cryptosystems," Comm. ACM, vol. 21, no. 2, pp. 120-126, 1978, available: <http://doi.acm.org/10.1145/359340.359342>.
- 7) S.C. Pohlig and M.E. Hellman, An Improved Algorithm for Computing Logarithms over GF(p) and Its Cryptographic Significance," IEEE Trans. Information Theory, vol. IT-24, pp. 106-110, 1978.
- 8) O. Goldreich, Encryption Schemes," (working draft), Mar. 2003, available: http://www.wisdom.eizmann.ac.il/~oded/PSBookFr_ag/enc.ps.
- 9) Y. Lindell and B. Pinkas, Privacy Preserving Data Mining," Advances in Cryptology (CRYPTO 2000), pp. 36-54, 2000, available: <http://link.springer.de/link/service/series/0558/bibs/1880/18800036.htm>.
- 10) I. Ioannidis and A. Grama, An Efficient Protocol for Yao's Millionaires' Problem," Proc. Hawaii Int'l Conf. System Sciences (HICSS-36), 2003.
- 11) M.K. Reiter and A.D. Rubin, Crowds: Anonymity for Web Transactions," ACM Trans. Information

- and System Security, vol. 1, no. 1, pp. 66-92, Nov. 1998, available: <http://doi.acm.org/10.1145/290163.290168>.
- 12) J.C. Benaloh, Secret Sharing Homomorphisms: Keeping Shares of a Secret Secret," Advances in Cryptography (CRYPTO86): Proc., A. Odlyzko, ed., pp. 251-260, 1986, available:
 - 13) B. Chor and E. Kushilevitz, A Communication-Privacy Tradeoff for Modular Addition," Information Processing Letters, vol. 45, no. 4, pp. 205-210, 1993.
 - 14) C. Clifton, M. Kantarcioglu, and J. Vaidya, Defining Privacy for Data Mining," Proc. US Nat'l Science Foundation Workshop on Next Generation Data Mining, H. Kargupta, A. Joshi, and K. Sivakumar, eds., pp. 126-133, 2002.
 - 15) T. ElGamal, A Public Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms," IEEE Trans. Information Theory, vol. IT-31, no. 4, pp. 469-472, July 1985.
 - 16) A. Shamir, R.L. Rivest, and L.M. Adleman, Mental Poker," Technical Memo MIT-LCS-TM-125, Laboratory for Computer Science, MIT, Fe

Theory of Carry Value Transformation (CVT) and its Application in Fractal formation

Pabitra Pal Choudhury¹, Birendra Kumar Nayak², Sudhakar Sahoo³, Sk. Sarif Hassan⁴ {GJCST Classification
G.1.2}

Abstract- In this paper the theory of Carry Value Transformation (CVT) is designed and developed on a pair of n-bit strings and is used to produce many interesting patterns. One of them is found to be a self-similar fractal whose dimension is same as the dimension of the Sierpinski triangle. Different construction procedures like L-system, Cellular Automata rule, Tilling for this fractal are obtained which signifies that like other tools CVT can also be used for the formation of self-similar fractals. It is shown that CVT can be used for the production of periodic as well as chaotic patterns. Also, the analytical and algebraic properties of CVT are discussed. The definition of CVT in two-dimension is slightly modified and its mathematical properties are highlighted. Finally, the extension of CVT and modified CVT (MCVT) are done in higher dimensions. **Keywords-** Carry Value Transformation, Fractals, L-System, Cellular Automata and Tilling, Discrete Dynamical System.

I. INTRODUCTION

Benoit Mandelbrot coined the word fractal from the Latin adjective fractus. The corresponding Latin verb frangere means to break to create irregular fragments. The precise definition of Fractal according to Benoit Mandelbrot is as a set for which the Hausdorff Besicovitch dimension strictly exceeds the topological dimension [1]. Many things in nature are very complex, chaotic but exhibit some self-similarity. The complexity of fractals and the property of self-similarity have a large set of real world applications. Fractals can be generated using construction procedure/algorithms simply by repetition of mathematical formula that are often recursive and ideally suited to computer. In this paper a new transformation named as Carry Value Transformation (CVT) is defined in binary number system and using CVT we are trying to explore that behind the complexity of nature there remains a simple methodology about which most often we are ignorant. Earlier, in [2] we have used CVT with Cellular Automata in efficient hardware design of some basic arithmetic operations. But in this paper although we emphasize the formation of self-similar-fractals, the algorithm using CVT also produces periodic and chaotic patterns. The underlying development of CVT is same as the concept of Carry Save Adder (CSA) [3] where carry or overflow bits generated in

the addition process of two integers are saved in the memory. Here we perform the bit wise XOR operation of the operands to get a string of sum-bits (ignoring the carry-in) and simultaneously the bit wise ANDing of the operands to get a string of carry-bits, the latter string is padded with a 0 on the right to signify that there is no carry-in to the LSB. The organization of the paper is as follows. Section-2 discusses some of the basic concepts on fractals, L-systems, tilling problem, Cellular Automata etc. which are used in the subsequent sections. The concept of CVT is defined in section-3. It can be found in section-4 that CVT generates a beautiful self-similar fractal whose dimension is found to be same as that of monster fractal, Sierpinski triangle. Section-5 deals with the various ways like L-System, Cellular Automata and Tillings by which the same fractal in binary number system can be constructed. CVT can also be used for the production of periodic as well as chaotic patterns are shown in section-6. The definition of CVT in two-dimension is slightly modified and its mathematical properties are highlighted in section-7. Section-8 deals with the extension of CVT and modified CVT (MCVT) in higher dimensions. In section-9 we have explored discrete dynamical property of modified CVT (MCVT). On highlighting other possible applications of CVT and some future research directions a conclusion is drawn in section 10. It should be noted that a very preliminary version of this paper has been published in an international conference [13].

II. FRACTAL BASICS

This sections deals with the basics of Fractals [1, 4, 5] and the various ways the fractals may be constructed like L-Systems [6], Cellular Automata [7, 8] and synthesis of Tilling [9] etc. It is intended for readers who are not conversant with the fundamentals of these concepts.

2.1 Why we study fractals

We feel very much worried due to our inability to describe using the traditional Euclidean Geometry, the shape of cloud, a mountain, a coastline or a tree. In nature, clouds are not really spherical, mountains are not conical, coastlines are not circular, even the lightning doesn't travel in a straight line. More generally, we could be able to conclude that many patterns of nature are so irregular and fragmented, that, compared with *Euclid Geometry* a term, can be used in this regard to denote all of standard geometry. Mathematicians have over the years disdained this challenge and have increasingly chosen to flee nature by devising theories unrelated to natural objects we can see or feel.

About^{1&4}- Applied Statistics Unit, Indian Statistical Institute, Kolkata, 700108, INDIA Email: pabitrpalchoudhury@gmail.com & sarimif@gmail.com

About²- Institute of Mathematics and Applications, Andharua, Bhubaneswar-751003

Email: sudhakar.sahoo@gmail.com

About³- P.G. Department of Mathematics, Utkal University, Bhubaneswar-751004 Email: bknatu@yahoo.co.uk

After a long time, responding to this challenge, Benoit Mandelbrot developed a new geometry of nature and implemented its use in a number of diverse arenas of science such as Astronomy, Biology, Mathematics, Physics, and Geography and so on [1, 4, 5, 10, 11]. This new-born geometry can describe many of the irregular and fragmented (chaotic) patterns around us, and leads to full-fledged theories, by identifying a family of shapes, now-a-days which we people call 'FRACTALS'.

2.2 Measuring Fractal dimension

The fractal dimension alone does not give an idea of what "fractals" are really about. Mandelbrot founded his insights in the idea of self similarity, requiring that a true fractal "fracture" or break apart into smaller pieces that resemble the whole. This is a special case of the idea that there should be a dynamical system underlying the geometry of the set. This is partly why the idea fractals have become so popular throughout science; it is a fundamental aim of science to seek to understand the underlying dynamical properties of any natural phenomena. It has now become apparent that relatively simple dynamics, more precisely dynamical system can produce the fantastically intricate shapes and behavior that occur throughout nature. Now let us try to define what fractal dimension (self-similarity dimension) is. Given a self-similar structure [1], there is a relation between the reduction factor (scaling factor) 'S' and the number of pieces 'N' into which the structure can be divided; and that relation is...

$$N = 1/SD, \text{ equivalently, } D = \log(N)/\log(1/S)$$

This 'D' is called the Fractal dimension (Self-similarity dimension)

2.3 Ways to construct fractals

2.3.1 Lindenmayer Systems (L-system) produces fractals

As a biologist, Aristid Lindenmayer [6] studied growth patterns in various types of algae. In 1968 he developed Lindenmayer systems (or L-Systems) as a mathematical formalism for describing the growth of simple multi-cellular organisms. The central concept of L-System is that of rewriting. In general rewriting is a technique for defining complex object by successively replacing parts of a simple initial object using a set of production rules. Definition of an L-System: An L-system is a formal grammar consisting of 4 parts: A set of *variables*: symbols that can be replaced by production rules. A set of *constants*: symbols that do not get replaced. An *axiom*, which is a string, composed of some number of variables and/or constants. The axiom is the initial state of the system. A set of *production rules* defining the way variables can be replaced with combinations of constants and other variables. A production consists of two strings - the predecessor and the successor.

2.3.2 Fractals by Cellular Automata rules

The scientific output of Wolfram's [7, 8] work played a central role in launching Cellular Automata (CA) as a new field of science to understand the complexity of nature.

Starting from an initial seed he studied the space-time diagram of all the 256 three-neighborhood elementary CA rules and classified the rules into four distinct classes according to the complexity of the pattern. According to him the Class 2 rules deals with the periodic and fractal patterns.

In 1-D the *global state* or simply *state* of a CA at any time-instant t is represented as a vector $X^t = (x_1^t, x_2^t, \dots, x_n^t)$ where x_i^t denotes the bit in the i^{th} cell x_i at time-instant t . If the "present state" of an n -bit CA (at time t) is X^t , its "next state" (at time $t+1$), denoted by X^{t+1} , is in general given by the *global mapping* $F(X^t) = (f^1(lb^t, x_1^t, x_2^t), f^2(x_1^t, x_2^t, x_3^t), \dots, f^n(x_{n-1}^t, x_n^t, rb^t))$, where f^i is a local mapping to the i^{th} cell and lb and rb denote respectively the left boundary of x_1 and right boundary of x_n in case of periodic boundary CA and those values are 0 in case of null boundary CA. If the same local mapping (rule) determines the "next" bit in each cell of a CA, the CA will be called a Uniform CA, otherwise it will be called a Hybrid CA. For our purpose, we have used one-way CA, which allows only one-way communication, i.e., in a 1-D array each cell depends only on itself and its left neighbor. One can also consider dependence on the cell and its right neighbor. However both sides dependence is not allowed. Just like L-system and Cellular Automata, fractals can also be obtained by Iterated Function Systems [4] and using different Tiles [9]. A complex figure can be easily (in most of the cases) synthesized by using of tiles. Next section discusses a new and efficient construction tool named as CVT by which uncountable number of fractal patterns can be generated.

III. CARRY VALUE TRANSFORMATION (CVT)

The carry or overflow bits are usually generated at the time of addition between two n -bit strings. In the usual addition process, carry value is always a single bit and if generated then it is added column wise with other bits and not saved in its own place. But the carry value defined here are the usual carries generated bit wise and stored in their respective places as shown in "Fig. 1".

$$\begin{array}{rcl} \text{carry value} & = & c_n \quad c_{n-1} \dots c_1 \quad 0 \\ a & = & a_n \quad a_{n-1} \dots a_1 \\ b & = & b_n \quad b_{n-1} \dots b_1 \\ a \oplus b & = & a_n \oplus b_n \quad a_{n-1} \oplus b_{n-1} \dots a_1 \oplus b_1 \end{array}$$

Figure 1: Carry generated in i^{th} column is saved in $(i-1)^{\text{th}}$ column

Thus to find out the carry value we perform the bit wise XOR operation of the operands to get a string of sum-bits (ignoring the carry-in) and simultaneously the bit wise ANDing of the operands to get a string of carry-bits, the latter string is padded with a '0' on the right to signify that there is no carry-in to the LSB. Thus the corresponding decimal value of the string of carry bits is always an even integer. Now we can give a precise definition of CVT as follows: Let $B = \{0,1\}$ and CVT is a mapping defined as $CVT: (B_n \times B_n) \rightarrow B_{n+1}$ where B_n is the set of strings of length n on $B = \{0,1\}$. More specifically, if

$a = (a_n, a_{n-1}, \dots, a_1)$ and $b = (b_n, b_{n-1}, \dots, b_1)$ then

$CVT(a, b) = (a_n \wedge b_n, a_{n-1} \wedge b_{n-1}, \dots, a_1 \wedge b_1, 0)$ is an $(n+1)$ bit string, belonging to Z , set of non-negative integers, and can be computed bit wise by logical AND operation followed by a 0, which denotes no carry is generated in the LSB at the time of addition procedure. In other words, CVT is a mapping from $Z \times Z \rightarrow Z$ where Z is set of non-negative integers.

Illustration:

Suppose, we want the CVT of the numbers $(13)_{10} \equiv (1101)_2$ and $(14)_{10} \equiv (1110)_2$. Both are 4-bit numbers. The carry value is computed as follows:

Carry: 1 1 0 0 0

Augend: 1 1 0 1

Addend: 1 1 1 0

XOR: 0 0 1 1

Figure 2: Carry generated in i th column is saved in $(i-1)$ th column

Conceptually, in the general addition process the carry or overflow bit from each stage (if any) goes to the next stage so that, in each stage after the first (i.e. the LSB position with no carry-in), actually a 3-bit addition is performed instead of a 2-bit addition by means of the full adder. Instead of going for this traditional method, what we do is that we perform the bit wise XOR operation of the operands (ignoring the carry-in of each stage from the previous stage) and simultaneously the bit wise ANDing of the operands to get a string of carry-bits, the latter string is padded with a '0' on the right to signify that there is no carry-in to the LSB (the overflow bit of this ANDing being always '0' is simply ignored). In our example, bit wise XOR gives $(0011)_2 \equiv (3)_{10}$ and bit wise ANDing followed by zero-padding gives $(11000)_2 \equiv (24)_{10}$. Thus $CVT(1101, 1110) = 11000$ and equivalently in decimal notation one can write $CVT(13, 14) = 24$. In the next section we have used the carry value in decimal to construct the CV table.

IV. GENERATION OF SELF SIMILAR-FRACTAL USING CVT

A table is constructed that contains only the carry values (or even terms) defined above between all possible integers a 's and b 's arranged in an ascending order of x and y -axis respectively. We observe some interesting patterns in the table. We would like to make it clear how the CV-table is constructed.

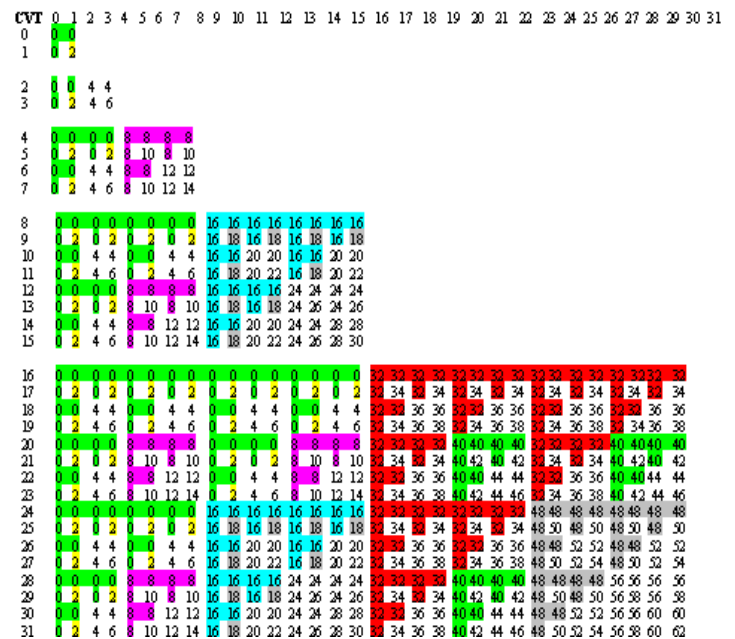
Step 1. Arrange all the integers 0 1 2 3 4 5 6 ... (as long as we want) in ascending order and place it in both, uppermost row and leftmost column in a table.

Step 2. Compute $CVT(a, b)$ as mentioned in section III and store it in decimal form in the (a, b) position.

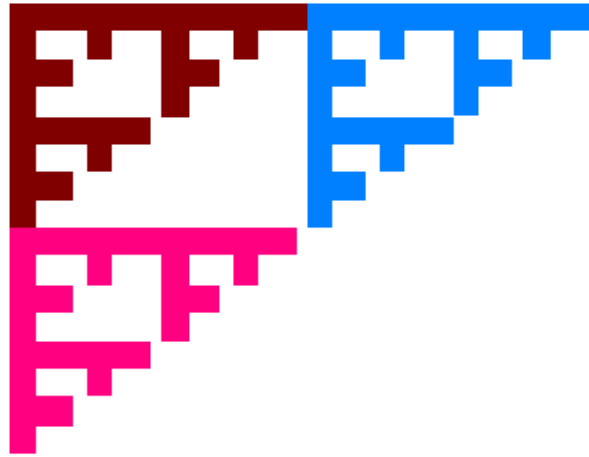
Step 3. Then we look on the pattern of any integer, and we have made it color. This shows a very beautiful consistent picture, which we see as a fractal as shown in table I followed by Fig. 3".

Choosing different sequence of rows and columns in the CV-table in step 1, one can obtain uncountable number of patterns only by computing a single operation $CVT(a, b)$ in each entry position (a, b) of the table. Thus one of the advantages of CVT over other construction tools lies in the number of cell evolutions for the formation of fractals. In this case only one uniform evolution is required at each cell position to obtain the required fractal pattern where as in case of L-System, Cellular Automata, Iterated Function System etc many iterations are required.

Illustration:



[Figure 3: A fractal structure on using CVT of different integer values]



filling the values in its right, bottom and diagonal (bottom-right) positions and recursively this will leads to (4x4), (8x8), (16x16) matrices...etc that can be seen in fig-7.4 and can be constructed by the following recursive formula.

$$A_0 = 0_{(1 \times 1)}$$

$$A_k = \begin{bmatrix} A_{k-1} & A_{k-1} \\ A_{k-1} & A_{k-1} + 2^k \end{bmatrix}_{(2^k \times 2^k)} \quad \text{for } k \geq 1$$

Where, A_k is a matrix of dimension $(2^k \times 2^k)$.

3. A recursive pattern exists in each square block of size $(2^k \times 2^k)$ for $k=0, 1, 2, 3, \dots$ etc. starting from (0, 0) position (in —Fig3— viewing as a matrix) and if

$$1^{st} \text{ quadrant: } a_{ij} \begin{matrix} 0 \leq i \leq 2^{k-1}-1 \\ 2^{k-1} \leq j \leq 2^k-1 \end{matrix} = a_{i(j \bmod 2^{k-1})}$$

$$2^{nd} \text{ quadrant: } a_{ij} \quad 0 \leq i, j \leq 2^{k-1}-1$$

$$3^{rd} \text{ quadrant: } a_{ij} \begin{matrix} 2^{k-1} \leq i \leq 2^k-1 \\ 0 \leq j \leq 2^{k-1}-1 \end{matrix} = a_{(i \bmod 2^{k-1})j}$$

$$4^{th} \text{ quadrant: } a_{ij} \quad 2^{k-1} \leq i, j \leq 2^k-1 = a_{(i \bmod 2^{k-1})(j \bmod 2^{k-1})} + 2^k$$

4. One can partition a square block of size $(2^k \times 2^k)$ for $k=0, 1, 2, 3, \dots$ etc. into two different classes. Where the elements of 1^{st} , 2^{nd} and 3^{rd} are in Class-I and the elements of 4^{th} quadrant is in Class-II. Further Class-I elements can be obtained from Class-II by element wise modulo 2^k operations ($b = (a_{ij}) \bmod 2^k$). Thus 2^k can act as a pivot.

Dimension of this fractal

For this fractal, $N=3$, $S=1/2$, where 1 is the initial length.

Fractal dimension D is given by ...

$$3 = 1/(1/2)^D \text{ or } D = \log 3 / \log 2 \approx 1.585$$

This is same as the dimension of Sierpinski triangle. Thus CVT fractal as obtained by us can be regarded as a relative of Sierpinski triangle [5].

Interestingly, in Fig. 3— if we multiply the row and column elements by 2^n ($n \geq 0$), and construct the CV table using the above construction procedure then also

this block is partitioned in the middle (both in row and column) into 4 equal sub blocks of size $(2^{k-1} \times 2^{k-1})$ for $k > 1$ then also the same pattern can be observed. If the 4-blocks are treated as 4 quadrants (1^{st} , 2^{nd} , 3^{rd} and 4^{th}) in the two dimensional Euclidean space then the patterns in first three quadrants are exactly same and in the 4^{th} quadrant the pattern is same only values are different and that can also be easily constructed by adding 2^k to each element of any of the three quadrants. That is if a_{ij} denote the entry of i^{th} row and j^{th} column of the CV-table of size $(2^k \times 2^k)$, then

we will get the same Sierpinski fractal as shown in Fig4—.

If an odd sequence i.e. $\{1, 3, 5, 7, 9, 11, 13, 15, \dots\}$ is taken in both row and column of Fig. 3—, the pattern that emerges using the similar construction procedure as above is same as the relatives of a Sierpinski triangle. Now, if we multiply each element of the sequence by 2^n ($n \geq 0$), then this sequence also lead to the same pattern. Therefore, one can say that Sierpinski gasket thus generated is very stable.

In Bogomolny's website <http://www.cut-the-knot.org/ctk/Sierpinski.shtml>, a list of different techniques are given for the construction of Sierpinski gasket (also, the Sierpinski triangle) and also the author conjectured that there are undoubtedly more. Thus CVT defined in this paper is another way for the construction of this kind of fractal. Next section discusses other various ways like L-system, Cellular Automata rules etc. to construct the self-similar CVT fractal as obtained in Fig. 4—.

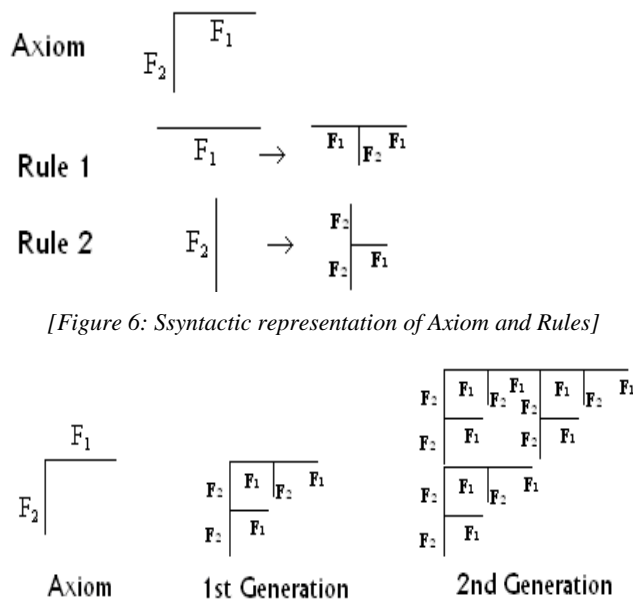
V. VARIOUS WAYS TO CONSTRUCT THE CVT FRACTAL

5.1 L- System to generate CVT fractal

Let F_1 , F_2 denote horizontal and vertical line segments respectively. Starting from the axiom as shown in table II and Fig 5— two rules are defined which produces the above fractal generated by CVT, after infinitely many iterations.

Table 1

L-System for CVT fractal	
Variables:	F_1, F_2
Constants:	-
Axiom:	$F_2 - F_1$
Rules:	$F_1 = F_1 - F_2 - F_2 - F_1$ $F_2 = F_2 - F_1 - F_1 - F_2$
Angle increment:	90 degrees



[Figure 6: Ssyntactic representation of Axiom and Rules]

[Figure 7: L-System is used to generate the above fractal]

5.2 Cellular Automata rules to construct CVT fractal

In [4] we found that the fractal generated from one-dimensional, two-neighborhood, binary Cellular Automata rule starting from an initial sheet, if rotated by an angle of

180 degree is same as our CVT fractal. According to Wolfram naming convention [10], [11] this is rule number 6, whose truth table and initial sheet on which rule 6 would be applied are given below...

0	0
	0

0	1
	1

1	0
	1

1	1
	0

[illegible]

[Figure 8: The rules and the initial seed of CA to generate the CVT fractal]

After finitely many steps, the space-time pattern for the above CA rule is shown in Fig. 8”.

0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
0	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0
0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0

[Figure 9: CA evolution from the initial seed generates the CVT fractal]

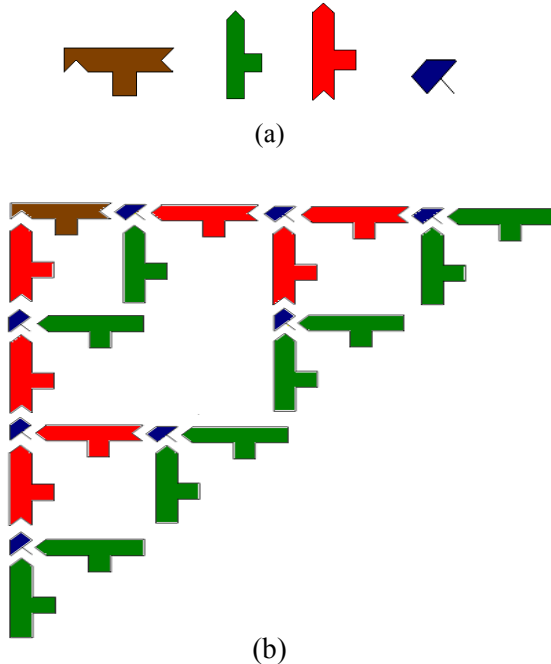
Fig9” shows the same CA evolution by replacing 1 by a black cell and 0 by a white cell as usual done by Wolfram. Now by flipping vertically (same as the vertical rotation) this CA evolution we get the same CVT fractal as shown in Fig. 3.

[illegible]

[Figure 10: The rules and the initial seed of CA to generate the CVT fractal]

5.3 Synthesis of CVT fractal by Tiling

The constituent parts on synthesis can give rise to the fractal picture. Here we have used four key tiles those are used to generate CVT fractal shown in Fig. 10".



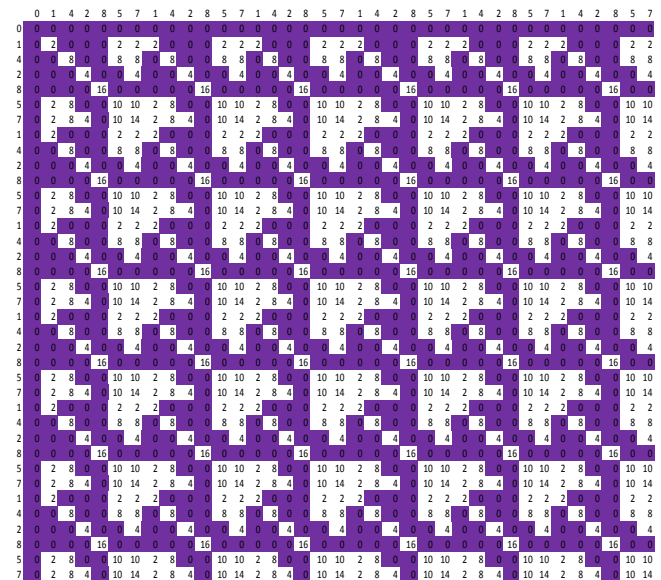
[Figure 11: (a) shows four key tiles those are used to generate the CVT fractal and (b) shows the arrangements of these keys.]

Different ways of construction of the self-similar CVT fractal signifies that like other tools CVT can also be used as a construction tool for the formation of different kinds of self-similar fractals. Next section shows that not only the CV-table produces self-similar fractals but CVT can also be used to generate periodic as well as chaotic patterns.

VI. CVT FOR GENERATING PERIODIC AND CHAOTIC PATTERNS

Already we have shown in section 4, that CVT can produce a self-similar pattern (fractal). Here we would like to produce another type of source to have a periodic pattern. For that, first of all we need a periodic sequence of numbers where we can be able to apply CVT to obtain the pattern. And we have a renowned domain of such sequences those are obtained considering decimal representation of rational numbers. Let us consider an example of rational number $1/7$. The decimal representation of $1/7$ is $0.142857142857142857142857 \dots$ (142857 is being repeated).

Taking the above decimal representation as a sequence of integers, let us construct the CV-Table in binary number system as follows...



[Figure 12: a fractal structure on using CVT of different integer values]

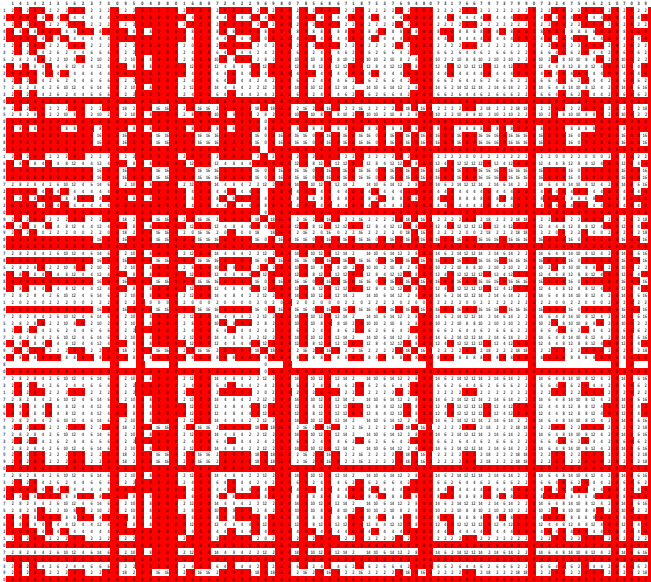
Here we have considered the pattern of zeros, which we have made it colored. Clearly this pattern of zeros is periodic. Here, we can conclude that for every rational number we can be able to have periodic pattern using CVT, sometimes it may be followed by some chaotic patterns as if starts from an explosion then things becomes stable. That is, here we can encounter a source of countably infinite number of almost periodic patterns corresponding to rational numbers because we know that the set of rational numbers is countable and infinite.

6.1 Formation of Chaotic Pattern

We have not yet seen that whether CVT can produce a non-periodic or chaotic pattern or not! Now we are ready to demonstrate an example where we could have a chaotic pattern using CVT. First of all we need a random sequence of integers, and then we will be applying CVT as we have applied earlier. Here also we have a good domain of such sequences, which can be considered the decimal representation of irrational numbers.

Let us consider an example of an irrational number $\sqrt{2}$. The decimal representation of $\sqrt{2}$ is $1.414213562373095048801688724229698078568671875376948073176679737990 \dots$ (Non-periodic, non-recurring).

Considering the above sequence let us try to construct the CV-table in binary number system as shown in table IV.



[Figure 13: a fractal structure on using CVT of different integer values]

Here we have considered the pattern of zeros, which we have made it colored. Clearly this pattern of zeros is a non-periodic (chaotic).

Here, we can conclude that for every irrational number we can be able to have non-periodic (chaotic) pattern using CVT. That is here from we can be able to conclude that an uncountable number of periodic and non periodic pattern corresponding to any real numbers can be obtained through CVT. As we know that the set of real numbers is uncountable.

It can be noted that this definition of CVT when treated as a binary operation doesn't satisfy algebraic properties such as Closure, Associative, Existence of identity and Existence of inverse except Commutativity. But as we have seen in section 4 CVT could produce fractals. Now we can slightly modify our CVT definition and redefine it so that without affecting the fractal formation it also satisfies some of those interesting algebraic properties.

VII. MODIFIED CARRY VALUE TRANSFORMATION (MCVT)

Formal definition of MCVT

MCVT is a mapping defined as $MCVT: (B_n \times B_n) \rightarrow B_n$ where B_n is the set of strings of length n on $B = \{0,1\}$.

More specifically, if $a = (a_n, a_{n-1}, \dots, a_1)$ and $b = (b_n, b_{n-1}, \dots, b_1)$ then

$MCVT(a, b) = (a_n \wedge b_n, a_{n-1} \wedge b_{n-1}, \dots, a_1 \wedge b_1)$ is an n bit string and can be computed bit wise by logical AND operation, which denotes no carry, is generated in the LSB at the time of addition procedure.

$$\begin{aligned} \text{carry value} &= c_n && c_{n-1} \dots \dots \dots c_1 \\ a &= a_n && a_{n-1} \dots \dots \dots a_1 \\ b &= b_n && b_{n-1} \dots \dots \dots b_1 \\ a \oplus b &= a_n \oplus b_n && a_{n-1} \oplus b_{n-1} \dots \dots \dots a_1 \oplus b_1 \end{aligned}$$

[Figure 14: Carry generated in i th column is saved in the same i th column.]

Theorem 2: $(MCVT, B_n)$ is a commutative monoid.

Proof: As the range of MCVT is B_n so it satisfies Closure property. All other properties can be proved as follows.

For Associative Property

Claim: $Modified\ CVT(M\ odified\ CVT(a, b), c)$
 $= Modified\ CVT(a, M\ odified\ CVT(b, c))$

L.H.S

$$\begin{aligned} &Modified\ CVT(Modified\ CVT(a, b), c) \\ &= Modified\ CVT((a_n \wedge b_n, a_{n-1} \wedge b_{n-1}, \dots, a_1 \wedge b_1), (c_n, c_{n-1}, \dots, c_1)) \\ &= (a_n \wedge b_n \wedge c_n, a_{n-1} \wedge b_{n-1} \wedge c_{n-1}, \dots, a_1 \wedge b_1 \wedge c_1) \end{aligned}$$

R.H.S

$$\begin{aligned} &Modified\ CVT(a, Modified\ CVT(b, c)) \\ &= Modified\ CVT((a_n, a_{n-1}, \dots, a_1), (b_n \wedge c_n, b_{n-1} \wedge c_{n-1}, \dots, b_1 \wedge c_1)) \\ &= (a_n \wedge b_n \wedge c_n, a_{n-1} \wedge b_{n-1} \wedge c_{n-1}, \dots, a_1 \wedge b_1 \wedge c_1) \end{aligned}$$

For Existence of identity

$$\begin{aligned} &Modified\ CVT((a_n, a_{n-1}, \dots, a_1), (1, 1, \dots, 1)) \\ &= (a_n \wedge 1, a_{n-1} \wedge 1, \dots, a_1 \wedge 1) = (a_n, a_{n-1}, \dots, a_1) \end{aligned}$$

For Commutative

$$\begin{aligned} &Modified\ CVT(a, b) = (a_n \wedge b_n, a_{n-1} \wedge b_{n-1}, \dots, a_1 \wedge b_1) \\ &= (b_n \wedge a_n, b_{n-1} \wedge a_{n-1}, \dots, b_1 \wedge a_1) \\ &= Modified\ CVT(b, a) \end{aligned}$$

Hence, $(MCVT, B_n)$ is a **commutative monoid**.

In the modified CV table the decimal values in each entry are exactly half i.e. because only the LSB position 0 is not padded in this type of vectors. Thus the fractals generated by both these CVT operations are exactly same but the modified CVT has an additional advantage of getting some algebraic properties than the conventional CVT. Following section deals with the extension of both CVT as well as MCVT in higher dimensions.

VIII. ANALYTICAL AND ALGEBRAIC PROPERTIES OF MCVT

It is shown in earlier in section the fractal obtained by CVT (thus by MCVT) can also be constructed by 1-D CA rules,

which is basically a sub class of Discrete Dynamical System. This motivates us to study about the dynamical properties of CVT or MCVT. Interestingly (Z, MCVT) is a Discrete Dynamical System. Its proof is based on the following definition of dynamical system. Let us make you recalled the definition of a Dynamical system.

Dynamical system: A dynamical system is a semi-group G acting on a space M , i.e. there is a map

$$T: G \times M \rightarrow M \\ (g, x) \rightarrow T_g(x) \text{ such that } T_g * T_h = \dots \dots \dots (1)$$

If G is a group, then $(Z, MCVT)$ is called Invertible dynamical system. In another form we can formalize the condition – (1) as follows.

A dynamical system is semi-group G acting on a space M , i.e. there is a map

$$T: G \times M \rightarrow M \\ (g, x) \rightarrow T_g(x)$$

such that there exist a function f for which

$$f(T_g, T_h) = T_{f(g,h)} \dots \dots \dots (2)$$

Let us come to in our arena...

Let us come to in our arena and let us define MCVT in decimal number system as $MCVT: Z \times Z \rightarrow Z$; Here

$$G = Z, \quad M = Z \quad \text{Such that} \\ MCVT_a(x) = MCVT(a, x) \quad a \in Z, x \in Z$$

Theorem 3: $(Z, MCVT)$ is a **Discrete Dynamical System**.

Proof: $(Z, MCVT)$ is a semi-group, as it's a commutative monoid by theorem 7.2. To show $(Z, MCVT)$ is a Dynamical system we have to prove that

$$MCVT(MCVT_a, MCVT_b) = MCVT_{MCVT(a,b)}$$

Let $a, b, x \in Z$ and $a = (a_n, a_{n-1}, \dots, a_1)_2$, $b = (b_n, b_{n-1}, \dots, b_1)_2$ and $x = (x_n, x_{n-1}, \dots, x_1)_2$ Now,

$$\begin{aligned} & MCVT(MCVT_a(x), MCVT_b(x)) \\ &= MCVT((a_n \wedge x_n, a_{n-1} \wedge x_{n-1}, \dots, a_1 \wedge x_1), (b_n \wedge x_n, b_{n-1} \wedge x_{n-1}, \dots, b_1 \wedge x_1)) \\ &= ((a_n \wedge x_n) \wedge (b_n \wedge x_n), (a_{n-1} \wedge x_{n-1}) \wedge (b_{n-1} \wedge x_{n-1}), \dots, (a_1 \wedge x_1) \wedge (b_1 \wedge x_1)) \\ &= ((a_n \wedge b_n \wedge x_n), (a_{n-1} \wedge b_{n-1} \wedge x_{n-1}), \dots, (a_1 \wedge b_1 \wedge x_1)) \\ &= MCVT((a_n \wedge b_n, a_{n-1} \wedge b_{n-1}, \dots, a_1 \wedge b_1), (x_n, x_{n-1}, \dots, x_1)) \\ &= MCVT(MCVT(a, b), x) \\ &= MCVT_{MCVT(a,b)}(x) \end{aligned}$$

That is, $MCVT(MCVT_a, MCVT_b) = MCVT_{MCVT(a,b)}$ Hence, $(Z, MCVT)$ is a **Discrete Dynamical System**.

IX. EXTENSION OF CVT AND MCVT IN HIGHER DIMENSIONS

Let us define CVT (or MCVT) recursively in higher dimensional space Z^k . CVT (or MCVT) is defined as

$$CVT(or MCVT): (B_n \times B_n \times B_n \times \dots \times B_n) \rightarrow B_{n+1}, \quad \text{Where } B_n \text{ is the set of strings of length } n \text{ on } B = \{0, 1\}$$

More specifically,

$$CVT(x_1, x_2, \dots, x_k) = CVT(CVT(x_1, x_2, \dots, x_{k-1}), x_k)$$

In a similar fashion, we can write

$$MCVT(x_1, x_2, \dots, x_k) = MCVT(MCVT(x_1, x_2, \dots, x_{k-1}), x_k)$$

Where k is a positive integer for both CVT and MCVT. It is to be noted that $k = 2$ is the terminating condition for the above recursive procedures.

In particular for $k = 3$ $CVT(x_1, x_2, x_3) = CVT(CVT(x_1, x_2), x_3)$ where $CVT(x_1, x_2)$ could be evaluated as defined above in section 3.

This definition helps us to generate fractals in space, where as earlier we have got fractals in plane. On exploring these ideas we can build up fractals in n-dimensional space.

10. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

This paper presents a new transformation named as Carry Value Transformation (CVT) applied on a pair of integers. Previously we have used this CVT for Efficient Hardware design of arithmetic operations [2]. On further investigation of this transform in binary number system produces a beautiful pattern, which is found to be a fractal having dimension 1.585, same as that of Sierpinski triangle. Further CVT can be applied for the production of periodic and chaotic patterns. Interestingly, it is proved that $(Z, MCVT)$ is a Discrete Dynamical System. Further, the definition of CVT is slightly modified and its mathematical properties are highlighted where we have shown that $(MCVT,)$ is a commutative monoid. Finally, the extension of CVT and modified CVT (MCVT) are done in higher dimensions.

Authors are of firm conviction that CVT/MCVT can be used in the way the other mathematical transforms (e.g., Fourier, Discrete Cosine, Laplace, Wavelet, Cellular Automata Transforms etc.) are used; only the domains will vary from one transform to another. Further studies of algebraic and analytical properties of these transformations are highly needed for the complete exploration of this potential area. Further, authors are expecting to get a close relationship between the CVT/MCVT with many different application areas like Computational geometry, Data compression, Quad trees, Defective chessboards, Pattern Classification, Theory of Computation, and Analysis of Cellular Automata Rules etc. Exploring all these will be our immediate future research directions.

X. REFERENCES:

- 1) B.B. Mandelbrot, 1982 The fractal geometry of nature. New York,
- 2) P.P. Choudhury, S. Sahoo, M. Chakraborty, 2008 Implementation of Basic Arithmetic Operations Using Cellular Automaton, ICIT-08, *IEEE CS Press*, 11th International Conference on Information Technology, pp 79-80,
- 3) J. L. Hennessy and D. A Patterson, 1996 Computer Architecture: A Quantitative Approach" (2nd edition), Morgan Kaufmann, San Francisco.
- 4) M. Barnsley, 1988 *Fractals Everywhere*, Academic Press Inc. (London) Ltd., San Diego.
- 5) H. O. Pietgen, H. Jurgeens, D. Saupe, 1992 *Chaos and Fractals New Frontiers of Science*, ISBN 3-540-97903-4, Springer Verlag.
- 6) P. Prusinkiewicz, A. Lindenmayer, 1996 *Algorithmic Beauty of Plants*, ISBN: 0387946764.
- 7) S. Wolfram, 1994, *Cellular Automata*, Wesley Publishing Company, ISBN 0-201-62664-0.
- 8) S. Wolfram, Statistical Mechanics of Cellular Automata, Reviews of Modern Physics, vol. 55, pp. 601-644, 1983.
- 9) C. Bandt, P. Gummelt, 1997 *Fractal Penrose Tilings. I. Construction and matching rules*. Aequationes Mathematicae 53, pp. 295-307.
- 10) A. K. Ghosh, P. P. Choudhury, A. Basuray, 2007. *Chaotic Fractals with Multi valued Logic in Cellular Automata*, CISSE 2006, IEEE, University of Bridgeport, Innovations and Advanced Techniques in Computer and Information Sciences and Engineering, pp.77-82.
- 11) A. K. Ghosh, P. P. Choudhury, R. Choudhury, 2003 *Production of fractals by various means and measuring their dimensions with probable explanation*, *Laser Horizon*, Journal of Laser Science and Technology Centre (LASTEC), vol. 6/No. 2, pp. 50-59.
- 12) A. Bogomolny, *Cut The Knot! An interactive column using Java applets* <http://www.cut-the-knot.org/ctk/Sierpinski.shtml>
- 13) P. P. Choudhury, S. Sahoo, B. K Nayak and Sk. S. Hassan, 2009 *Carry Value Transformation: Application in Fractal Formation*, 2009 IEEE International Advanced Computing Conference (IACC 2009),

Papers, Addison-

Research on Automatic Fuzzy Ontology Generation for World Wide Web

GJCST Classification
H.5.3, H.5.4

K.Srikanth¹, J. Anitha², Ch. Heymaraju³, A.Harish⁴, B.Sasikanth⁵

Abstract-Ontology is an effective conceptualism commonly used for the World Wide Web. Fuzzy logic can be incorporated to ontology to represent uncertainty information. To tackle this problem, this paper proposes the FOGA (Fuzzy Ontology Generation framework) for automatic generation of fuzzy ontology on uncertainty information. The FOGA framework comprises the following components: Fuzzy Formal Concept Analysis, Concept Hierarchy Generation, and Fuzzy Ontology Generation. We also discuss approximating reasoning for incremental enrichment of the ontology with new upcoming data. Finally, a fuzzy-based technique for integrating other attributes of database to the ontology is proposed. **Index Terms**-Intelligent Web services and World Wide Web, ontology design, uncertainty, "fuzzy," probabilistic, knowledge representation formalisms and methods, concept learning.

Keywords - Ontology Generation, Fuzzy Ontology.

I. INTRODUCTION

Ontology is a conceptualization of a domain into a human understandable, machine-readable format consisting of entities, attributes, relationships, and axioms [1]. It is used as a standard knowledge representation for the World Wide Web [2]. However, the conceptual formalism supported by typical ontology may not be sufficient to represent uncertainty information commonly found in many application domains due to the lack of clear-cut boundaries between concepts of the domains. For example, a document can be very relevant, relevant, or irrelevant to a research area. In addition, keywords extracted from scientific publications can be used to infer the corresponding research areas. However, it is inappropriate to treat all keywords equally as some keywords may be more significant than others. To tackle this type of problems, one possible solution is to incorporate fuzzy logic [3] into ontology to handle uncertainty data. Traditionally, fuzzy ontology is generated and used in text retrieval [4] and search engines [5], in which membership values are used to evaluate the similarities between the concepts in a concept hierarchy. However, manual generation of fuzzy ontology from a predefined concept hierarchy is a difficult and tedious task that often requires expert interpretation. So, automatic generation of concept hierarchy and fuzzy ontology from uncertainty data of a domain is highly desirable. In this paper, we propose a framework known as FOGA (Fuzzy Ontology Generation framework) that can automatically generate a fuzzy ontology from uncertainty data based on

Formal Concept Analysis (FCA) [6] theory. The generated fuzzy ontology is mapped to a World Wide representation in OWL (Web Ontology Language) [7]. The rest of this paper is organized as follows: Section 2 discusses related work on ontology generation and FCA. Section 3 gives some basic definitions and operators of the fuzzy theory. The FOGA framework is presented in Section 4. Section 5 discusses the approximating reasoning technique to incrementally furnish the generated ontology with new instance. The problem of integrating extra attributes in database to the ontology is given in Section 6. Performance evaluation of the proposed FOGA framework is given in Section 7. Finally, Section 8 concludes the paper.

II. RELATED WORK

2.1 Ontology Generation

Although editing tools [8], [9] have been developed to help users to create and edit ontology, it is a troublesome task to manually derive ontology from data. Typically, ontology can be generated from various data types such as textual data [10]. Compared to other types of data, ontology generation from textual data has attracted the most attention. Among techniques used for processing textual data, clustering is one of the most effective techniques for ontology learning. Conceptual clustering techniques such as COBWEB and CLASSIT are powerful clustering techniques that can conceptualize clusters for ontology generation. We have created DESK to fill this gap, using the Programming By Demonstration paradigm [2], [5], [6] to allow the page modification by a non-expert user. In programming by demonstration the system infers procedural information from examples of what the user wants to achieve. The programming by demonstration paradigm has an intrinsic ambiguity because general information has to be derived from particular cases provided by the user. To solve such ambiguity some strategies have been used, such as monitoring all user interaction (vs. watching only the initial and the final state), using multiple examples (e.g. negative examples), or interactively asking the user to help or decide. The extraction of structured information, like the difference and context model used by DESK, from a semi-structured document (HTML code) is very similar to the way wrappers operate [8]. Wrappers provide a uniform access to the information stored in heterogeneous repositories like data bases, files and so forth.

About^{1,2,3,4&5} -GIT, GITAM University, Visakhapatnam
¹srikanthwizard@gmail.com, ²anithanv28@gmail.com, ³heymanraju@gmail.com, ⁴saihari08@gmail.com, ⁵battula.sasi@gmail.com

III. FUZZY THEORY

In this section, we review some fundamental knowledge of fuzzy theory [3].

Definition 1 (Fuzzy Set). A fuzzy set A on a domain U , is defined by a membership function μ from U to $[0,1]$, i.e., each item in A has a membership value given by μ . We denote $\varphi(S)$ as a fuzzy set generated from a traditional set of items S . Each item in S has a membership value in $[0, 1]$. S can also be called as a crisp set.

Definition 2 (Fuzzy Relation). A fuzzy set A on a domain $G \times M$, where G and M are two crisp sets is a fuzzy relation on G, M .

Definition 3 (Fuzzy Sets Intersection). The intersection of fuzzy sets A and B , denoted as $A \cap B$, is defined by $\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x))$.

Definition 4 (Fuzzy Sets Union). The intersection of fuzzy sets A and B , denoted as $A \cup B$, is defined by $\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x))$.

Definition 5 (Fuzzy Set Cardinality). Let S_f be a fuzzy set on the domain U . The cardinality of S_f is defined as $|S_f| = \sum_{x \in U} \mu(x)$

Where $\mu(x)$ is the membership of x in S_f .

Definition 6 (Fuzzy Sets Similarity). The similarity between two fuzzy sets A and B is defined as

$$E(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

Definition 7 (Fuzzy Sets Subsethood). The subsethood of a fuzzy set A of a conceptual cluster B is calculated as

$$\text{Subsethood}(A,B) = \frac{|A \cap B|}{|B|}$$

Definition 8 (Fuzzy Set Max-min Composition). Let $P(X,Y)$ be a fuzzy relation on X, Y and $P(Y,Z)$ be a fuzzy relation on Y, Z . The max-min composition of $P(X,Y)$ and $Q(Y,Z)$, $P \cdot Q$, is defined by:

$$\mu_{P \cdot Q}(X,Y) = \max \min(\mu_p(X,Y), \mu_q(Y,Z)), \varphi_x \in X, y \in Y.$$

The max-min composition indicates the strength of relation between the element of X and Z .

IV. THE FOGA FRAMEWORK

Fig. 1 shows the proposed FOGA (Fuzzy Ontology Generation framework), which consists of the following components.

4.1 Fuzzy Formal Concept Analysis

The Fuzzy Formal Concept Analysis incorporates fuzzy logic into Formal Concept Analysis to represent vague Information.

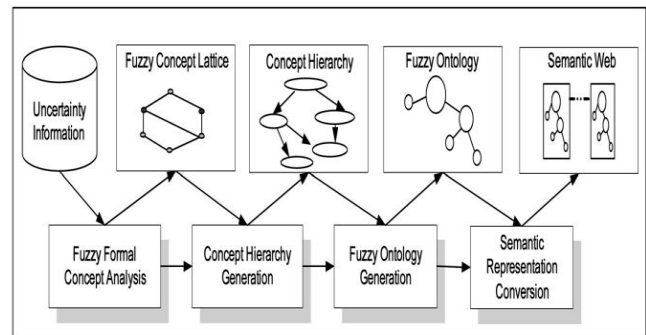


Fig 1. The FOGA framework.

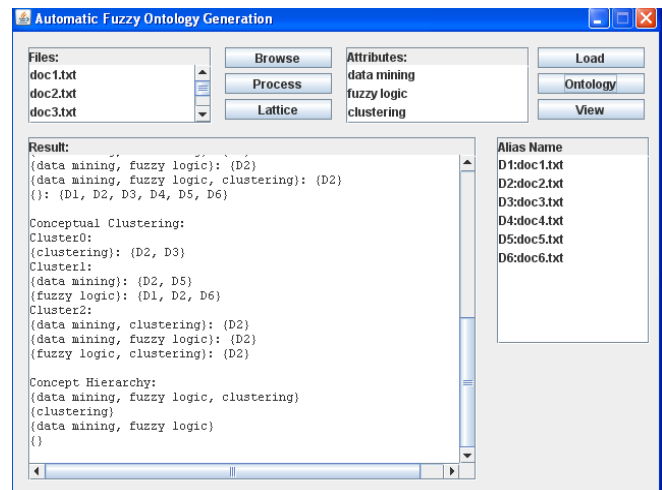


Fig.2.Ontology Generation Frame Work

TABLE 1

A Cross-Table of a Fuzzy Formal Context

	Data Mining	Clustering	Fuzzy Logic
D1	0.8	0.12	0.61
D2	0.9	0.85	0.13
D3	0.1	0.14	0.87

TABLE 2

Fuzzy Formal Context in Table 1 with an α cut $\alpha = 0.5$

	Data Mining	Clustering	Fuzzy Logic
D1	0.8	-	0.61
D2	0.9	0.85	-
D3	-	-	0.87

V. DESK AS AN AUTHORIZING TOOL

After the proposition extraction step, we have a set of propositions as fuzzy rules. The next step is to use the generated rules for reasoning new data. For example, assume that we have a fuzzy rule IF x is A THEN y is B , where A and B are fuzzy sets. Then, if we have a new proposition x is A , we need to find what conclusion we can get about y .

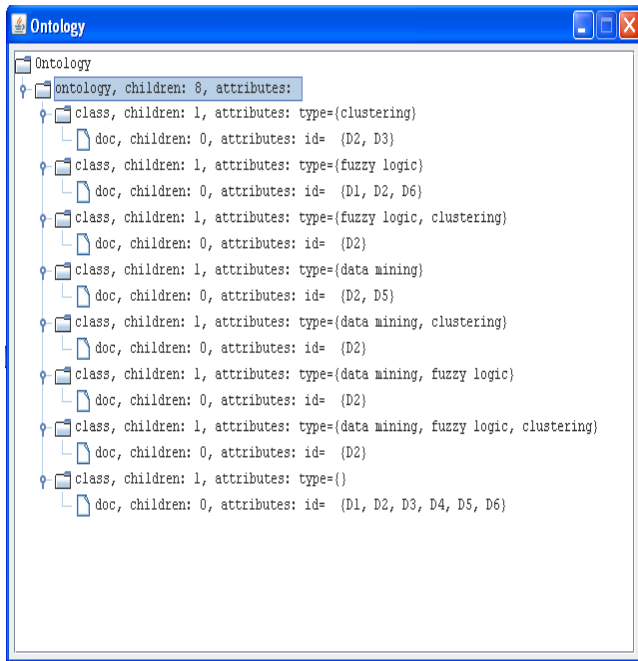


Fig.3. Ontology Tree Structure

TABLE 3

A Fuzzy Formal Context Having Cross Relation with the Fuzzy Formal Context in Table 2

	D1	D2	D3
Author1	1.0	-	1.0
Author2	0.5	1.0	-

The cross relation represents an intercontext relation that probably occurs between the fuzzy formal contexts when the set of objects of a context is regarded as the set of attributes of an other contexts. For example, the context represented by the cross table shown in Table 3 has cross relation with the context in Table 2, while the documents are used as attributes of the authors. The membership value of 1.0 implies that the author is the first author of the document, while 0.5 implies that the author is the second author.

VI. PERFORMANCE EVALUATION

6.1 Generating Ontology from Citation Database To evaluate the proposed FOGA framework for ontology

generation, we have collected a set of 1,400 scientific documents on the research area Information Retrieval published in 1987-1997 from the Institute for Scientific Information's (ISI) Web site [52]. The downloaded documents are preprocessed to extract related information such as the title, authors, citation keywords, and other citation information. The extracted information is then stored as a citation database. First, we construct a fuzzy formal context $K_f = \{G, M, I\}$, with G as the set of documents and M as the set of citation keywords. The membership value of a document D on a citation keyword CK in K_f is computed as $\mu(d, C_k) = n1/n2$, where $n1$ is the number of documents that cite D and contain CK and $n2$ is the number of documents that cite D . This formula is based on the premise that the more frequent a keyword occurs in the citing paper, the more important the keyword is in the cited paper.

VII. EVALUATION USING RECALL, PRECISION, AND F-MEASURE

We have classified manually the documents downloaded from ISI into classes based on their research themes. These classes are used as a benchmark to evaluate the clustering results in terms of recall, precision, and F-measure. As discussed earlier, we extract citation keywords of documents as their attributes. Since these attributes are descriptors for the generated clusters, if more keywords are extracted and used, the more meaningful the cluster descriptors are constructed. To verify this, we vary the number of keywords N extracted from documents from 2 to 10, and the similarity threshold T_s from 0.2 to 0.9 when performing conceptual clustering. The measured precision, recall and F-measure are presented in Table 7, respectively.

TABLE 7
Performance Results Using Recall Measurement

	$T_s=0.2$	$T_s=0.3$	$T_s=0.4$	$T_s=0.5$	$T_s=0.6$	$T_s=0.7$	$T_s=0.8$	$T_s=0.9$
N=2	0.99	0.99	0.99	0.99	0.99	0.98	0.98	0.98
N=3	0.99	0.99	0.99	0.99	0.98	0.98	0.97	0.97
N=4	0.98	0.98	0.97	0.97	0.94	0.95	0.94	0.94
N=5	0.89	0.87	0.87	0.88	0.87	0.89	0.89	0.89
N=6	0.8	0.81	0.83	0.83	0.83	0.85	0.85	0.85
N=7	0.81	0.8	0.82	0.82	0.83	0.84	0.86	0.86
N=8	0.79	0.79	0.81	0.82	0.82	0.84	0.85	0.85
N=9	0.76	0.77	0.8	0.8	0.81	0.83	0.84	0.84
N=10	0.73	0.75	0.78	0.78	0.79	0.81	0.83	0.83

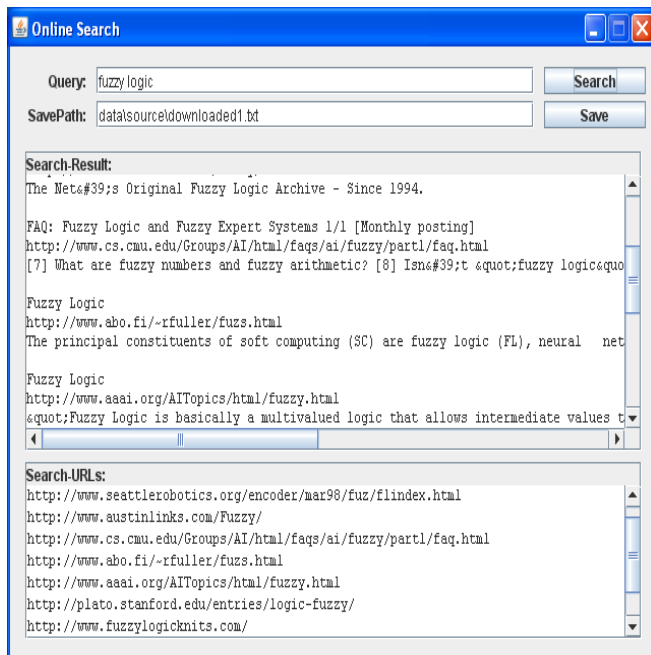


Fig 4. Performance Result Using World

Wide Web Mining Techniques

VIII) CONCLUSIONS

In this paper, we have proposed the FOGA framework for fuzzy ontology generation on uncertainty information. FOGA consists of the following steps: Fuzzy FormalConcept Analysis, Fuzzy Conceptual Clustering, FuzzyOntology Generation, and World Wide Web Representation Conversion. In addition, we have also proposed an approximating reasoning technique that allows the generated fuzzy ontology to be incrementally furnished with new instances. Finally, we have also proposed a technique to integrate extra attributes in a database to the ontology. Our authoring tool provides automatic support for the customization of dynamic web documents based on comparing the pages generated by the system with a modified version provided by the end-user. DESK is based on PEGASUS, a system used to represent the World Wide Web information structured by models that allow a clear separation between contents and presentation. DESK uses domain information stored in PEGASUS and presentation models for finding the context of changes made by user. Our authoring tool also determines whether the user is enabled to do these modifications depending on a user model. With DESK the user only needs to take care of editing HTML pages using any standard HTML editing tool such as PageMaker or Netscape Composer.

IX) REFERENCES

- 1) N. Guarino and P. Giaretta, *Ontologies and Knowledge Bases: Towards a Terminological Clarification. Toward Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing*. Amsterdam: IOS Press, 1995.

- 2) Castells, P., Szekely, P.: *Presentation Models by Example*. En: Duke, D.J., Puerta A. (eds.): *Design, Specification and Verification of Interactive Systems '99*. Springer-Verlag, 1999, pp. 100-116.
- 3) L.A. Zadeh, *Fuzzy Logic and Approximate Reasoning*, Synthese, vol. 30, pp. 407-428, 1975.
- 4) Castells, P. Macías, J.A.: *An Adaptive Hypermedia Presentation Modeling System for Custom Knowledge Representations*. *Proceedings of the World Conference on the WWW and Internet*
- 5) D.H. Widyantoro and J. Yen, *A Fuzzy Ontology-Based Abstract Search Engine and Its User Studies*, Proc. 10th IEEE Int'l Conf. Fuzzy Systems, pp. 1291-1294, 2001.
- 6) B. Ganter and R. Wille, *Formal Concept Analysis: Mathematical Foundations*. Springer, 1999. [7] W3C, *Web Ontology Language Overview*, "http://www.w3.org/TR/owl-features/", 2006.
- 7) *OntoWeb deliverable 1.3*, http://www.ontoweb.org/, 2006.
- 8) A. Gomez-Perez, O. Corcho, and M. Fernandez-Lopez, *Ontological Engineering: With Examples from the Areas of Knowledge Management, e-Commerce and the Semantic Web (Advanced Information and Knowledge Processing)*. Springer, 2004.
- 9) Macías, J. A., Castells, P.: *Adaptive Hypermedia Presentation Modeling for Domain Ontologies*. To appear in proceedings of 10th International Conference on Human-Computer Interaction (HCII '2001). New Orleans, Louisiana, 2001.
- 10) Murray, T.: *Authoring Knowledge Based Tutors: Tools for Content, Instructional Strategy, Student Model, and Interface Design*. *Journal of the Learning Sciences* 7, 1, 1998, 5-64.
- 11) Muslea, I.: *Extraction Patterns for Information Extraction Tasks: A Survey*. *Proceedings of AAAI Workshop on Machine Learning for Information Extraction*. Orlando, Florida, July 1999.
- 12) Vassileva, J.: *Dynamic Course Generation on the WWW*. *Actas 8th World Conference on Artificial Intelligence in Education (AIED'97)*. Kobe, Japan, 1997, 498-505.
- 13) Weber, G. and Specht, M.: *User modeling and Adaptive Navigation Support in WWW-based Tutoring Systems*. *Proceedings 6th International Conference on User Modeling (UM97)*. Sardinia, Italy, 1997.

Global Journals Inc. (US) Guidelines Handbook 2011

www.GlobalJournals.org

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

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

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

- AICCT title will be awarded to the person/institution after approval of Editor-in-Chief 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.
- A professional email address will be allotted with free 1GB email space.
- AICHSS will be authorized to receive e-Journal GJHSS for lifetime.



Auxiliary Memberships

ANNUAL MEMBER

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

PAPER PUBLICATION

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



Process of submission of Research Paper

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

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

The Author can submit the paper either online or offline. The authors should prefer online submission.

Online Submission: There are three ways to submit your paper:

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

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

2. ETHICAL GUIDELINES

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

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

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

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

- 1) Substantial contributions to conception and acquisition of data, analysis and interpretation of the findings.
- 2) Drafting the paper and revising it critically regarding important academic content.
- 3) Final approval of the version of the paper to be published.

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

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

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

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

Please mention proper reference and appropriate acknowledgements wherever expected.

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

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

3. SUBMISSION OF MANUSCRIPTS

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

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



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

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

4. MANUSCRIPT'S CATEGORY

Based on potential and nature, the manuscript can be categorized under the following heads: Original research paper: Such papers are reports of high-level significant original research work.

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

Research articles: These are handled with small investigation and applications

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

5. STRUCTURE AND FORMAT OF MANUSCRIPT

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

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

- (a) Title should be relevant and commensurate with the theme of the paper.
- (b) A brief Summary, "Abstract" (less than 150 words) containing the major results and conclusions.
- (c) Up to ten keywords, that precisely identifies the paper's subject, purpose, and focus.
- (d) An Introduction, giving necessary background excluding subheadings; objectives must be clearly declared.
- (e) Resources and techniques with sufficient complete experimental details (wherever possible by reference) to permit repetition; sources of information must be given and numerical methods must be specified by reference, unless non-standard.
- (f) Results should be presented concisely, by well-designed tables and/or figures; the same data may not be used in both; suitable statistical data should be given. All data must be obtained with attention to numerical detail in the planning stage. As reproduced design has been recognized to be important to experiments for a considerable time, the Editor has decided that any paper that appears not to have adequate numerical treatments of the data will be returned un-refereed;
- (g) Discussion should cover the implications and consequences, not just recapitulating the results; conclusions should be summarizing.
- (h) Brief Acknowledgements.
- (i) References in the proper form.

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

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

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 $1.4 \times 10^{-3} \text{ m}^3$, or 4 mm somewhat than $4 \times 10^{-3} \text{ m}$. Chemical formula and solutions must identify the form used, e.g. anhydrous or hydrated, and the concentration must be in clearly defined units. Common species names should be followed by underlines at the first mention. For following use the generic name should be constricted to a single letter, if it is clear.

Structure

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

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

Abstract, used in Original Papers and Reviews:

Optimizing Abstract for Search Engines

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

Key Words

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

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

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

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

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

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

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



Acknowledgements: Please make these as concise as possible.

References

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

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

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

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

Tables, Figures and Figure Legends

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

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

Preparation of Electronic Figures for Publication

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

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

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

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

6. AFTER ACCEPTANCE

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



6.1 Proof Corrections

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

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

(Free of charge) from the following website:

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

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

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

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

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

6.3 Author Services

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

6.4 Author Material Archive Policy

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

6.5 Offprint and Extra Copies

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

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 through which your research is going to be in print to the rest of the crowd. Care should be taken to categorize your thoughts well and present them in a logical and neat manner. A good quality research paper format is essential because it serves to highlight your research paper and bring to light all necessary aspects in your research.

INFORMAL GUIDELINES OF RESEARCH PAPER WRITING

Key points to remember:

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

Final Points:

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

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

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

General style:

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

To make a paper clear

· Adhere to recommended page limits

Mistakes to evade

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



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

In every sections of your document

- Use standard writing style including articles ("a", "the," etc.)
- Keep on paying attention on the research topic of the paper
- Use paragraphs to split each significant point (excluding for the abstract)
- Align the primary line of each section
- Present your points in sound order
- Use present tense to report well accepted
- Use past tense to describe specific results
- Shun familiar wording, don't address the reviewer directly, and don't use slang, slang language, or superlatives
- Shun use of extra pictures - include only those figures essential to presenting results

Title Page:

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

Abstract:

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

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

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



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

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

Approach:

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

Introduction:

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

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

Approach:

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

Procedures (Methods and Materials):

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



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

Materials:

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

Methods:

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

Approach:

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

What to keep away from

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

Results:

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

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

Content

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

What to stay away from

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



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

Approach

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

Figures and tables

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

Discussion:

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

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

Approach:

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

ADMINISTRATION RULES LISTED BEFORE SUBMITTING YOUR RESEARCH PAPER TO GLOBAL JOURNALS INC. (US)

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

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



- The **major constraint** is that you must independently make all content, tables, graphs, and facts that are offered in the paper. You must write each part of the paper wholly on your own. The Peer-reviewers need to identify your own perceptive of the concepts in your own terms. NEVER extract straight from any foundation, and never rephrase someone else's analysis.
- Do not give permission to anyone else to "PROOFREAD" your manuscript.
- **Methods to avoid Plagiarism is applied by us on every paper, if found guilty, you will be blacklisted by all of our collaborated research groups, your institution will be informed for this and strict legal actions will be taken immediately.)**
- To guard yourself and others from possible illegal use please do not permit anyone right to use to your paper and files.



CRITERION FOR GRADING A RESEARCH PAPER (COMPILATION)
BY GLOBAL JOURNALS INC. (US)

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

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

Index

A

Access · 8, 2, 52, 55, 62
advancement · 38
AES · 8, 7, 8, 9, 10, 11
akeyword · 110
algorithms · 13, 30, 31, 32, 47, 48, 49, 58, 59, 61, 63, 69, 70, 87, 95, 96, 98
alleviating · 76
analysis · 5, 37, 38, 40, 41, 42, 43, 45, 47, 50, 51, 59, 74, 78, 86, 87, 90, 95, 96
approximating · 31, 108, 111
architecture · 3, 4, 14, 32, 35, 62, 63
Assessing · 42, 44
attacks · 32, 33, 35, 58
attainable · 4
authority · 4, 33, 35
Automata · 98, 99, 100, 102, 103, 107

B

Bayes · 46, 57, 58, 59, 60, 61
Beauty · 18, 107
biometric · 69, 72, 73
Bluetooth · 51, 52, 53, 54, 56

C

capillaries · 57
Carry · 8, 98, 99, 100, 105, 107
Cellular · 98, 99, 100, 102, 103, 107
characterization · 59
checksum · 21
Children · 82, 84
class · 5, 12, 14, 15, 16, 46, 47, 48, 64, 69, 79, 82, 83, 84, 85, 86, 87, 89, 94, 106
Classification · 46, 49, 50, 58, 107
classifier · 46, 48, 58, 92
Code · 14, 51, 82
cohesive · 84
Collaborative · 35
Commutativity · 95, 105
comparison · 4, 10, 19, 23, 25, 43, 46, 47, 48, 56, 59, 68, 76, 93

complicated · 61, 86
complimentary · 56
Complimentary · 51
compute · 3, 4, 20, 47, 71, 83, 84, 93, 94, 95
concentrating · 94
conceptual · 39, 44, 45, 78, 79, 108, 109, 110
connected · 4, 15, 35, 37, 71
consequences · 57, 75
conservation · 54
consoles · 7
control · 2, 3, 4, 5, 6, 19, 21, 26, 52, 53, 55, 56, 62, 63, 75, 76, 82
Control · 8, 2, 19, 29, 52, 62, 92
conventional · 18, 90, 105
conversion · 12, 14, 15, 16
customary · 37
customer · 37, 39, 42, 43, 44, 45, 69, 94

D

data · 3, 4, 7, 11, 12, 13, 14, 15, 18, 19, 20, 21, 22, 23, 25, 26, 28, 32, 33, 34, 35, 38, 40, 46, 47, 48, 49, 51, 52, 53, 54, 55, 57, 58, 59, 61, 62, 63, 64, 68, 70, 71, 75, 76, 77, 78, 79, 80, 93, 94, 95, 96, 108, 110
Data · 8, 21, 25, 35, 36, 42, 43, 47, 52, 56, 61, 63, 79, 93, 96, 97, 107
decaying · 71
Declaration · 16
Decryption · 7, 9, 96
Definition · 14, 16, 95, 99, 109
Defuzzification · 75, 78
denote · 57, 98, 99, 102, 109
depending · 2, 38, 53, 73, 111
depicting · 42
Depth · 82, 83
DES · 64
design · 2, 5, 13, 32, 35, 38, 39, 40, 42, 43, 44, 51, 82, 83, 84, 85, 86, 98, 107, 108
designed · 4, 5, 7, 13, 14, 15, 21, 38, 54, 55, 63, 82, 83, 98
detection · 54, 69, 70, 74, 87, 88, 89, 90, 91, 92
development · 2, 12, 13, 18, 19, 29, 35, 38, 40, 57, 75, 76, 77, 78, 79, 82, 84, 98
Direct-Sequence · 51
Discrete · 31, 97, 98, 106, 107
dissemination · 20
Distributed · 12, 13, 52, 94

documentation · 14, IV
downlink · 62, 63
DWRR · 62, 63, 64, 65, 66, 68
Dynamical · 98, 106, 107

E

Eavesdropping · 32, 33, 34
Edge · 8, 87, 91, 92
effort · 63, 64, 75, 76, 77, 78, 79, 80, 81
effortlessly · 76
E-Governance · 8, 27
embedded · 7, 14, 21
Embedded · 8, 7, 10, 11
emphasize · 98
Employment · 29
emulator · 21
Encryption · 7, 11, 33, 93, 95, 96
enthusiastic · 27
entropy · 87, 94
estimation · 50, 75, 76, 77, 78, 79, 80, 81
Ethernet · 52
evaluated · 2, 3, 4, 46, 47, 48, 57, 59, 82, 88, 95, 106
exceeding · 8, 10
exhaustive · 90
Experiments · 39, 46, 48
Extensible · 12
extensions · 52, 53
extracting · 12, 14, 39, 70, 71
extraction · 12, 13, 14, 15, 69, 71, 72, 108, 110

F

Face · 69, 70, 74
Facial · 69
features · 69
File · 12, 13, 16
FOAF · 2, 3, 4
fractal · 98, 99, 100, 101, 102, 103, 104, 105, 107
framework · 12, 13, 14, 15, 36, 39, 43, 45, 75, 77, 108, 109, 110, 111
Francis · 56, 68
Frequency · 51, 91
fuzzy · 2, 46, 50, 69, 75, 76, 77, 78, 80, 81, 87, 88, 89, 90, 91, 92, 108, 109, 110, 111
Fuzzy · 8, 75, 76, 77, 78, 79, 80, 81, 87, 89, 90, 92, 108, 109, 110, 111

G

Generation · 8, 94, 97, 108, 109, 111

GPS · 8, 7, 11, 19, 20, 21, 22, 23, 24, 25, 26, 63
Grayhole · 32, 33
group · 3, 4, 5, 12, 31, 32, 38, 44, 76, 87, 106

H

Hadoop · 12, 13, 18
Hbase · 12, 13
hierarchy · 14, 43, 44, 83, 84, 85, 101, 108
Higher · 85
highway · 12
human · 12, 13, 37, 46, 57, 69, 71, 75, 76, 108

I

IASTED · 74
ICT · 8, 27, 28, 29, 30
identity · 2, 33, 69, 105
Impersonation · 32, 33
impersonator's · 33
implementation · 2, 4, 5, 12, 17, 52, 85
inference · 13, 78, 87, 88, 89
Information · 4, 2, 5, 18, 27, 29, 30, 40, 41, 42, 43, 44, 45, 49, 50, 54, 68, 80, 81, 92, 96, 97, 107, 109, 110, 111
infrastructure · 32, 51
Infrastructure · 29, 32, 33, 35
inorder · 40, 42
Integration · 52
intend · 80
intersection · 94, 95, 109
introduced · 19, 46, 52, 62, 70, 75, 82, 88
itemsets · 93, 94, 95, 96

J

Jharkhand · 8, 27, 28, 29, 30

K

keying · 51
Keying · 51
K-Nearest · 8, 46, 48
KNN · 46, 47, 48, 49
Knowledge · 8, 2, 5, 6, 12, 57, 69, 80, 96, 111
KStar · 57, 59, 60, 61

L

Language · 12, 14, 18, 108, 111

languages · 13, 14
latitude · 19, 20, 21, 22, 23, 24
launched · 19, 20
located · 19, 21, 69, 71, 101
logic · 75, 76, 77, 78, 87, 90, 92, 108, 109
Logic · 8, 75, 76, 77, 78, 79, 80, 81, 87, 92, 107, 111
luminance · 70, 71

M

Magnitude · 78
malicious · 32, 33
management · 2, 5, 12, 35, 38, 45, 52, 55, 61, 62, 75, 76, 80, 82
Markup · 12
matching · 69, 107
Mathematics · 4, 5, 31, 50, 98, 99
maximum · 20
mechanisms · 11, 35, 55, 57, 63, 64, 75
Metadata · 8, 12, 14, 18
Method · 3, 17, 49, 50, 72, 82, 83, 84, 85, 92, 95, 96
metrics · 6, 43, 44, 76, 82, 83, 84, 85, 86
migraine · 58
mining · 2, 3, 58, 59, 61, 93, 94, 95
Mining · 8, 4, 6, 93, 94, 96, 97, 111
MKNN · 46, 47, 48, 49
Mobile · 8, 7, 11, 12, 26, 32, 34, 36, 51, 55, 56
models · 1, 20, 35, 43, 59, 75, 76, 77, 78, 79, 81, 111
Modified · 46, 48, 92, 105
Monk's · 47, 48
multidimensional · 38, 46
multiplication · 8, 9, 10, 47, 48

N

Naive · 57, 58, 59, 60, 61
Neighbor · 8, 4, 46, 48, 50
neighboring · 4, 69, 71
Network · 6, 30, 32, 33, 35, 36, 50, 51, 56, 64, 68
Number · 5, 7, 10, 16, 69, 72, 79, 82, 84, 92

O

Offset · 51
Ontology · 8, 12, 13, 15, 17, 18, 108, 109, 110, 111
optimized · 11, 56
organogram · 3, 4
Orthogonal · 51

P

parameters · 14, 55, 58, 62, 64, 69, 75, 77, 78, 79, 82, 83, 85, 86, 87
Parsing · 12
patterns · 93, 98, 99, 100, 101, 102, 104, 107
Phase · 51, 95, 96
phase-shift · 51
physiological · 69
pipeline · 29
plotting · 22, 23
precision · 20, 76, 110
Privacy · 36, 42, 93, 96, 97
Process · 9
processor · 5, 7
Processors · 8, 7, 11
promising · 12, 77
published · 14, 55, 98, 110

Q

Quadrature · 51
quality · 37, 38, 39, 40, 42, 43, 44, 45, 46, 53, 54, 61, 62, 64, 78, 82, 87, 90
quantities · 47, 93, 94
questionnaire · 39, 40

R

Random · 93, 96
randomly · 48, 77
receiver · 19, 20, 21, 22, 23, 24, 25, 26, 62
recognition · 46, 49, 69, 90
refinement · 82, 86
refinements · 46, 86
relational · 13, 14
represent · 4, 12, 14, 57, 70, 71, 76, 84, 87, 108, 109, 111
Repudiation · 32
REQUESTs · 34
Requisition · 4
Response · 32, 82, 85
Rijndael · 7
robustness · 46, 48, 69, 70, 87
Rushing · 32, 34

S

Satisfaction · 29, 37, 41, 43, 44
sBlackhole · 32
scenarios · 11, 64

searching · 12, 37
security · 2, 4, 5, 7, 11, 32, 33, 34, 35, 38, 42, 51, 53, 55, 62, 69, 73, 93, 94, 95
Security · 8, 2, 4, 6, 11, 32, 33, 35, 36, 55, 56, 62, 93, 94, 96, 97
semantic · 2, 3, 12, 14
semihonest · 94, 95
sentence · 19, 20, 21, 22, 23, 24, 25, 26
separating · 5, 47, 71
shape · 57, 69, 70, 71, 87, 98
simulation · 19, 21, 23, 64, 90
simulator · 19, 21
smoothing · 47, 87
software · 12, 13, 16, 18, 19, 21, 26, 27, 37, 75, 76, 77, 78, 80, 81, 82, 83, 95
specialists · 1, 61
stations · 19, 55, 62, 64
steatorrhoeic · 57
strategy · 3, 12, 63
strength · 87, 88, 90, 109
Student · 37, 39, 79, 111
substitution · 7
summarizes · 15, 38
switches · 64
system · 2, 3, 4, 5, 12, 19, 20, 22, 25, 26, 27, 32, 35, 36, 37, 46, 55, 57, 58, 69, 70, 71, 73, 82, 84, 85, 87, 88, 89, 90, 93, 98, 99, 102, 104, 106, 107, 108, 111
System · 6, 12, 13, 19, 26, 29, 30, 31, 33, 44, 45, 52, 84, 96, 97, 98, 99, 100, 102, 103, 106, 107, 111
systems · 2, 6, 7, 11, 13, 19, 35, 42, 43, 46, 69, 71, 75, 80, 82, 92, 98, 99

T

Task · 37, 40, 41
technique · 2, 3, 5, 31, 45, 47, 63, 77, 78, 87, 90, 99, 108, 111
techniques · 2, 3, 12, 50, 75, 76, 77, 79, 82, 87, 90, 93, 102, 108
Tidwell · 18
Tilling · 98, 104

tool · 13, 14, 15, 16, 37, 39, 43, 57, 58, 59, 75, 78, 82, 83, 86, 87, 93, 95, 99, 104, 110, 111
tooth · 69, 71
toothbrushes · 58
tracking · 19, 20, 21, 25, 26, 69, 95
transformation · 7, 13, 15, 71, 98, 107
Transformation · 8, 7, 8, 9, 10, 98, 99, 105, 107
transmission · 20, 32, 35, 52, 55, 62, 63
transparency · 27, 28, 29, 77
Tree · 57, 58, 59, 60, 61, 82, 83, 110
tremendous · 27
Tutoring · 111

U

unauthorized · 32, 34, 35, 73
uncountable · 99, 100, 105
uplink · 62, 63
usability · 2, 5, 38, 40, 42, 43, 44, 45, 56

V

validation · 45, 47, 57, 59, 75
Validation · 46, 77
Value · 8, 64, 98, 99, 105, 107
Versicolour · 47

W

web · 2, 3, 4, 6, 12, 37, 38, 40, 42, 43, 44, 45, 111
WEBQUAL · 37, 40, 42
Website · 7, 8, 6, 37, 39, 40, 42, 45
Weighted · 46, 47, 63, 66, 82, 83
WEKA · 57, 58, 59
width · 47, 70
wireless · 4, 32, 51, 53, 54, 55, 62
Wormhole · 32, 33



save our planet



Global Journal of Computer Science and Technology

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



ISSN 9754350