

GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY

discovering thoughts and inventing future

Volume 10 Issue 7 Version 1.0

Online ISSN: 0975-4172

Print ISSN: 0975-4350

highlights

Extraction of Polygon Regions

HE-Advisor

Design and Implementation of RADIUS

VANET Parameters and Applications

16 Technology
Reforming
Ideas

September 2010



Global Journal of Computer Science and Technology



Global Journal of Computer Science and Technology

Volume 10 Issue 7 (Ver. 1.0)

Global Academy of Research and Development

© 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.”

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.

License No.: 42125/022010/1186
Registration No.: 430374
Import-Export Code: 1109007027

Global Academy of Research and Development

Publisher's correspondence office

Global Journals, Headquarters Corporate Office,
United States

Offset Typesetting

Global Journals, City Center Office,
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: investers@globaljournals.org
Technical Support: technology@globaljournals.org
Media & Releases: media@globaljournals.org

Pricing (Including by Air Parcel Charges):

For Authors:

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

Yearly Subscription (Personal & Institutional):

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

Editorial Board Members

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
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 Neurosciences
Northwestern University Feinberg School
of Medicine

Dr. Pina C. Sanelli

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

Dr. Roberto Sanchez

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

Dr. Wen-Yih Sun

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

Dr. Michael R. Rudnick

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

Dr. Bassey Benjamin Esu

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

Dr. Aziz M. Barbar, Ph.D.

IEEE Senior Member
Chairperson, Department of Computer
Science
AUST - American University of Science &
Technology
Alfred Naccash Avenue – Ashrafieh

Chief Author

Dr. R.K. Dixit (HON.)

M.Sc., Ph.D., FICCT

Chief Author, India

Email: authorind@computerresearch.org

Dean & Editor-in-Chief (HON.)

Vivek Dubey(HON.)

MS (Industrial Engineering),

MS (Mechanical Engineering)

University of Wisconsin

FICCT

Editor-in-Chief, USA

editorusa@computerresearch.org

Sangita Dixit

M.Sc., FICCT

Dean and Publisher, India

deanind@computerresearch.org

Er. Suyog Dixit

BE (HONS. in Computer Science), FICCT

SAP Certified Consultant

Technical Dean, India

Website: www.suyogdixit.com

Email: suyog@suyogdixit.com,

dean@computerresearch.org

Contents of the Volume

- 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 Novel Approach for Extraction of Polygon Regions **2-7**
 2. Functionality and Security Analysis of ORACLE, IBM-DB2 & SQL Server **8-12**
 3. Securing Retina Fuzzy Vault System using Soft Biometrics **13-18**
 4. Performance Improvement of UDP and TCP Traffic by CJM Algorithm in Voice Transmission **19-23**
 5. Fuzzy Set in Business Process Management as Reference Model **24-27**
 6. Impact of Aspect Oriented Programming on Software Development Quality Metrics **28-35**
 7. HE-Advisor: A Multidisciplinary Web-Based Higher Education Advisory System **36-47**
 8. Design and Implementation of RADIUS – An Network Security Protocol **48-54**
 9. DiRiboPred: A Web Tool for Classification and Prediction of Ribonucleases **55-60**
 10. Detecting Copy-Move Forgery in Digital Images A Survey and Analysis of Current Methods **61-65**
 11. Configuration Complexity: A Layered based Configuration Repository Architecture for Conflicts Identification **66-71**
 12. VANET Parameters and Applications: A Review **72-77**
 13. Parametric-Activity Based System Development Method Selection **78-80**
 14. Analyzing Horizontal Integration Among Soa, Cloud Computing And Grid Computing **81-85**
 15. Multi Domain Metadata Standard Harvesting **86-89**
 16. Maximum Likelihood Criterion Based On Polygon Inequality **90-92**
- 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 Novel Approach for Extraction of Polygon Regions

{ GJCST Computing Classification
15.1,15.4 }

Kemal Yükses¹ Metin Turan²

Abstract-This paper presents a new algorithm to find out whether a polygon exists around a reference point given within the graphical domain. The algorithm is based on creating discrete line segments and then searching them using the orientations formed at segments intersections. The computational complexity of the searching algorithm has been determined as $O(n^2)$

I. INTRODUCTION

One of the most important problems that must be solved when developing graphical based system is the determination of simple graphical objects (such as line, circle, etc.) within the drawing area. Although polygons are known as simple graphical elements, they are composed of more simple graphical objects as line segments. In order to understand all the properties of a polygon, one must know the each line segment of that object [1]. There are widely used methods finding line segments of each of the polygon inside the whole structure [2]. Polygonal objects can be classified as convex or concave in shape according to the connection of the points inside of them. Most of the algorithms have been developed up to now especially deals with convex polygons [3]. So there are a number of algorithms to create convex polygons from concave ones [4]. Most of the polygon algorithms depend on shapes. Many studies on computer vision and robotics are used a ray of a single flash light to detect the edges of polygon regions. There have been such algorithms using the flash light approach [5,6,7]. The following studies deal with vector and/or raster graphics. They follow certain algorithms to detect polygon boundaries. Polygons created with vector graphics have been used on various applications of manufacturing industries to Geographical Information System (GIS) [8, 9, 10]. Raster based algorithms have been used in areas such as remote sensing to extract land cover data from satellite images and pattern recognition [11, 12]. The studies on polygons concentrate on the techniques such as computing the centre of a polygon region, calculation the area size and finding the centre of gravity of the polygon, whether a certain point is located within the polygon or not, intersection points of two polygons, area size (Hidden regions) of two overlapping polygons, defining the location of polygons and the triangulation of a simple polygon [13, 14, 15]. The proposed method in this study is based on an algorithm designed to determine the polygon region characteristics[16]. The initial stage of the algorithm has been inspired of the

well-known scan line filling algorithm [17]. The developed algorithm has already been applied for calculating the heat requirement of a building project.

II. THE PROBLEM DEFINITION AND PROPOSED SOLUTION

Searching a polygon around a reference point within the drawing domain is an important issue in computer graphics. The problem may become more complex if the assumption about the type (convex or concave) of the polygon is ignored. The following part of the study has been organized to construct the requirements of the proposed method first, and then algorithm itself.

Polygon boundaries are constituted by line segments which are basic design elements. These line segments could be defined by a user at random, or the result of a straight line recognition algorithms on an image.

One instinctively may think that it is a very trivial problem. If all the cases are considered, the problem becomes more complex to solve. Especially the polygons in the drawing area may be concave or convex in shape. So, the algorithm must should solve for both of them.

The method in this article consists of two steps. The first step is to create individual discrete line elements which are separated at the intersection points. The second step is to distinguish those line segments that construct the polygon region around a given reference point.

In the process of creating discrete line segments, it would be necessary to use mathematical representation of line segments. Although there are two ways for the representation; analytical one and parametrical one. The first one suffers from the intersection testing [18]. In order to avoid this shortcoming, parametrical representation of line segments has been preferred. In this representation, the start and end points of two line elements P1, P2 and P3, P4 can be defined as follow

$$P_i = (x_i, y_i), \quad i=1,2,3,4;$$

$$P(s) = P_1 + (P_2 - P_1)s \quad 0 \leq s \leq 1 \quad (3)$$

$$P(t) = P_3 + (P_4 - P_3)t \quad 0 \leq t \leq 1 \quad (1)$$

As a result of this representation, the intersection of the two line elements $Z = (x_z, y_z)$ is found from the following equation:

$$P(s) = P(t) \quad (2)$$

In this method, the user can select any line element at random, and then, this line element is checked against all the others in the graphical database to test if it intersects any of them. If there are intersection points, then, the line elements at each intersection points are broken down as individual discrete line elements on the basis of this intersection point. Then the same process is repeated for the other intersection points. As a result of this process, all the line elements in the

graphical database would be in the form of discrete line segments. Finally, the intersection points become vertices of the polygons in the system if there are any. Figure 1 illustrates the process of how to create line segments. In case of a discrete line drawn at random intersects with other discrete lines in the graphical database. Figure 2 shows the possible segmentations.

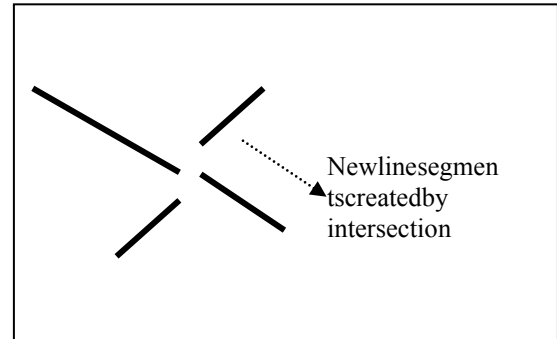
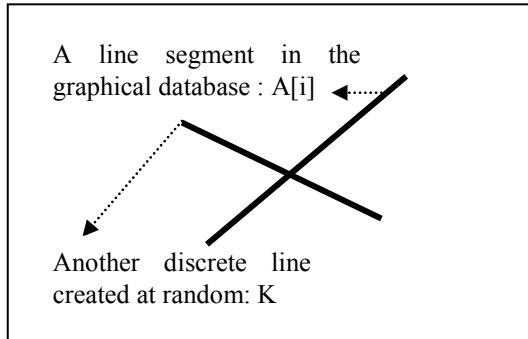


Fig. 1. The process of creating discrete line elements

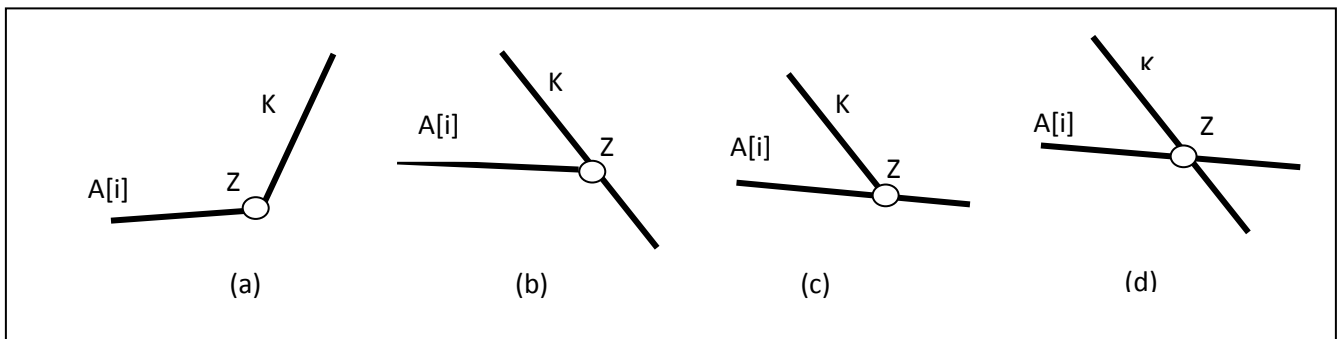
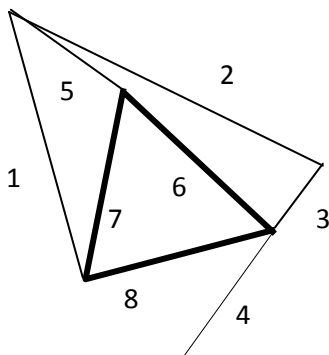


Fig. 2. The cases of intersection of two discrete lines in the graphical database

A new entry is inserted in adjacent matrix for each new line segment created by intersection. Adjacent line segments are stored in this bit matrix. Figure 3 shows an example of

adjacent matrix. It is used to search the possible candidates for the next coming line segment on the polygon which accelerates the SearchPoly algorithm



	1	2	3	4	5	6	7	8
1	0	1	0	0	1	0	1	1
2	1	0	1	0	1	0	0	0
3	0	1	0	1	0	1	0	1
4	0	0	1	0	0	1	0	1
5	1	1	0	0	0	1	1	0
6	0	0	1	1	1	0	1	1
7	1	0	0	0	1	1	0	1
8	1	0	1	1	0	1	1	0

Fig. 3. A simple polygonal region and related adjacent matrix

The pseudo-code for creating line segments is given below:

n : the number of the line segments in the graphical database
A[i]: the set of the line segments in the system ($i=1,2,\dots,n$)
A[i].first :one end of the line segment A[i]
A[i].second :the other end of the line segment A[i]
K: current line which is just drawn by the user (selected line)

K.first: one end of the K

K.second: the other end of the K

Z: intersection point

Procedure createSegments(K)

```
{
  For each line segment A[i],  $i=1,\dots,n$ 
  {
    If A[i] line segment intersects with discrete K line at a point Z
    {
      If Z is start or end point of A[i] {
        If Z is start or end point of K then save K as line segment
      }
      Else {divide K into two separate line segments
        A[n+1]=(K.first,Z),
        A[n+2]=(Z,K.second)
        update adjacent matrix}
    }
  }
  Else {
    If Z is start or end point of K then
    {divide A[i] into two separate line segments
      A[i]=(A[i].first,Z), A[n+1]=(Z,A[i].second)
      update adjacent matrix}
  }
}
```

Else {divide K and A[i] into two separate line segments

A[i]=(A[i].first,Z),

A[n+1]=(Z,A[i].second)

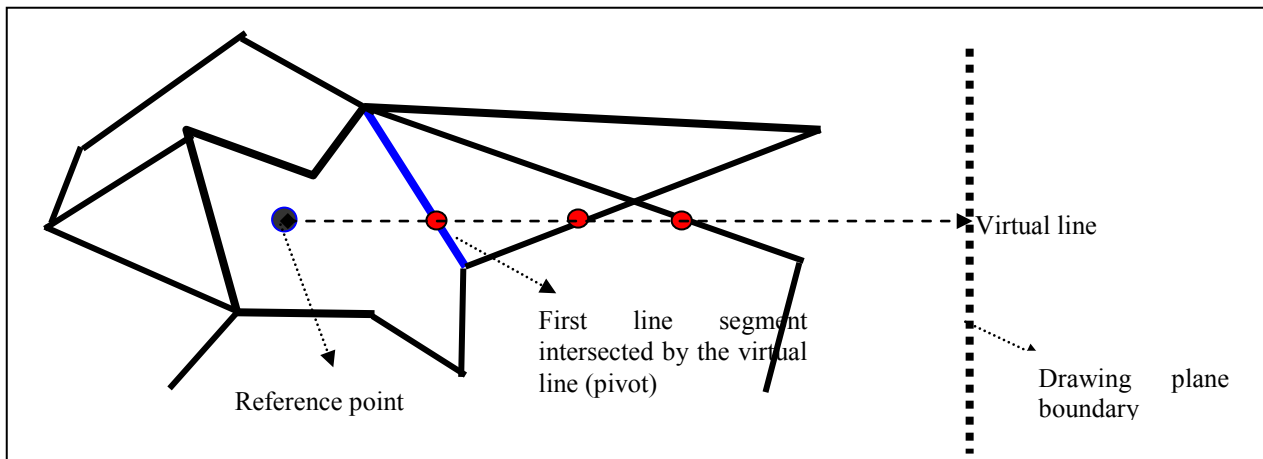
A[n+2]=(K.first,Z),

A[n+3]=(Z,K.second)

update adjacent matrix}

Finally, the model is ready to check a given reference point if there is a polygon region around it or not. The second step is to take a reference point. Once the reference point has been selected, a vertical virtual line from the reference point is drawn to the right boundary of the drawing plane. The virtual line intersects some line segments and results a set of intersection points along the same line. The line segment, which has an intersection point that has the minimum distance to the reference point, has been selected. This is the first line segment of the polygon region.

Figure 3 illustrates the selection of the reference point, the virtual line and determining the first line segment of the polygon region. One end of the first line segment is taken as a basis (in other words base point) for executing the algorithm further. As there are two ends of the first line segment, obviously there are two alternative directions to go forward. Selection of the end point also defines the direction of the algorithm. Figure 4 shows decision of one end of the first line segment (in other words the direction of algorithm).



Once the algorithm direction has been defined, it remains same through the whole algorithm. After deciding the base point of line segment, the other line segments starting from or ending at the same point are selected from the graphical database. The angles created between the based line segment

and the other selected line segments are calculated using opposite direction of the algorithm direction. For example, if the algorithm direction is clock-wise, the angles are evaluated in anti-clock-wise direction.

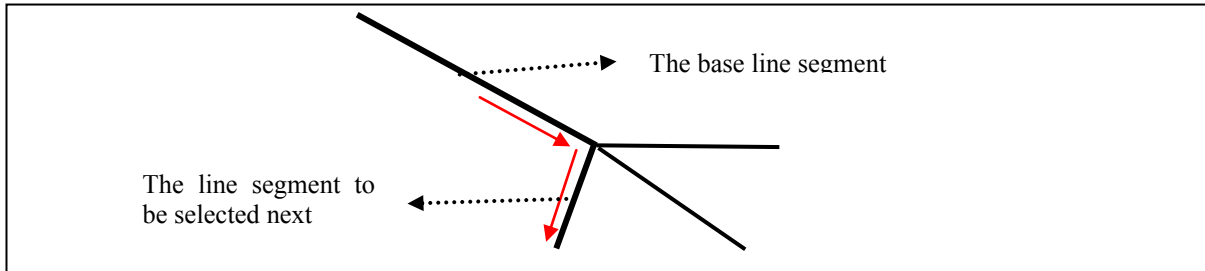


Fig. 4. Selection of algorithm direction

Figure 5 shows the direction of computation and the angles created by the line segments joining together at a base point. The line segment which gives the smallest angle at a base point is the new base line segment of this polygon region. And the same operations are performed repeatedly.

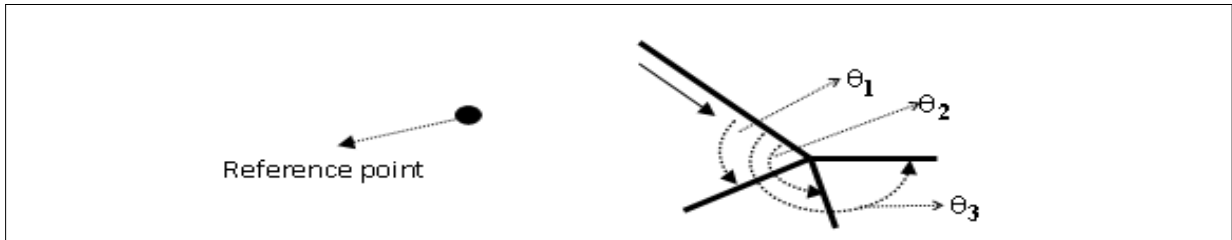


Fig. 5. Angular calculations at the base point (opposite to the algorithm direction)

The algorithm ends when the repeated operation reaches the first line segment or there isn't any line segment to go forward (selection of new base line segment).. If it reaches the first line segment, it means that all the line segments traced up to now construct the desired polygon. If it fails to find a next line segments, it shows that there is no closed polygon containing the starting point or there is an open door in the selected area. Both of these cases might have significant meanings depending on the application. The pseudo-code for searching polygon region is given below. Moreover, Figure 6 illustrate execution of searchPoly(B) algorithm

```

{
  Set BLS to CLS
  while (TRUE)
  {
    If (bit sum of the CLS row in the adjacent matrix is zero) return false
    /* One may eliminate (put zero) the columns in the adjacent matrix whose line segments searched before */
    Calculate the angles between the CLS and its adjacent line segments (use adjacent matrix)
    Choose the smallest ALS measured in the opposite direction to the algorithm direction
    Find new BP for selected ALS
    Save selected ALS into the line segments list of possible polygon
    If selected ALS joins with other end of BLS return true
    Set ALS to CLS
  }
}
    
```

BLS: base line segment
 ALS: adjacent line segment
 CLS: current line segment
 BP : base point
 boolean **Function searchPoly**(BLS , BP)

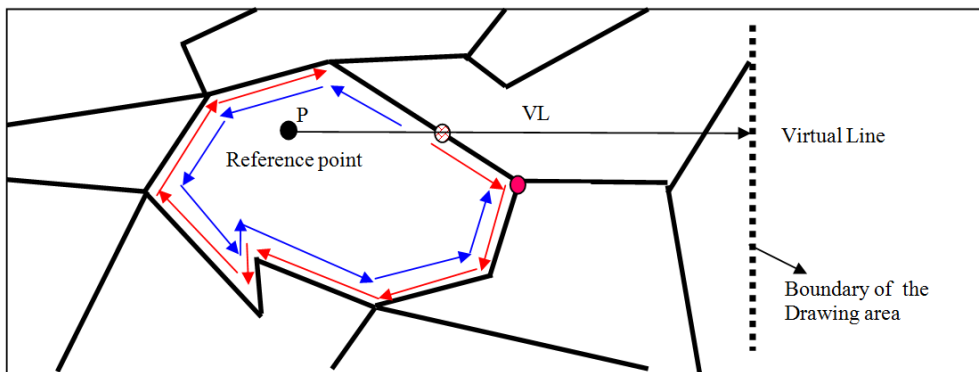


Fig. 6. Execution of the searchPoly() function

III. ANALYSIS OF THE ALGORITHMS

The total cost of constructing line segments must be examined for worst and best cases. For each case, It is assumed that there n line segments are drawn. In best case, it is assumed that there isn't any intersection causes fragmentation of discrete lines. The size of the Line segment database determines the complexity as $O(n)$. For the scan operation on the Adjacent matrix causes $O(n^2)$. As a result of this, the complexity of the searchpoly method is $O(n^2)$. In worst case, it is assumed that each line drawn $i, i=2,3,..n$ intersects all previously drawn lines. Then the new line i will cause to generate at most $2i-1$ new line segments (where $(i-1)$ of them generated by segmenting previous lines, and i due to new line drawn). Therefore the total number of line segments is

$$1 + \sum_{i=2}^n 2i - 1 \quad (4)$$

which is $O(n^2)$. By adding adjacency matrix calculation the total complexity will be $O(n^2)$.

Composing algorithm must be applied during the drawing or defining the discrete lines. So, complexity of composing algorithm does not have any effect on performance of searching polygon region.

The complexity of searchPoly(BLS,BP) algorithm can be decreased to $n*n$ by using the adjacency matrix. It is clear that, if search space can be limited to some line segments, the algorithm works faster. However, the cost of the search algorithm is still acceptable for the application to the complex cases.

IV. IMPLEMENTATION

The algorithm has been implemented within a CAD based heat requirement calculation of a building project. It has been developed using Visual Basic programming language and its graphics tools. Figure 7.a and Figure 7.b show examples of composing segments while Figure 8 demonstrates how to select a specific polygonal region.

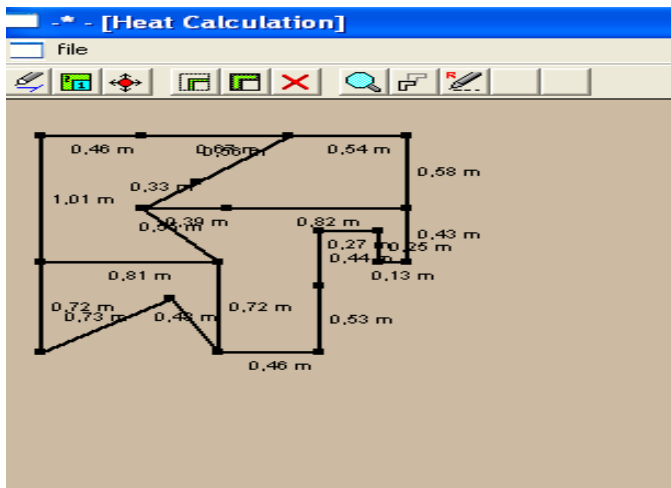


Fig. 7.a. Composing segments within the CAD based Heat Evaluation Program (before the intersection)

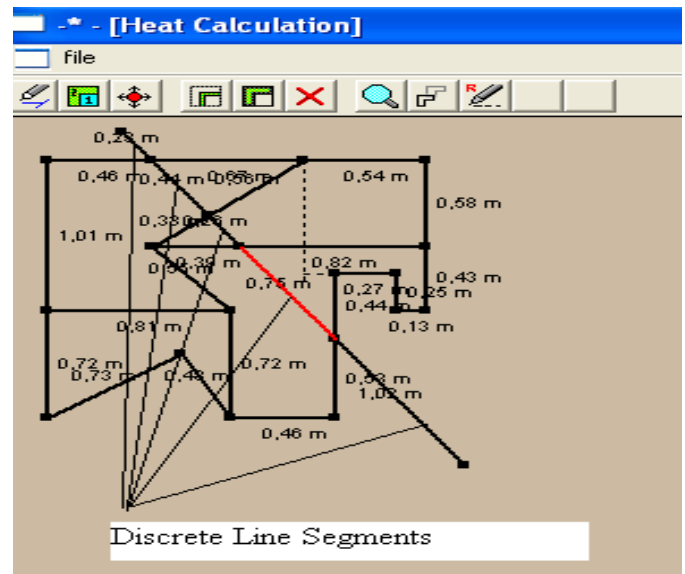
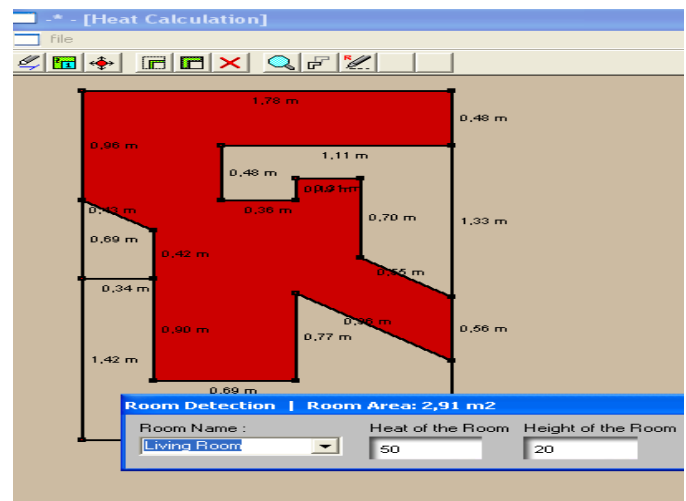


Figure 7.b. Composing segments within the CAD based Heat Evaluation Program (after the intersection)



V. CONCLUSION

In the light of the literature review has been made in this particular field, this method and algorithms seem to be a unique solution for the definition of the polygon region characteristics. The analysis of the algorithms indicates that they have got a fast and robust structure. This method can be used a wide variety of application areas. These applications may spread to the region definition, finding orientations and graphical object recognition. Such applications may be implemented in the Computer Graphics Industries, particularly, Computer Aided Design (CAD), Computer Aided Manufacturing (CAM), Robotics and Geographical Information System (GIS).

For the future studies, this approach can be extended for the 3D model in order to obtain the edges or surfaces of a polyhedral structure.

VI. REFERENCES

- 1) J. Pach, E. R. Campo, On circumscribing polygons for line segments, *Computational Geometry* 10 (1998) 121-124
- 2) [X.Li, T.W.Woon, T.S.Tan, Zhuang, Decomposing Polygon Meshes For interactive Applications, *Proceedings of the 2001 symposium on Interactive 3D graphics*,2001,35-42
- 3) M.A. Lopez, S. Reisner, Efficient approximation of convex polygons, *International Journal of Computational Geometry & Applications* Vol. 10, No. 5, (2000), pp 445-452
- 4) J. Fernandez, L. Canovas, B. Pelegrin, Algorithms for the decomposition of a polygon into convex polygons, *European Journal of Operational Research* 121 (2000) 330-342
- 5) [5] S. Y. Shin, T. J. Woo, An Optimal Algorithm for Finding all Visible Edges in a Simple Polygon, *IEEE Transactions on robotics and automation*. Vol 5. Num. 2, April ,1989
- 6) S. M. Lavalley,B.H.Simov, G.Slutzki, An algorithm for searching a polygonal region with a flashlight, *International Journal of Computational Geometry & Applications* Vol. 12, Nos. 1 & 2 (2002) 87-113.
- 7) M.Keil, J. Snoeying, On the time bound for convex decomposition of simple polygons *Int. Journal of Computational Geometry & Applications* Vol. 12, No:3. (2002).
- 8) P. K. Agarwal,N. Amenta, M. Sharir, Placement of one convex polygon inside another, *Discrete Computational Geometry*, Vol 19, 1999,pp 95-104
- 9) [9] Y. Ishigami, How Many Diagonal Rectangles Are Needed to Cover an Orthogonal Polygon?, *Discrete Computational Geometry*, Vol 24,2000,pp117-140
- 10) K.Jha Manoj, C. McCall, P. Schonfeld, Using GIS, Generic Algorithms, and Visualization in Highway Development, *Computer-Aided Civil and Infrastructure Engineering* 16(6), (2001), 399-414.
- 11) V.Sridhar,M.A. Nascimento, ,Xiaobo Li, Region-based image retrival using multiple-features, in "5th International Conference VISUAL 2002", Vol 2314, pp 61-75.
- 12) C. Huggel, A. Käab, W. Haeberli, B. Krummenacher, Regional-scale GIS- models for assessment of hazards from glacier lake outbursts: evaluation and application in the Swiss Alps, *Natural Hazards and Earth Sciences* (2003), 647-662.
- 13) J. L. A reliable test for inclusion of a point in a polygon, *ACM SIGCSE Bulletin* Volume 34 , Issue 4 (December 2002),81 - 84
- 14) D. M. Mount, Intersection Detection and Seperators for simple polygons, *Annual Symposium on Computational Geometry Proceedings of the eighth annual symposium on Computational geometry* Berlin, Germany, (1992), 303 - 311
- 15) W. Lenhart, R. Pollack, J. Sack, R. Seidel, M. Sharir, S. Suri ,G. Toussaint, S. Whitesides, C. Yap, Computing the Link Center of a Simple Polygon, *Annual Symposium on Computational Geometry Proceedings of the eighth annual symposium on Computational geometry*, Waterloo, Ontario, Canada,(1987),1-10
- 16) D. P. Luebke, A Developer's Survey of Polygonal Simplification Algorithms, *IEEE Computer Graphics and Applications*, May 2001, pp. 24-35
- 17) T. Whitted, A scan line algorithm for computer display of curved surfaces, *ACM SIGGRAPH Computer Graphics*,vol. 12,issue 3, 1978,pp. 26-31
- 18) D.F.Rogers, Lines intersections Algorithms, *Procedural Elements for Computer Graphics*, 3rd printing,1988,pp 172-183

Functionality and Security Analysis of ORACLE, IBM-DB2 & SQL Server

A. Jangra¹ D. Bishla² Komal Bhatia³ Priyanka⁴

GJCST Computing Classification
H 2.7,D 4.6

Abstract-Information may be the most valuable commodity in the modern world as it takes many application dependent different forms. We need to store too much data in file cabinets or cardboard boxes. The need to safely store large collections of persistent data, efficiently “slice and dice” it from different angles by multiple users and update it easily when necessary is critical for every enterprise. Security pin the main feature of DBMS like Encryption, Authentication, Proxy Authentication, Authorization, Auditing, LDAP Support, etc. Functionality is the most important feature of any DBMS. How data function in different-different situation like Concurrency Model, index capabilities, partitioning options, Parallel execution, Clustered configurations, Additional data warehousing capabilities, Self tuning capabilities, Array, Trigger, Procedures, Tables etc. In this paper we compare the globally recognized database’s to get the details of all these above features and also some extra comparative parametric features.

I. INTRODUCTION

An organization must have accurate and reliable data for a better decision making. The objective of the DBMS is to provide a convenient and effective method of defining, sorting, and retrieving the information contained in the database. The DBMS interfaces with application programs, so that the data contained in the database can be used by multiple application and multiple users. The database system allow these users to access and manipulate the data contained in the database in a convenient and effective manner. Every organization choose the database management system according to there need and requirement. Security and functionality is the biggest issue in any DBMS. There are many DBMS in market. Example of Database Management Systems is Alpha Five, DataEase, Oracle database, IBM DB2, Adaptive Server Enterprise, FileMaker, Firebird, Ingres, Informix, Mark Logic, Microsoft Access, Microsoft SQL Server, Microsoft Visual FoxPro, MonetDB, MySQL, PostgreSQL, Progress, SQLite, Teradata, CSQL, OpenLink Virtuoso, Daffodil DB, OpenOffice.org Base etc. But we select three type of database management system to compare there security and functionality these are Oracle, IBM DB2 and Microsoft SQL Server, we preferred these three because these three DBMS preferred widely in the market (as show in table 1)

About-^{1,2} CSE deptt. U.I.E.T. Kurukshetra University, India (telephone +91-9466027922, e-mail- er_jangra@yahoo.co.in,) (telephone +91 9996817678, e-mail- d.bishla@gmail.com)

About-³ CSE deptt. YMCA University of Sc. & Technolgy, (Hr.) India, (telephone +91-99953637670, e-mail- komal_bhatia1@yahoo.com)³

About-⁴ ECE deptt. K.I.T.M. Kurukshetra (Hr.) India. hhhhhhhh (telephone +91- 9466751345, priyanka.jangra@gmail.com)⁴

and follow the more security and functionality rather than other. I read & surveyed the many research paper and company profile where I get 93% were used oracle and 3% used IBM DB2 & SQL Server and only 4% other database management system. [2,8,15]

Feature	Oracle	DB2	SQL Server
Concurrency Model	Multi-version read consistency Non-Escalating row level Locking	No Locks escalate	Shared read locks or dirty reads Locks escalate
Indexing capabilities	B-Tree indexes Index-organized Tables Bitmap indexes Bitmap Join Indexes	Only B-Tree and Dynamic Bitmap Index	B-Tree indexes Clustered Indexes Not supported Not supported
Partitioning options	Range, hash, list and composite partitioning Local and global indexes	Hash partitioning and Local index	Not supported Only local indexes with member tables
Parallel execution	Queries, INSERT, UPDATE, DELETE	Queries only	Queries only
Clustered configurations	Transparent scalability with Real Application Clusters	Rigid Data partitioning required with DB2 EEE	Requires data partitioning in member tables and Distributed Partitioned Views
Additional data warehousing capabilities	Materialized Views MERGE Multi-table INSERT Pipelined table Functions	Not Supported	Indexed Views Not supported Not supported Not supported
Self tuning capabilities	Automatic Performance diagnosis Automatic SQL Tuning Self-tuning memory, free space and I/O management	No equivalent or limited capabilities	No equivalent or limited capabilities
Array	Supported	Supported	Not Supported
Trigger	BEFORE triggers, AFTER triggers, INSTEAD OF triggers, Database Event triggers	BEFORE triggers, AFTER triggers, INSTEAD OF triggers	AFTER triggers, INSTEAD OF triggers
Procedures	PL/SQL statements, Java methods, third-generation language (3GL) routines	DB2 SQL dialect statements, Java methods, third-generation language (3GL) routines	T-SQL statements
Tables	Relational tables, Object tables, Temporary tables, Partitioned tables, External tables, Index organized tables	Relational tables, Object tables, Temporary tables	Relational tables, Temporary tables

Table 2 – Databases functionality comparison summary.

[11, 12, 13]

Market structure

Given below is a list of top RDBMS vendors in 2009 with figures in millions of United States Dollars published in an IDC study.

Table 1 – Market worth of database

Vendor	Global Revenue
Oracle	8,800
IBM	3,483
Microsoft	3,052
Sybase	524
Teradata	457
Others	1,624
Total	16,452

II. COMPARATIVE ANALYSIS

We compare Oracle, DB2 & SQL Server under the parameter of functionality and security

A. Functionality

The main features differences between three databases are summarized in the table 2: [3, 5]

OLTP environment have large volume of data for sort and frequent update and insert the data. So OLTP require high throughput, index strategy and excellent data concurrency etc. **Concurrency control** in multi-user environments ensures that data updates made by one user do not affect those made by other users. Oracle, DB2 and SQL Server differ greatly in their implementation of concurrency control. The main differences show in the below table 3.

Oracle	SQL Server	DB2
Multi-version read consistency	Not available	Not available
No read locks	Requires shared read locks to avoid dirty reads	Requires read locks to avoid dirty reads
No dirty reads	Dirty reads if not using shared locks	Dirty read if not using read lock
Non-escalating row-level locking	Locks escalate	Locks escalate
Readers don't block writers	Readers block writers	Readers block writers
Writers don't block readers	Writers block readers	Writers block readers
Minimal deadlocks under load	Deadlocks can be a serious problem under load	Deadlocks can be a problem under load

Table 3 - Sub table of functionality comparison (Concurrency Model). [3, 9, 14]

Oracle fully support the mix workload of simultaneously query and insert, update commands. In Oracle no read locks, no dirty reads, reader cannot block the writer, writer cannot block the readers. Data is always available for user

because no deadlocks occur. Oracle's implementation of multi-version read consistency always provides consistent and accurate results. When an update occurs in a transaction, the original data values are recorded in the database's undo records. Oracle uses the current information in the undo records to construct a read-consistent view of a table's data, and to ensure that a version of the information, consistent at the beginning of the uncommitted transaction, can always be returned to any user.

DB2 block the reader for writers and writers for readers. DB2 does not provide multi-version read consistency. DB2 block the readers while writing and block the writers while reading. DB2 not allow the mix of read and write request. IBM also accepts this fact in their own documents.

SQL Server also does not provide multi-version read consistency. Instead it requires applications to either use shared locks for read operations, with various levels of isolation, or to accept dirty reads. Shared locks prevent data that is read from being changed by concurrent transactions. Shared lock's to ensure that data readers only see committed data. These readers take and release shared locks as they read data. These shared locks do not affect other readers. A reader waits for a writer to commit the changes before reading a record. A reader holding shared locks also blocks a writer trying to update the same data. Important thing is that ~~releasing~~ the locks quickly for other users in SQL Server than in Oracle.

Non-Escalating Row-Level Locking Row-level locking ensures that any user updating a row in a table will only lock that row, leaving all other rows available for concurrent operations. Oracle uses row-level locking as the default concurrency model and stores locking information within the actual rows themselves. By doing so, Oracle can have as many row level locks as there are rows or index entries in the database, providing unlimited data concurrency. Oracle never locks and as a consequence oracle users never face the situation of deadlock due to lock escalation.[7,3]

DB2 also supports row level locking by default. Lock list is an additional memory structure these lock lists have limited size so that limited number of lock are reside in memory structure or lock list. Lock escalation is an internal mechanism that is invoked by the DB2 lock manager to reduce the number of locks held in lock list. Escalatin occur from row locks to a table lock when the number of locks held exceed the threshold defined by the database configuration parameter Lock list. [5]

SQL Server also supports row-level locking as the default concurrency model. However, because it was not the default level of lock granularity in earlier versions of the database, the late addition of row-level locking was made possible only through the use of additional, separate pools of lock structures.

Indexes are basically used for sorting operation on table columns and provide a faster path to data. Using indexes can reduce disk I/O operations, so that increasing the performance of data retrieval. Oracle, DB2 and SQL Server support traditional B-Tree indexing schemes, which are

ordered lists of key values, associated with the storage location of the table row that contains these values.

Oracle support index-organized tables, (IOT itself a table space) bitmap and bitmap join index. DB2 support the dynamic bitmap index and clustered indexes in SQL server. Index-organized tables provide fast access to table data for queries involving exact match and/or range search on the primary key because table rows are stored in the leaf nodes of the primary key index. For example English dictionary that themselves an indexed.[1,5]

A bitmap index uses a bitmap (or bit vector) for each key value instead of a list of the table rows' storage locations (ROWIDs). Each bit in the bitmap corresponds to a row in the table. The bit is set when the table's row contains the key value.

In Oracle, it is also possible to create bitmap indexes on index-organized tables, thereby allowing index-organized tables to be used as fact tables in data warehousing environments. A bitmap join index is a bitmap index for the join of two or more tables. A bitmap join index can be used to avoid actual joins of tables, or to greatly reduce the volume of data that must be joined, by performing restrictions in advance. Queries using bitmap join indexes can be sped up via bit-wise operations.

Bitmap join indexes, which contain multiple dimension tables, can eliminate bitwise operations, which are necessary in the star transformation with bitmap indexes on single tables. Performance measurements performed under various types of star queries demonstrate tremendous response time improvements when queries use bitmap join indexes. DB2 and SQL Server do not support IOT, bitmap indexes and bitmap join indexes.[1,3,5]

Partitioning allow the large database in to small pieces and also store the pieces in different-different location. So that data can retrieve and stored fast with more I/O process.

Oracle hold the all Partitioning options like Range, hash, list and composite partitioning Local and global indexes. DB2 hold the Hash partitioning and Local index. And SQL Server hold the Only local indexes with member tables.

Here Oracle keep the more partitioning option that is not in DB2 neither in SQL Server.

Cluster is an group of independent server connected via a private network. All server work as a single system. Oracle and DB2 support the cluster but SQL Server does not support the Cluster. Oracle use the Real application cluster (RAC) to support the hardware cluster. RAC adopts a shared link approach for this database file are logically shared among the nodes of a loosely coupled system with each instance having access to all the data. RAC use the patented cache fusion architecture, a technology that utilizes the interconnected cache of all the nodes in the cluster to satisfy database request for any type of application (OLTP, DSS, Packaged application). RAC is unique feature of oracle and make it best. DB2 adopts the shared nothing approach. In this database file are partitioned among the instances running on the nodes of a multi computer system. Each instances on different subnet

of the data and all access to this data is performed exclusively by this owing instance.[1,3,4,5]

Additional data warehousing capabilities Extraction, Transformation and loading (ETL) Oracle provide the additional feature of data warehousing environment like materialized view, Merge, Multi table insert, Pipelined table function. Both DB2 and SQL Server does not support the these additional features. Oracle need not tune the database because Oracle have self tuning means automatic storage management (ASM), Automatic work repository (AWR), Automatic database diagnostic monitor (ADDM), Automatic SQL tuning And automatic back and recovery management like this all work of tuning done in oracle by automatic but in the DB2 and SQL Server so the database tuning manually. [1, 3]

B. Security

Whenever any company plan to purchase DBMS firstly company think about security. Security feature at the top of buyer's list. DBMS purchasing decision must factor in security. A company reputation and livelihood may be ruled if it does not protect the data or customer information held inside. Today is e-business world, so security issues have become more complicated than ever.

Mainly security consists of these major factors specify in below table 4.

Authentication is ensuring only right user is connect to the database and use to prove the identity of the user. In large enterprises applications expand day-to-day, therefore need for strong user authentication techniques grows up. User-id/password, keys and biometric authentication are some of the solutions used to address the authentication problem. Authentication have three levels network, DBMS and in the operating system.

Security feature	Oracle	IBM DB2	SQL Server
Authentication	Yes	Yes	Yes
Proxy Authentication	Yes	No	No
Authorization	Yes	Yes	Yes
Encryption	Yes	Yes	No
LDAP Support	Yes	Yes	Yes
Auditing	Yes	Yes	No
Fine grained auditing	Yes	No	No
RACF Support	Yes	Yes	No

Table 4 – DBMS Security features. [1,3]

Oracle support strong authentication at the network and database layers by supporting X.509v3 digital certificate and also integrated with third party network authentication services i.e. token card smart card Kerberos, DCE, biometrics and Cyber Safe. To connect with remotely to oracle DBMS using RADIUS that is most secure in all database using the SecureID tokens, secure computing safeword token, smart card and active card tokens built-in into the database. Oracle's have many authentication methods like internal user authentication, OS authentication and network authentication methods.[3,9]

IBM support strong authentication at the database and operating system layers and in various Tivoli application.

That is 3rd party application. IBM support service such as DCE, Kerberos, and RACF. Tivoli secure way policy support only secureID the leading token or also called hard token, but leaving customers with only one choice. That use X.509v3 certification for strong authentication over SSL. DB2 have external authentication. DB2 always passes authentication requests to the operating system and/or 3rd party products such as IBM's Tivoli. To support this, DB2 has employed in version 8.2 an open plugin architecture. This architecture allows for easy integration of 3rd party or custom authentication plugins that allow for the extension of DB2's authentication capabilities. Operating system and Kerberos (previously only available on Windows, now available on UNIX platforms too) have been re-implemented using the plug-in architecture. On the Windows platform, the depth of integration into the OS authentication capabilities has been depend.

SQL Server offers the DBMS authentication via the operating system security (similar to DB2's implementation). Microsoft use the Active Directory components of Microsoft admin server using the LDAP and Kerberos protocols. The operating system authentication integrates extremely well with the operating system security features. The Kerberos framework is supported in conjunction with Active Directory. Oracle's strategy provides the most authentication alternatives. DB2 now has the architecture to accelerate the integration of new authentication methods. This applies not only for standardized 3rd party products but also for custom-developed authentication methods. SQL Server has tight integration with products supplied by Microsoft itself.

Authorization: Once a user is authenticated to the DBMS, A user's authorization tells us what data he should have to and what types of operation he can perform on those objects. It has to be verified that the user is authorized for accessing the queried data and/or function in the request. Authorization is normally bound to users and groups or roles. Oracle and DB2 both have the same definition of privileges and use standard SQL.

In DB2, authorization can be provide to users or groups. Roles are only supported in the sense of predefined system roles (Roles is the set of privilege). When utilizing group authorization, attaching users to groups is done outside the DBMS, i.e. in the operating system or through custom plugins. Oracle authorizes based on users and roles. Oracle roles can be local to a database or enterprise-wide when managed with an LDAP compliant server. For SQL Server, there is a distinction between a server login and database users, a user connection to a database is only allowed when the server login is mapped to a database user. SQL Server utilizes operating system group information directly without need for mapping. [3,5,10,14]In Oracle and SQL Server, can define application roles that are only used for applications. These prevent direct user access to tables and views. Although DB2 does not support roles, applications can connect to the database with a virtual user-id that implements the necessary level of access security. Content

and functionality control can be implemented at 2 different levels: object and row level. Object level security is a central component of RDBMS technology and has been covered by all 3 DBMS' sufficiently for years. Views have been the traditional answer to row level security and are supported by all DBMS. DB2 and Oracle both support the use of view to limit access the data. Oracle offers an additional, integrated implementation of row level security with Virtual Private Database (VPD) and Label Security (OLS). VPD enables implementation of row level security into the database, OLS manages the actual labeling of both users and data. These labels are directly compared when accessing the data. This approach promises high performance during runtime security checks. However, the manageability is at a disadvantage when security definitions are updated (i.e. department split or join). In this case, labeling must be physically redone, meaning explicit data updates. For further performance improvement Oracle added static and context-sensitive policy execution strategies with 10g.DB2 and Oracle take the advantage of Resource Control Facility (RACF) for access control in mainframe environment. [3,6,7,10,14]

Encryption: Protecting data stored in the database against unauthorized user is enabled for both DB2 and Oracle by data or column encryption. Only Oracle support tablespace level encryption. SQL Server does not support data encryption. However, encryption of login and application role passwords stored at the server and catalog information (such as view and trigger definitions) is enabled. DB2 use the function that enable an application to encrypt data using an RC2 block cipher with a 128-bit key and using an MD2 message digest. It provides column-level encryption, enabling all values in a column to be encrypted with the same key an encryption password. Oracle provide DES (56-bit), 2-key and 3-key Triple-DES (112 and 168 bits) in an encryption toolkit package that enable application to encrypt within the database. DB2 password based key provide flexibility if not a slight overburden on the end user to choose a strong key. Where oracle has made stored data encryption enhancements in four development cycles.

Today everybody wants encrypting data passing over a network. Network encryption is addressed by all DBMS: DB2 has added encryption of network traffic in version 8.2 with DES and RC2 but customer must purchase additional IBM product to encrypt various network layers, SQL Server has integrated SSL encryption into its net libraries and Oracle provides SSL, _DES in 56 bit and 40 bit key length', _RC4 in 256 bit 128 bit 56 bit and 40 bit key length', _2key 112 bit' and _3key 168 bit' encryption with their Advanced Security option. Wherever the database is available Oracle provide the oracle advance security to protect all communication with the oracle database. to prevent modification or replay of data during transmission oracle use an MD5 or SHA-1 message digest included in each network packet. In short oracle provides a variety of ways to encrypt communication over all protocols with any database communication.[5,7,3]

LDAP integration for centralized user management: Oracle, IBM and SQL server are turning to Lightweight Directory Access Control (LDAP) directories to centrally store and manage users. Tivoli secure way user Administration provides an LDAPv3 compliant directory service. Oracle offer an LDAPv3 compliant directory service, Oracle internet directory, and many oracle product are it as a scalable, secure central information repository. SQL server also work with LDAP only support the Microsoft admin server at active directory. DB2 support the LDAP on OS/400, AIX, OS/390 and window. And oracle support the LDAP at all platforms plus Oracle internet directory.

Auditing is keeping the record of user activities in a table of database to track the user activity to ensure that user has done the right action. That is done by DBA. Flexible reporting on audit data is possible in all three databases.

Oracle use the LogMiner utility that is always available and does not drop records of any change made to it that use for recover the database and allow the customer to audit the database by system privileges, statements, by object and user. Oracle keep the all record whether the operation is successful or unsuccessful. Oracle also use the statistics report to audit the database. Their 15 security certificates are seen as the assurance that Oracle is unbreakable. No competitor is as active as Oracle in this area - DB2 and SQL Server have 1 certificate each. DB2's Common Criteria certification applies to the current version, 8.2. At the end of September 2004, we found 6 alerts for Oracle database and Enterprise Manager in 2004, more than 20 vulnerability and incident notes for SQL Server and no alerts for DB2. DB2 provide an administrator tool called db2audit for use by the DBA. DB2 capture the audit record at database level and instance level. DB2 has option of configure to audit trail synchronously or asynchronously. DB2 use Tivoli product to enhance the auditing features namely secure way security manager (login and access to various resources) and secure way PKI (PKI services). But oracle has mandatory log file to record the all entries of database. Oracle use the fine-grained auditing policies which specify the data access condition that audit event.

Oracle comes with its support for multiple authentication methods, its unique row level security, its unique proxy authentication and its support for enterprise users and roles. DB2 comes with advantages to SQL Server due to its new plug-in authentication architecture and the possibility to store data encrypted. [1, 3 5]

III. CONCLUSION

We compare Oracle, IBM DB2 and SQL Server DBMS with function and security parameter and we found that Oracle seems more secure DBMS in comparison to DB2 and SQL server. Because it support all the security feature like Authentication, Proxy Authentication, Authorization, Encryption, LDAP Support, Auditing, Fine grained auditing, RACF Support but on the other hand DB2 not support the Fine grained auditing, Proxy Authentication so

DB2 is the challenger of Oracle but SQL server is behind of them because that support only Authentication, Authorization, LDAP Support.

Functionality is the most important feature of any database. Based on functionality we found that oracle has more function than other database. Oracle use some unique features like multi version read consistency and merge etc. Both oracle and DB2 use more fine-tuning to the configuration can be done via start-up parameters. DB2 is runner up and SQL Server behind of them.

IV. REFERENCE

- 1) IBM DB2 UDB V8.2, Oracle 10g and Microsoft SQL server 2000, a technical comparison by bad homburg, Germany November 2004 Jacqueline Bloemen, Guido Brunner. Version 2.0.
- 2) Kevin Loney, George Koch, and the experts at TUse -Oracle 9i complete reference" oracle press. McGraw-Hill/Osborne, p. 1214 Copyright 2002 act of 1976.
- 3) An Oracle White Paper Aug. 2002,2004, 2005,2007,2008.
- 4) [4.http://www.wisdomforce.com/dweb/resources/docs/MSSQL2005_ORACLE10g_compare.pdf](http://www.wisdomforce.com/dweb/resources/docs/MSSQL2005_ORACLE10g_compare.pdf).
- 5) Oracle University press release notes.
- 6) Understanding DB2® Learning Visually with Examples ByRaul F. Chong, Clara Liu, Sylvia F. Qi, Dwaine R. Snow.
- 7) DB2 ® SQL PL, Second Edition: Deployment and Advanced Configuration Essential Guide for DB2 ® UDB on Linux., UNIX™, Windows ® , i5/OS™, z/OS ® By Zamil Janmohamed, Clara Liu, Drew Bradstock, Raul Chong, Michael Gao, Fraser McArthur, Paul Yip.
- 8) 2780A Maintaining a Microsoft SQL server Database, part no.X12-11927, 02/2006.
- 9) Oracle Database High Availability Overview 11g Release 1(11.1) http://download.oracle.com/docs/cd/B28359_01/se rver.111/b28281/toc.htm.
- 10) Overview: Oracle Maximum Availability Architecture(MAA) <http://www.oracle.com/technology/deploy/availability/htdocs/maa.htm>.
- 11) Overview: Oracle Database High Availability <http://www.oracle.com/technology/deploy/availability>
- 12) A technical white paper by davidgornshtein and boristamarkin, revision 1.3
- 13) 13. Oracle to DB2conversion guide: compatibility made easy.
- 14) comparing SQL server 2008 to IBM DB2 9.5, SQL server technical article.
- 15) Introduction to IBM DB2 UDB, IBM data management solutions education services. Course FN 625.2,IBM part Z251-1599-00, September 5, 2003

Securing Retina Fuzzy Vault System using Soft Biometrics

{ GJCST Computing Classification
D 4.6, I 2.3 }

N. Radha¹ Dr. S. Karthikeyan² P. Anupriya³

Abstract-The major concern of almost all the verification system is user authentication and security. This necessitates the development of a mechanism that ensures user security and privacy. A lot of research has been carried on this developing field and numerous techniques have been proposed earlier in literature. These traditional methods use tokens and passwords to provide security to the users. Uncertainly, it can be easily compromised by attackers and therefore it is significant to design verification system that ensures authentication. In recent years, technology has turned in favor of combining soft biometrics and cryptographic key generation technique. The principal feature of using soft biometric template is that it cannot be easily revoked by any unauthorized user. Most commonly used soft biometric features are iris, retina, face, fingerprint, voice and so on. Fuzzy vault is the framework which comprises of the combination of soft biometrics and cryptographic key generation technique. This fuzzy vault acts as an additional layer of security. This overcomes the limitation met by a biometrics system when implemented individually. This paper proposes a biometric verification system investigating the combined usage of soft biometrics features hardened by fuzzy vault scheme. This approach uses retina as a soft biometric since it is capable of providing best results. Experiments were conducted to investigate the performance of the proposed authentication system in ensuring the user security and privacy.

Keywords-Authentication, Cryptography, Fuzzy Vault Scheme, Retina Feature Extraction, Retinal Soft Biometrics.

I. INTRODUCTION

Biometric technology identifies individuals automatically by using their biological or behavioral characteristics. There is growing interest in the use of biometrics for a large spectrum of applications, ranging from governmental programs to personal applications such as logical and physical access control. Since biometric properties cannot be lost or forgotten in contrast to tokens and passwords, they offer an attractive and convenient alternative to identify and authenticate user information.

The initial step of providing biometric authentication to user is enrollment. In this enrollment stage a user registers with the system where one or more measurements of user biometric data are obtained. Each such measurement is then processed by some algorithm to obtain a "template", and stored in a database. Some of the user biometrics that is

extensively used for authentication is face, fingerprint, hand geometry, keystroke dynamics, hand vein, iris, retina, signature, voice, facial thermogram, and DNA.

The use of above mentioned biometrics for recognizing individuals is becoming increasingly accepted and many applications are already accessible. These applications can be hardly classified in to one of the following two categories, verification and identification [1] [2] [3]. The verification systems validate a person's identity by comparing the captured biometric characteristic with that of person's own biometric template previously stored in the system whereas the identification systems recognize an individual by searching the entire template database for a match with the captured biometric characteristic.

The substitution of biometric features in the place of passwords provides an assortment of advantages in verification systems such as access control and so on. Although biometrics provides a variety of advantages it has some limitations. Once a biometric image or template is stolen, it is stolen forever and cannot be reissued, updated, or destroyed. An additional problem associated with the use of biometrics is that once a biometric is chosen, the same biometric will be used to access many different systems. This means that, if it is compromised, the attacker will have right to use all the accounts/services/applications of that particular user [6]. This is the correspondent of using the same password across multiple systems, which can lead to some very serious problems in terms of security. Recently, novel cryptographic techniques such as fuzzy commitment and fuzzy vault were proposed to provide a secure storage for the reference biometric template [4] [5]. The soft biometric template of the user is vault with the randomly generated key by a cryptographic framework so called "Fuzzy Vault Scheme." This overcomes the limitation met by a biometric system when implemented individually. Moreover it improves user authentication and security.

This paper proposes a biometric verification system, exploring the combined usage of soft biometrics features hardened by fuzzy vault scheme. This proposed approach, which will have enhanced security on comparison with the traditional systems. The soft biometric feature used in this method is retina, since it has been reported to provide some of the best results for verification systems and it remains fairly unaltered during a person's lifetime. Experiments were conducted to examine the performance of the proposed authentication system in ensuring security and privacy.

The remainder of this paper is organized as follows. Section 2 discusses the related work proposed earlier in literature for soft biometric authentication systems. Section 3 explains our proposed system for providing authentication-using retina as

About-¹Ph.D.Scholar,Department of Computer Science Karpagam University,Coimabtoire-21 (e-mail-Lakshmin07@sify.com)

About-²Director,Department of Computer Science Karpagam University,Coimabtoire-21(e-mail- skarthi@gmail.com)

About-³Lecturer,MCA Department PSGR Krishnammal college for women Peelamedu,Coimabtoire-4 (e-mail-anupriya@grgsact.com)

soft biometric feature by hardening the fuzzy vault scheme. Section 4 illustrates the experimental results with necessary explanations and Section 5 concludes the paper with fewer discussions.

II. RELATED WORK

Numerous research works has been proposed previously, which suggests the combination of biometrics and cryptography for developing a verification system [7] [8]. These are referred to as cancelable biometrics since it makes use of a one way transformation to convert the biometric signal into irreversible form. This section of the paper discusses some of the relevant work proposed earlier in literature for developing a user authentication system using soft biometric characteristics and fuzzy vault scheme. The hardening of soft biometric features with fuzzy vault scheme improves user security and privacy.

Moi et al. in [9] put forth an approach for identity document using iris biometric cryptography. They presented an approach to create a distinctive and more secure cryptographic key from iris template. The iris images are processed to generate iris template or code to be utilized for the encryption and decryption tasks. The international standard cryptography algorithm – AES has been adopted in their work to produce a high cryptographic strength security protection on the iris information. Their proposed approach comprises of two processes. They are encryption and decryption process. Template matching is the process used for pattern recognition. The utilization of biometric as a key is to enhance security in a more efficient way, decrease human mistakes during identification, increase user convenience and automation of security function. Their experimental results revealed that their proposed approach out performed some of the traditional techniques in providing authentication for the user.

A two-phase authentication mechanism for federated identity management systems was described by Abhilasha et al. in [10]. The first phase consists of a two-factor biometric authentication based on zero knowledge proofs. They employed techniques from vector-space model to engender cryptographic biometric keys. These keys are kept secret, thus preserving the confidentiality of the biometric data, and at the same time make use of the advantages of a biometric authentication. The second authentication combines several authentication factors in concurrence with the biometric to make available a strong authentication. A key advantage of their approach is that any unexpected combination of factors can be used. Such authentication system leverages the information of the user that is available from the federated identity management system. Their proposed approach improves privacy, reliability, security of the biometric data.

Uludag et al. in [11] discussed the issues and challenges in implementing the biometric system for user authentication. They presented a variety of methods that monolithically combine a cryptographic key with the biometric template of a user stored in the database in such a manner that the key cannot be revealed without a successful biometric authentication. They assessed the performance of one of

these biometric key binding/generation algorithms using the fingerprint biometric. Moreover they illustrated the challenges involved in biometric key generation principally due to extreme acquisition variations in the representation of a biometric identifier and the imperfect nature of biometric feature extraction and matching algorithms. They sophisticated on the suitability of these algorithms for digital rights management systems. Experiments were conducted to explore the performance of there discussed methods in improving user security.

A Biometric Verification System was proposed by Cimato et al. in [12]. In their proposed work they presented a biometric authentication technique based on the combination of multiple biometric readings. The authentication control can be performed offline and the stored identifier does not disclose any information on the biometric traits of the identified person, so that even in case of loss or steal of the document, privacy is guaranteed. Their proposed approach ensures high level of security because of the association of multiple biometric readings. Biometric techniques are more and more exploited in order to fasten and make more consistent the identification process. The combination of cryptography and biometrics increases the confidence in the system when biometric templates are stored for verification.

Sunil et al. in [13] put forth a novel methodology for the secure storage of fingerprint template by generating Secured Feature Matrix and keys for cryptographic techniques applied for data Encryption or Decryption with the aid of cancelable biometric features. They proposed a technique to produce cancelable key from fingerprint so as to surmount the limitations of traditional approaches. Cryptography is merged with biometrics in Biometric cryptosystems, otherwise known as crypto-biometric systems [11]. They have introduced the concept of cancelable biometrics that was earlier proposed in [14]. Their approach facilitates the every incidence of enrollment to utilize a distinct transform thus making expose cross matching unachievable. Generally, the transforms utilized for distortion are chosen to be non-invertible. Thus it is not possible to recover the original (undistorted) biometrics despite knowing the transform method and the resulting transformed biometric data.

An effective authentication scheme by combining crypto with biometrics was projected by Hao et al. in [15]. They projected the first practical and secure way to integrate the iris biometric into cryptographic applications. A repeatable binary string, which we call a biometric key, is generated reliably from genuine iris codes. The key is generated from a subject's iris image with the support of auxiliary error-correction data, which do not disclose the key and can be saved in a tamper-resistant token, such as a smart card. The reproduction of the key depends on two factors: the iris biometric and the token. The attacker has to get hold of both of them to compromise the key. Moreover they evaluated the technique using iris samples from 70 different eyes, with 10 samples from each eye. As a result they found that an error-free key can be reproduced reliably from genuine iris

codes with a 99.5 percent success rate. One can generate up AES. The extraction of a repeatable binary string from biometrics opens new possible applications, where a strong binding is required between a person and cryptographic operations.

Apart from above mentioned works numerous researches has been done in this field of combining fuzzy and cryptographic key generation techniques [23, 24, 25]. Establishing the identity of a person is a critical task in any identity management system. Karthick Nandakumar et al. in [16] [17] showed the password hardened finger print fuzzy vault in which password acts an additional layer of security. This additional layer of security improves the security and privacy of users' biometric template data. The same concept mentioned in [16] was suggested to iris based hardened fuzzy vault scheme [17]. The approach discussed in [17] applies a sequence of morphological operations to extract minutiae points from the iris texture. Chen et al in [26] proposes the use of a Higher Order Spectral (HOS) Transform that can be applied to biometric data as a secure hash function. This HOS transform is non-invertible, is robust to noise in the input allowing it to tolerate the natural variations present in a biometric and can be made to produce a large number of significantly different outputs given an identical input.

III. OUR PROPOSED APPROACH

Our proposed methodology of fuzzy vault construction using retina as a soft biometric feature involves three steps. In the initial step the retinal template is subjected to undergo a random transformation. The approach makes use of the advantages provided by both the fuzzy framework and the soft biometrics, thereby enhancing the security and privacy. In the next step the obtained transformed template is secured with the assistance of constructing a fuzzy vault. The final step comprise of hardening the constructed fuzzy vault by encrypting the vault with the key randomly generated from soft biometric features and the user password. The password pretends as an additional layer of security. Fig 1 shows the soft biometric hardening of retina-based fuzzy vault scheme

A. Retinal Bifurcation Feature Point Extraction

The technique described by Chen et al. in [19] is utilized in this paper, for extracting the bifurcation feature points from retina. The retinal bifurcation points are extracted to improve the security and privacy of the user. The combination of soft biometrics characteristics and fuzzy vault scheme exploit the performance of the authentication system that was developed in recent years. In our approach the bifurcation feature of retina were obtained form vascular pattern of retina. The two major operations to be performed on the retinal template are thinning and joining operation, in order to extract the retina vascular pattern. As a result of this operation the bifurcation feature points are extracted from the retinal template. Fig .2 (a) represents the original retinal template. Fig.2 (b) shows the highlighted bifurcation feature points in a retinal vascular tree after performing thinning and joining operations

to 140 bits of biometric key, more than enough for 128-bit

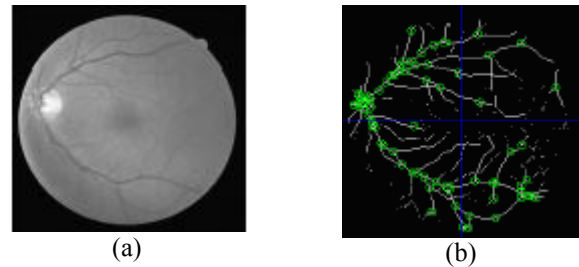


Fig. 2 (a). Original Retinal Template (b) Highlighted Bifurcation Feature

B. Hardening the retinal fuzzy vault using password

This is the significant step in the design of an authentication system. This makes use of the retinal template samples obtained from the database. The proposed system is implemented using MATLAB. The retinal samples that are obtained from the database are first resized as per our requirement. By highlighting the retinal bifurcation feature points the proposed method identifies the lock/unlock data. The bifurcation feature points are subjected to mathematical operation like permutation and translation using password. The principal requirement of this step is to achieve the three tuple parameters (u, v, θ) . In which u and v signifies the row and column indices respectively of the image found out and θ symbolizes the orientation parameter. These transformed feature points are then secured in the fuzzy vault using the 128 bit randomly generated key. A 64 bit user password is used to transform the randomly generated key. Additionally, the same can be used to encrypt the vault.

C. Transformation of Extracted Bifurcation Feature Points

As mentioned previously the retinal vascular tree holding the bifurcation points are destined to under go mathematical operations like permutation and translation. As a result of this process the original bifurcation points will get transformed into new points. There is a constraint on the number of characters used for user password. The user password is of 8 characters in length. Therefore a total of 64

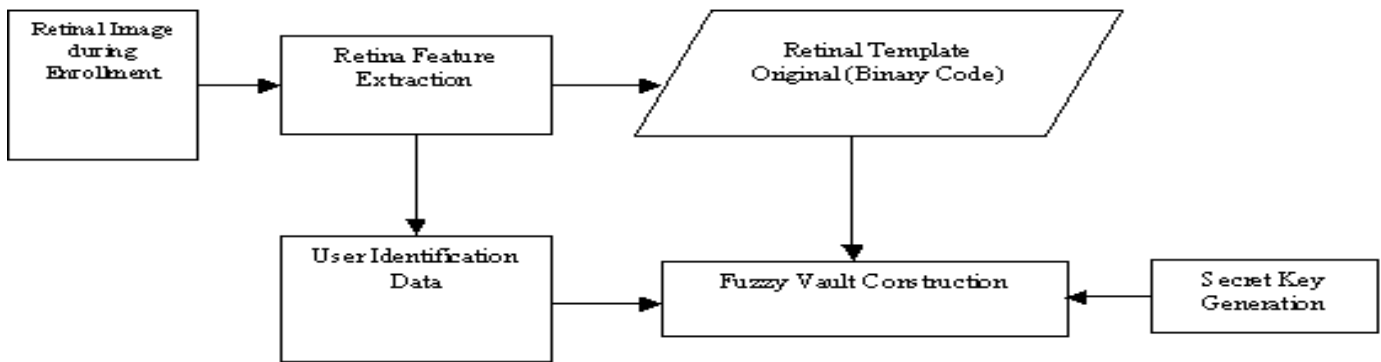


Fig.1 Soft biometric hardening of retina-based fuzzy vault scheme

bits are considered for randomization. These 64 bits are further divided into 4 blocks each block consisting of 16 bits. The first five characters resemble the password and the last three characters denote the soft biometric feature of the user. The five-character password used in our implementation is "FOKEN." The last three characters that indicate the user soft biometric characteristics are as follows. The sixth character denotes the height, the seventh stand for gender, and the eighth character resembles iris color of the user.

As an initial stage of implementation the bifurcation points are divided into 4 quadrants. Each quadrant is then processed with one password prior to permutation and translation operations. Care must be taken while applying the permutation operation. Note that there should not be any change in the relative position of the bifurcation points. The 16 bits of each quadrant is segmented into two bit block, one containing 9 bits and the other containing 7 bits. T_u denote the segment with 7 bits and T_v denote the segment with 9 bit length. T_u and T_v represents the amount of translation in the horizontal and in the vertical directions, respectively. Fig 3 shows transformed retinal bifurcation points.

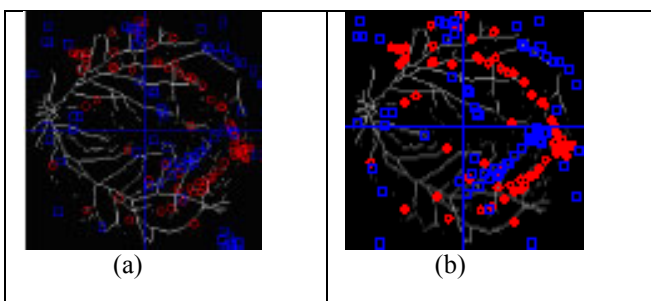


Fig 3. (a) and (b) Transformed bifurcation Points. (Blue-Transformed, Red-Original Points)

The transformation that is utilized to derive at the new retinal points is $X_u' = (X_u + T_u) \bmod (2^7)$

$$Y_v' = (Y_v + T_v) \bmod (2^9)$$

In which X_u and X_u' represents the horizontal distance before and after transformation respectively. In the similar manner, Y_v and Y_v' represents the vertical distance before and after transformation respectively.

D. Encoding the vault

This step secures the vault from being modified by an imposter from the knowledge of the password. The approach substitutes Reed-Solomon reconstruction step by Lagrange interpolation and cyclic redundancy check (CRC) based error detection. The obtained feature points are consistently quantized and articulated as binary strings. Therefore this can be represented as an element in Galois Field $GF(216)$. A large number of the chaff points are generated by the method mentioned in [4] [20]. Finally these chaff points are combined with the obtained feature points to make the imposter unaware of the genuine points in the retina.

E Decoding the vault

The user password is used to decrypt the encrypted fuzzy vault and the bifurcation feature points of the retina in this authentication phase. The helper data or a set of high curvature points are created in order to make possible the alignment of query minutiae to the biometric template. A transformation based on the password is implemented on the query feature points and the vault is unlocked.

IV. EXPERIMENTS AND RESULTS

The proposed work is implemented in MATLAB 7.0. The essential parameters used in this implementation are the number of chaff points (c), number of genuine points (r), and the total number of points ($r+c$). More the number of chaff points used, more is the privacy and security. It is remarkable that the number of chaff points introduced must be ten times the total number of genuine points that are available in the retinal template. The number of chaff points

TABLE I Bifurcation points before and after transformation

P ^t Quadrant and soft biometric features	Feature Points				Transformation code obtained from soft biometrics	
	Before Transformation		After Transformation			
	Horizontal Distance X_u	Vertical Distance Y_v	Horizontal Distance X_u'	Vertical Distance Y_v'	Row index with respect to horizontal axis T_u	Column index with respect to horizontal axis T_v
'VAULT' Height=157 Iris Color='B' Gender='M'	105	18	55	84	78	322

used determines the security and authentication provided by the developed system.

The revocability is evaluated by transforming the retinal (biometric) template for user password and soft biometric features. The proposed approach makes use of 8 characters to secure the vault as mentioned earlier. These 8 characters comprises of both the user password and the soft biometric characteristics of the user. The 8 characters can be grouped into two parts one containing the password of five characters. The sixth character denotes the height, the seventh and the eighth represents the gender and the color of iris respectively. Table 1 shows an example bifurcation points for one quadrant before the transformation and after performing the transformation for user password "VAULT", and user soft biometrics features namely height, gender and iris color.

The corresponding ASCII values of the 8 characters are utilized to secure the fuzzy vault. For the user password set as "VAULT" the corresponding ASCII values are determined as (86, 65, 85, 76 and 84). The remaining three characters are represented by the soft biometric features of the user. The value of the user height can be used as one parameter, and the remaining two ASCII values are calculated using the gender and the iris color of the user. With the change in the password variety of transformed templates can be obtained for same original biometric template. A variety of applications can use the soft biometric features with different passwords thus averting the cross matching.

V. CONCLUSION

As the decades pass by, improving the security and the privacy of the verification system is a challenging issue in recent years. Therefore, it is necessary to design a verification system that is more users friendly and secure. The proposed approach determines to combine the soft biometrics features and the cryptographic framework to develop a verification system that suits for a wide variety of applications. The biometric template that is taken into consideration in this approach is retina because of the advantage that

the retinal based genuine point determination pose a great challenge to all most all the attackers. Fuzzy vault is the framework which comprises of the combination of soft biometrics and cryptographic key generation technique. User password is used to improve the security and privacy of the authentication system. This password acts as an additional layer of security. Even if the password is compromised by an imposter it is hard to match the biometric template. Thereby, the security provided by biometric feature is not affected. In future, works to improve the performance of the vault can be carried out by applying non-invertible transformation and multiple biometric traits [21] [22]. This considerably reduces the failure to capture rate thus improving the performance of fuzzy vault

VI. REFERENCES

- 1) Jain, S. Pankanti, S. Prabhakar, L. Hong, and A. Ross, "Biometrics: A Grand Challenge", Proceedings of the International Conference on Pattern Recognition, Vol. 2, pp. 935-942, September 2004.
- 2) Wayman, A. Jain, D. Maltoni, and D. Maio, "Biometric Systems: Technology, Design and Performance Evaluation," Springer-Verlag, 2005.
- 3) Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, "Handbook of Fingerprint Recognition," Springer, 2003.
- 4) Juels, and M. Sudan, "A Fuzzy Vault Scheme", Proceedings of the International Symposium on Information Theory, p. 408, Lausanne, Switzerland, June 2002.
- 5) Juels, and M. Wattenberg, "A Fuzzy Commitment Scheme," Proceedings of the 6th ACM Conference on Computer and Communications Security, pp. 28-36, New York, NY, USA, 1999.
- 6) Tiago Santos, Gonalo Loureno, Lus Ducla Soares, and Paulo Lobato Correia, "Enhancing Biometrics Security," 2009.

- 7) N. K. Ratha, S. Chikkerur, J. H. Connell, and R. M. Bolle, "Generating Cancelable fingerprint templates," *IEEE Transactions on Pattern Analysis and Machine Learning*, vol. 29, no. 4, pp. 561-572, 2007.
- 8) Savvides, B. V. K. V. Kumar, and P. K. Khosla, "Cancelable Biometric filters for face recognition," *Proceedings of ICPR*, Vol. 3, pp. 922-925, Cambridge, UK, September 2004.
- 9) Sim Hiew Moi, Nazeema Binti Abdul Rahim, Puteh Saad, Pang Li Sim, Zalmiyah Zakaria, and Subariah Ibrahim, "Iris Biometric Cryptography for Identity Document," *IEEE Computer Society, International Conference of Soft Computing and Pattern Recognition*, pp. 736-741, 2009.
- 10) Abhilasha Bhargav-Spantzel, Anna Squicciarini, and Elisa Bertino, "Privacy preserving multi-factor authentication with biometrics," *Conference on Computer and Communications Security*, pp. 63-72, 2006.
- 11) U. Uludag, S. Pankanti, S. Prabhakar, A. K. Jain, "Biometric cryptosystems: issues and challenges," vol. 92, no. 6, pp. 948-960, 2004.
- 12) Stelvio Cimato, Marco Gamassi, Vincenzo Piuri, Roberto Sassi, and Fabio Scotti, "A Biometric Verification System Addressing Privacy Concerns," *International Conference on Computational Intelligence and Security (CIS 2007)*, pp.594-598, 2007.
- 13) Sunil V. K. Gaddam, and Manohar Lal, "Efficient Cancelable Biometric Key Generation Scheme for Cryptography," *International Journal of Network Security*, vol. 11, no. 2, pp. 57-65, 2010.
- 14) R. Ang, R. Safavi-Naini, and L. McAven, "Cancelable key-based fingerprint templates," *ACISP 2005*, pp. 242-252, 2005.
- 15) Feng Hao, Ross Anderson, and John Daugman, "Combining Crypto with Biometrics Effectively," *IEEE Transactions on Computers*, vol. 55, no. 9, pp. 1081-1088, 2006.
- 16) Karthik Nandakumar, Abhishek Nagar, and Anil K. Jain, "Hardening Fingerprint Fuzzy Vault Using Password," *International conference on Biometrics*, pp. 927 - 938, 2007.
- 17) Srinivasa Reddy, and I. Ramesh Babu, "Performance of Iris Based Hard Fuzzy Vault," *Proceedings of IEEE 8th International conference on computers and Information technology workshops*, pp. 248 - 253, 2008.
- 18) Karthick Nandakumar, Sharath Pankanti, and Anil K. Jain, "Fingerprint-Based Fuzzy Vault Implementation and Performance," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 4, pp.744 - 757, December 2007.
- 19) Li Chen, and Xiao-Long zhang, "Feature-Based Image Registration Using Bifurcation Structures," *Matlab Central*.
- 20) K. Jain, L. Hong, and R. Bolle, "Online Fingerprint Verification," *IEEE Transaction on Pattern Analysis and Machine Learning*, vol. 19, no. 4, pp. 302-314, April 1997.
- 21) Jain, Anil K. Jain and Arun Ross, "Multibiometric systems," *Communications of the ACM*, January 2004, Volume 47, no. 1, 2004.
- 22) K. Jain and A. Ross, "Learning User-specific parameters in a Multibiometric System," *Proceedings of IEEE International Conference on Image Processing (ICIP)*, Rochester, New York, pp. 57 - 60, 2002.
- 23) Monrose, M. K. Reiter, Q. Li, and S. Wetzel, "Cryptographic Key Generation from Voice," in *Proceedings IEEE Symposium on Security and Privacy*, Oakland, pp. 202-213, May 2001.
- 24) Dodis, L. Reyzin and A. Smith, "Fuzzy Extractors: How to generate Strong Keys from Biometrics and other Noisy Data," in *Proceedings of International Conference on Theory and Application of Cryptographic Techniques*, pp. 523-540, May 2004.
- 25) Jain, A. K., Nandakumar, K., and Nagar, A. Biometric template security. *EURASIP J. Adv. Signal Process* 2008, Jan 2008.
- 26) Chen, Brenden Chong and Chandran, Vinod Biometric template security using higher order spectra. In: *International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2010*, 14-19 March 2010

Performance Improvement of UDP and TCP Traffic by CJM Algorithm in Voice Transmission

IKram Ud Din¹ Saeed Mahfooz² Muhammad Adnan³

{ GJCST Computing Classification

C 2.5, C 2.2, C 4 }

Abstract-Different protocols are used at transport layer of the OSI model to transmit real time applications (i.e., video and voice). In this paper, two networks were analyzed in which the Transmission Control Protocol (TCP) and the User Datagram Protocol (UDP) were used for transportation. The purpose of this study is to evaluate the performance of TCP and UDP in real time transmissions for the parameters: Traffic Received, End-to-End Delay, and variation in delay (Jitter). At the receiver side, an algorithm called Chunk-based Jitter Management (CJM) [19] is applied on the buffer, which divides the packets into small size chunks and then plays them out. The evaluation of these transmission protocols has been done in OPNET. The graphs show the results in which the CJM algorithm performs well in end-to-end delay, jitter in voice packets, and voice packets receiving.

Keywords: TCP, UDP, Transmission Protocol, Delay, Jitter

I. INTRODUCTION AND RELATED WORK

The most vital transport protocol used nowadays is TCP. However, the vigorous transportation protocol for the streaming media is UDP. Two prime reasons that UDP is not being used commonly are: i). some organizations are blocking this protocol and ii). it is not friendly to other flows. In the meantime, TCP is obviously reliable and friendly to other flows, but with so many basic controls in the protocol such as flow control, congestion control, and others with the heavy acknowledgement mechanism, resulting jitters and delays. Thus it is unsurprisingly not friendly to the real time application [1]. Recent years have witnessed increasing demand for multimedia information services and explosive growth of the Internet. Transmission of the real time applications via the Internet has received magnificent attention. In term of transport protocol, the hereditary problems are lacking of variability in bit rates, throughput guarantees, jitters or delays, and packet loss [2], [3]. Those characteristics are not “friendly” to the real time data. Real time applications are able to compromise packet losses but sensitive to packet delays. Conventional perception holds that UDP is a better protocol than TCP for the transmission of critical data [4], [5]. This perception is straightforward to be understood because UDP is a best-effort delivery service. Theoretically, there will be less delay and provides better throughput. Unfortunately, this best-effort transport protocol potentially hinders the performance of other applications that employ TCP, or worse, jeopardize the stability of the Internet [6]. On the other hand, TCP

employs congestion control schemes that vary dynamically to network conditions, and thus, it often yields jitters and packet delays [7]. Its reliability is naturally unsuitable for critical data (i.e., video and voice) [3]. Although the above-mentioned grounds seem to indicate that the present transport protocols, namely UDP and TCP, are not appropriate for real time applications. Many researchers also found that TCP is a more fashionable transport protocol than UDP [8, 9]. The researchers have justified their views by giving a reason on congestion control mechanism used by TCP regulates rate as a function of the number of packets sent by the application [1]. Other studies have found that video clips on the Internet nowadays are encoded at bit rates of 89-300 Kbps [10], [11]. The other interesting studies have also found that TCP lost packet recovery mechanism or retransmission is not too severe for real time applications [12]. In recent years, Voice over IP (VoIP) has gained a lot of popularity. Session Initiation Protocol (SIP) developed by IETF for VoIP signaling is a communication control protocol used with transport layer's protocols, e.g., TCP and UDP. Today's SIP applications are mostly operating over the unreliable and connectionless transport protocol, UDP. Because SIP establishes connection and TCP is also a connection-oriented protocol which brings delay and jitter in real time transmission, therefore, UDP is a suitable protocol for video conferencing [13].

The remaining paper is organized as: Section II describes the structure of the network. Section III illustrates architecture of the proposed system. The OPNET simulations and results are given in section IV. Section V concludes our work. The future work is described in section VI, and references are given in section VII.

II. STRUCTURE OF THE NETWORK

In the given scenario (Fig. 1), a network is established between the two cities of Pakistan, i.e., Karachi and Lahore. The scenario consists of servers and clients. One server and a client are located at each site. The servers are named as VoIP_Karachi and VoIP_Lahore. The simulator used in this work is OPNET Modeler 14.0. The packets were sent and received for 10 minutes from one place to another. Once the network was tested for the normal flow, and second time the CJM algorithm was applied on the receiver's buffer. The networks were named as Normal_Flow for normal transmission of the data, and Chunk_based for the network on which the CJM algorithm was applied as in [14]. Routing Information Protocol (RIP) is the protocol implemented for routing on both side routers.

About -¹Department of Computer Science, University of Peshawar, Pakistan (e-mail-ikramuddin205@yahoo.com¹, saeedmahfooz@yahoo.com², adnan5283@yahoo.com³)

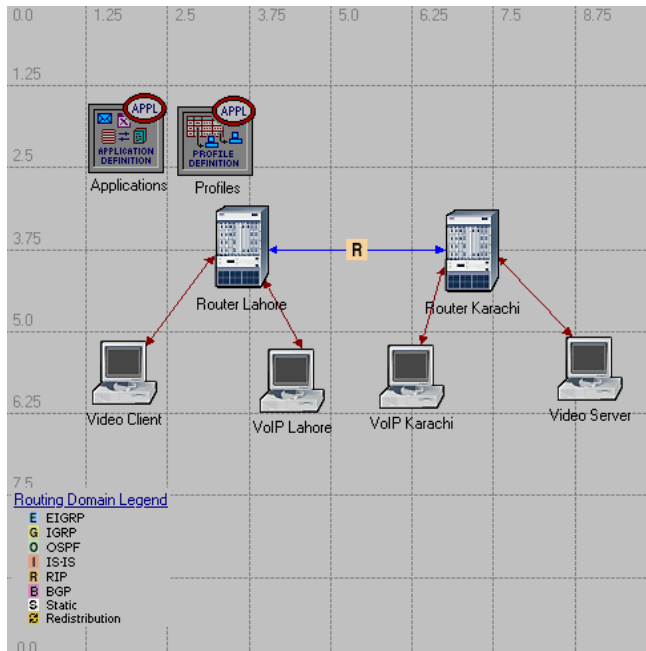


Fig. 1: Structure of the Network

Traffic Receive

Video or voice traffic is the total number of audio and video packets received during video conferencing or other type of real time communication (e.g., IP telephony) [14].

End-to-end Delay

End-to-end delay refers to the time taken by a packet to be reached across a network from source to the final destination. End-to-end delay depends on the end-to-end data paths/signal paths, the CODEC, and the payload size of the packets. Delay is the latency, one-way or round-trip, encounters when data packets are transmitted from one place to another. In order to maintain the expected voice quality for VoIP, the roundtrip delay must remain within approximately 120 milliseconds. [15], [14].

Jitter

In the context of voice over IP, jitter is the variation in delay of packets received, caused by network congestion or route changes. Jitter is a vital quality of service (QoS) factor in evaluation of network performance. It is one of the significant issues in packet based network for real time applications [16]. The variation of interpacket delay or jitter is one of the primary factors that agitates voice quality [17]. Jitter plays a vital role for the measurement of the Quality of Service of real time applications. The effect of end-to-end delay, packet loss, and jitter can be heard as: The calling party says –Good morning, everybody!” With end-to-end delay, the called party hears –.....Good morning, everybody!” With packet loss, the called party hears –Go....od... mor... ng ery body!” With jitter, the called

party hears –Good...morning, eve.....ry... body!” [18], [14].

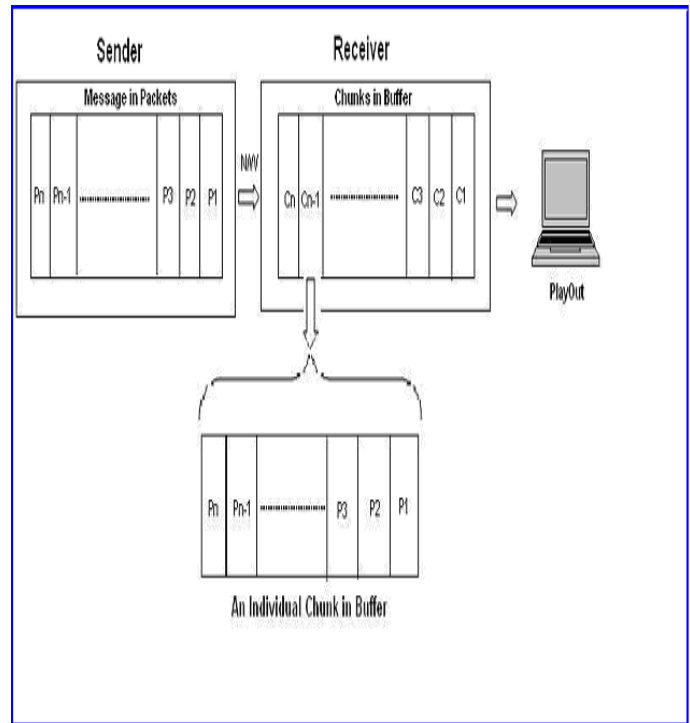


Fig. 2: Architecture of the Chunk-based Jitter Management Technique

III. SYSTEM'S ARCHITECTURE

Architecture of the CJM system comprises of three major components, i.e., sender, receiver, and network that are shown in fig 2.

A. Chunking of Packets

A voice message is mostly tightly coupled within a group of consecutive packets. Therefore, the proposed algorithm suggests chunking of the packets at the destination in the buffer. Thus, playing of a chunk, mostly, conveys a complete message without something missing. Hence, the QoS is improved while playing it out. All chunks are of the same size.

B. Algorithmic Outlines

The informal description of the proposed chunk-based jitter management algorithm in its pseudocodal form is given below [19]:

```
[Reading packets from network]
Read packets from the network and store in the buffer at the receiver.
[Chunking]
    Group the packets into same size chunks.
[PlayOut Chunks]
    Read chunks until buffer is empty i.e.
    do
    [Read Chunk]
    Read chunk from the buffer.
```

```

For (PacketCounter=1; PacketCounter<=
ChunkSize; PacketCounter++)
[PlayOut Packets]
StreamOut Packets
End of For
    ChunkCounter++;
While (ChunkCounter <= BufferSize)
[Store new Packets in the buffer from network]
Repeat step 1 to 3
[Stop]
Exit

```

IV. OPNET SIMULATION AND RESULTS

In this part, a scenario was examined in which jitter, delay, and packet receive rate were observed. The number of UDP and TCP packets received in Normal_Flow network is shown in figure 3. Figure 4 illustrates the voice traffic received when the CJM algorithm is applied on the receiver buffer. Jitter in the Normal_Flow network is given in figure 5, while it is shown in figure 6 when the CJM algorithm is applied on the buffer. The end-to-end delay in the

Normal_Flow and Chunk-based networks is shown in figure 7 and 8, respectively.

A. Performance Evaluation

The number of voice traffic received in the Normal_Flow and Chunk-based networks are shown in figure 3 and 4, respectively. In the Normal_Flow network, there is a slight difference between the packets received through UDP and TCP, but a huge difference was observed when the CJM algorithm was applied on the buffer. The voice packet delay variation or jitter was also minimized through CJM algorithm (figure 6) as compared to the normal flow of data (figure 5). The end-to-end delay in TCP and UDP is almost the same in the Normal_Flow network as shown in figure 7, but it is quite improved in the chunk-based network. In the given diagrams, the X-axis shows the amount of simulation time and the Y-axis shows the number of packets per second in figure 3 and 4, while the value of jitter in seconds in figure 5 and 6, and the value of delay in seconds in figure 7 and 8, respectively.

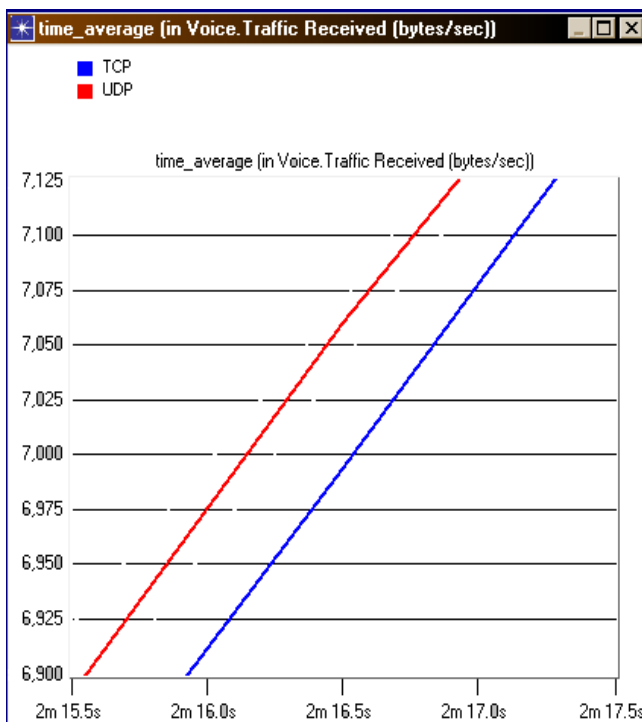


Fig. 3: Voice traffic received in Normal Flow

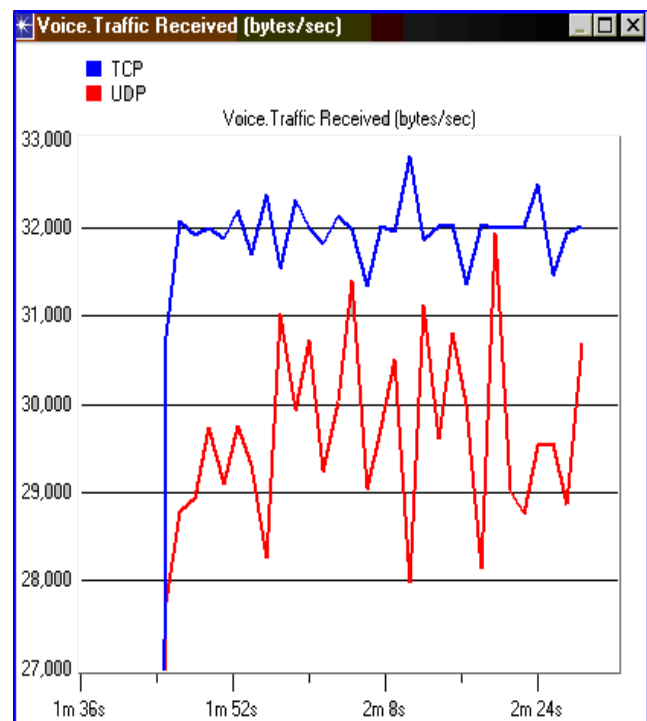


Fig. 4: Voice traffic received when CJM algorithm is applied on the buffer

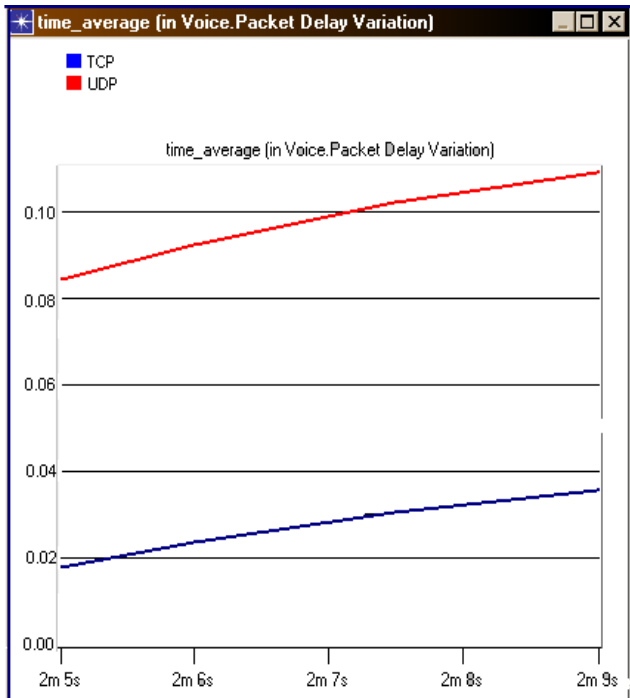


Fig. 5: Voice Packet delay variation in Normal Flow

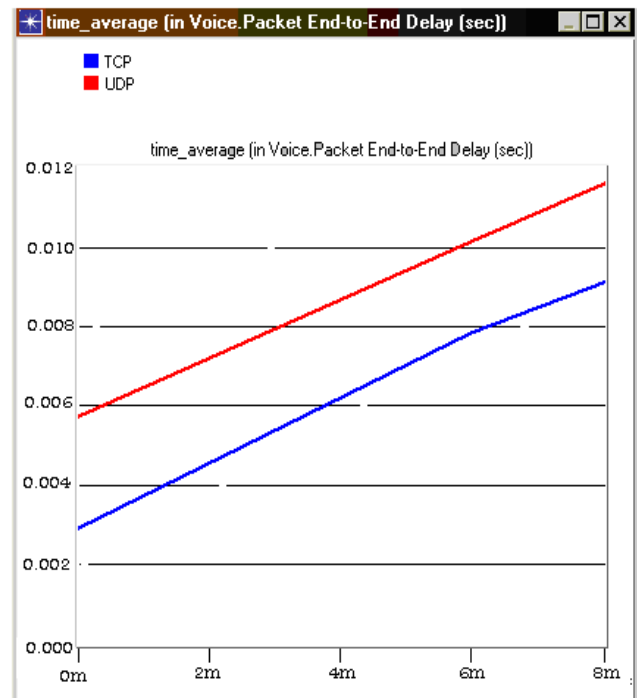


Fig. 7: Voice Packet end-to-end delay in Normal Flow

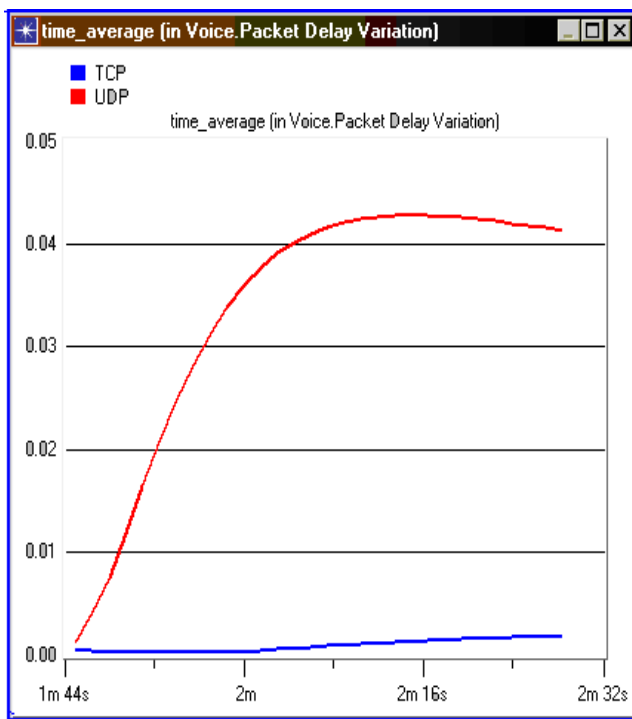


Fig. 6: Voice Packet delay variation after applying CJM algorithm on the buffer

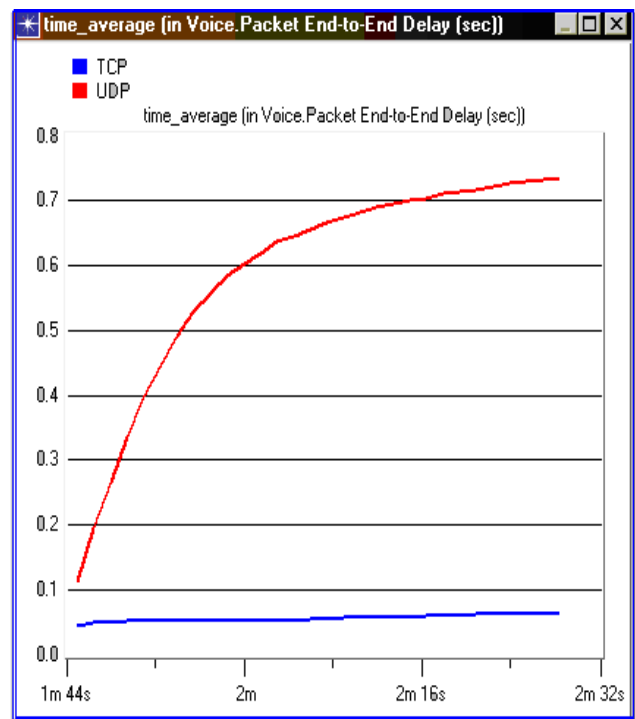


Fig. 8: Voice Packet end-to-end delay after CJM algorithm is applied on the buffer

V. CONCLUDING REMARKS

Today's networks size has been growing rapidly and support complicated applications, e.g., voice messages and video conferencing. Quality transmission is demand of the time. This needs some good results during transmission produced by transport protocol. The work done in this paper evaluates the available transport protocols: UDP and TCP for traffic receiving, jitter, and end-to-end delay. Our work for each of these parameters is based on OPNET simulation. The study presents a comprehensive result both for TCP and UDP against the parameters traffic received, jitter, and end-to-end delay one by one. After arriving packets at the destination, the Chunk-based Jitter Management (CJM) algorithm is applied on the buffer. The simulation results show that CJM algorithm performs better as compared to the normal flow of data

VI. FUTURE WORK

As for feasible future work, we are planning to carry on with the implementation of Stream Control Transmission Protocol (SCTP) which has both the qualities of UDP and TCP, and combine it with the existing cooperative transport protocols. In this way, we will implement accurate cooperative mechanisms that will further improve network performance.

VII. REFERENCES

- 1) Suki M. Arif, O.G., Suhaidi Hassan, A Survey on Buffer and Rate Adaptation Optimization in TCP-Based Streaming Media Studies, in International Conference on Network Applications, Protocols and Services. 2008, Executive Development Center: Malaysia.
- 2) M. Prangl, I.K. and H. Hellwagner, Towards QoS Improvements of TCP-Based Media Delivery, in Networking and Services (ICNS 2008). 2008. p. 188-193.
- 3) H. Sun, A. Vetro. and J. Xin, An Overview of Scalable Video Streaming: Research Articles. Wireless Communications & Mobile Computing, 2007. 7: p. 159-172.
- 4) E . Brosh, S. A. Baset, D. Rubenstein, and H. Schulzrinne, The Delay-Friendliness of TCP, in SIGMETRICS. 2008.
- 5) S. Boyden, A.M. and C. Williamson, TCP Vegas Performance with Streaming Media, in Performance, Computing, and Communications Conference (IPCCC 2007). 2007. p. 35-44.
- 6) S. Floyd and K. Fall, Promoting the use of end-to-end congestion control in the Internet. IEEE/ACM Trans. Netw, 1999. 7: p. 458-472.
- 7) Wang, W. Wei., et al, Application-layer multipath data transfer via TCP: Schemes and performance tradeoffs. Perform. Eval, 2007. 64: p. 965-977.
- 8) L. Guo, E. Tan, S. Chen, Z. Xiao, O. Spatscheck, and X. Zhang, Delving into internet streaming media delivery: a quality and resource utilization perspective, in 6th ACM SIGCOMM conference on Internet measurement. 2006, ACM: Rio de Janeiro, Brazil. p. 217-230.
- 9) Wang, J. Kurose, P. Shenoy, and D. Towsley, Multimedia streaming via TCP: an analytic performance study, in SIGMETRICS: Perform. Eval. Rev. 2004. p. 406-407.
- 10) M. Li, M. Claypool., R. Kinicki, and J. Nichols, Characteristics of streaming media stored on the Web. ACM Trans. Inter. Tech, 2005. 5: p. 601-626.
- 11) L. Guo, S. Chen., Z. Xiao, and X. Zhang, Analysis of multimedia workloads with implications for internet streaming, in 14th international conference on World Wide Web. 2005, ACM: Chiba, Japan. p. 519-528.
- 12) Loguinov and H. Radha, in INFOCOM. 2002: Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies.
- 13) Rosenberg, J., et al, SIP Session Initiation Protocol, in KFC3261, IETF, Editor. 2002.
- 14) IU. Din, S. Mahfooz, and M. Adnan, Analysis of the Routing Protocols in Real Time Transmission: A Comparative Study. Global Journal of Computer Science and Technology, 2010, Accepted for Publication.
- 15) Paul J. Fong, E. Knipp, et al, Configuring Cisco Voice Over IP, Second Edition.
- 16) IC. Demichelis and P. Chimento, IP Packet Delay Variation Metric for IP Performance Metrics (IPPM). Request for Comments: 3393, 2002.
- 17) Pedrasa, J.R.I. and C.A.M. Festin, An Enhanced Framing Strategy for Jitter Management, in TENCON 2005 IEEE Region 10. 2005: Melbourne, Qld. p. 1-6.
- 18) Cisco Systems, Inc, Cisco Voice Over IP. Student Guide, Edition V. 4.2. 2004.
- 19) IU. Din, S. Mahfooz, and M. Adnan, —AChunk-based Jitter Management Algorithm for Voice Transmission”, Global Journal of Computer Science and Technology, 2010. Accepted for Publication.

Fuzzy Set in Business Process Management as Reference Model

{ GJCST Computing Classification
1.2.3, J.1 }

Mr. Yugal Kumar¹ Mr. Samit Yadav² Rahul Yadav³

Abstract-The Central theme of this article or paper is to use the benefits or advantage of the Fuzzy set theory in Business to improve process, quality and extension of the business. In business the enterprise-specific for process improvement is characterized by decision-making premises. The decision does not exist in the form of mathematic models or numeric values. Decisions are characterized as thoughts, consideration and creativity. These are usually derived from fuzzy conditions such as low processing time and high quality. Although these conditions are not precise or accurate, but these conditions consist additional and important information for the understanding of concrete business situations that are related to the business or any process. Thus the verbal information as well as vaguely formulated statements, premises, objectives and restrictions are very important for reference model adaptation. The systematic consideration of fuzzy set or data in reference model adaptation can only succeed when the models to be allow the consideration of fuzzy data. The fuzzy set theory-based extension of information modeling therefore provides the foundation for the development of a methodology, as well as the realization of a tool for reference model adaptation with regard to fuzzy data in this article.

I. INTRODUCTION

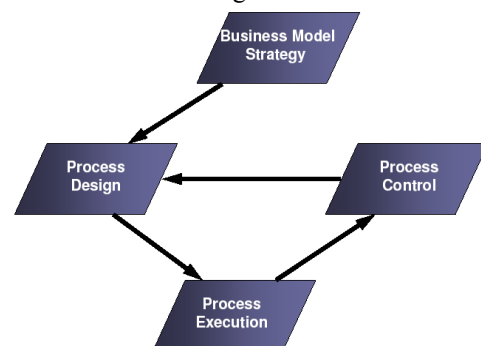
While companies try to improve key business processes to maximize their own values, business processes are mostly managed based on experts' or leaders' experiences in business practice. Non-value added processes are sometimes overcontrolled. In this environment, a structured framework for a systematic BPM is required. However, there is a surprising lack of an overall framework to support improvements based on the data about the performance of each business process. In service industries such as life insurance, BPM is more salient than in the manufacturing industry. The process of acquiring and receiving services is a product itself in the service industry.

There are some difficulties in applying BPM in a service industry

- 1) It is difficult to define processes and their flows. Flowcharts and process maps are hardly used in the service industry.
- 2) It is hard to measure process performance.
- 3) Some noisy or uncontrollable factors such as customer behavior influence service processes.

I. FUZZINESS IN BUSINESS PROCESS MANAGEMENT

The goals in current business engineering projects lay in the design of business processes and the analysis of requirements for their IT-support with regard to corporate strategies. Process design must follow a comprehensive approach, comprising planning and control, i. e. the management of operational processes. Modeling has proved to be helpful for the support of systematic procedures in process design. Modeling languages like the event-driven process chain (EPC) serve as an operationalized approach to model construction. Software tools for business process modeling support the business engineer with system components for the analysis, design and simulation of business process models. Many concepts that consider situation-specific problems have been developed for the collection and improvement of business processes, their generalization in reference models and their enterprise-specific adaptation in customizing. Many of these approaches focus on the user-friendly and intuitive usability of methods by modeling them on human ways of thinking. More important for making the required decisions are however, the exact quantification and formalization of decision rules. However, in many cases, only uncertain, imprecise and vague information about the often not technically determined procedures is available for business processes. This circumstance will be met here by extending process modeling through the consideration and processing of fuzziness using the fuzzy-set-theory. This fuzzy extension will be reproduced with the EPC. The EPC was chosen as a process modeling language due significantly to its popularity in modeling practice. We will specify the term "fuzziness" and motivate the consideration of fuzzy data using the fuzzy set theory. The life cycle of Business Process is mentioned in diagram.



About-1 Lecturer in CSE Dept., SDITM, Balana, Panipat

About-2 Senior Lecturer in CSE Dept., PDMCE, Bahudurgarh

About-3 Lecturer in CSE Dept., PDMCE, Bahudurgarh

Life Cycle of Business Processes

II. CRISP TO FUZZY SETS

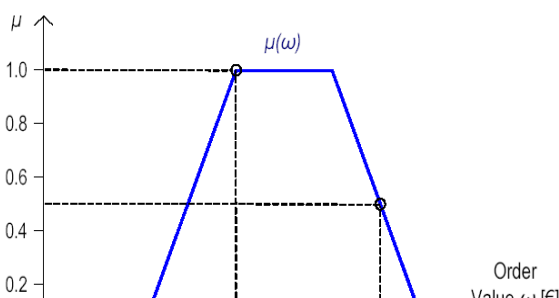
There is no standard definition for the term “fuzziness”. It almost seems as if the understanding of the term itself must remain fuzzy. Fuzziness is usually defined by way of differentiation with deterministic, stochastic and uncertain states of information. In this article, fuzziness is seen as uncertainty with regard to data and its interdependencies. Different reasons for fuzziness can be identified in the business context. First, fuzziness occurs due to the complexity of the environment and the limits in human perception when comprehending reality. The resulting informational fuzziness, determined by human language and thought, can be ascribed to a surplus of information. This happens when terms with a high level of abstraction are used. Thus for example, knowledge intensive processes contain short-lived information from a number of sources, which results in the fact that only one part of the total process can be covered at one point in time. This part however already becomes dated during the coverage of other sub-aspects. Many different attributes must be considered for the description of such complex terms. Fuzziness occurs because often, man is not capable of processing all of the relevant information and because, perhaps even the individual pieces of information are themselves already fuzzy. The descriptive attributes of the term are aggregated according to human information processing using linguistic terms. Fuzziness also exists in human preference and goal conceptions. In many situations, preferences human preference orders cannot be determined exactly.

III. FUZZY SET THEORY

Fuzzy set theory is the sub domain of Soft Computing i.e. Fuzzy Logic + Neural Networks + Evolutionary Computing + Probabilistic reasoning . The crucial point in the fuzzy theory is that it is not only to evaluate conditions of objects with “true” or “false”, but also rather to allow “intermediate stages”. So, the subsequent to Zadeh’s original idea, the classic set theory, i. e. the theory of crisp sets, is extended by the description and combination of fuzzy sets. The grade of membership for each element ω of a predetermined (crisp) basic set w to a subset A that’s belong to w , is expressed by a value $\mu_A(\omega)$ of a mapping

- 1) $\mu_A : w \rightarrow [0;1]$.
- 2) μ_A is called the membership function of the fuzzy set $\{(\omega; \mu_A(\omega)) \mid \omega \in w\}$.

The value of membership function is lies between 0 and 1. Which is defined with the help of graph as below:-

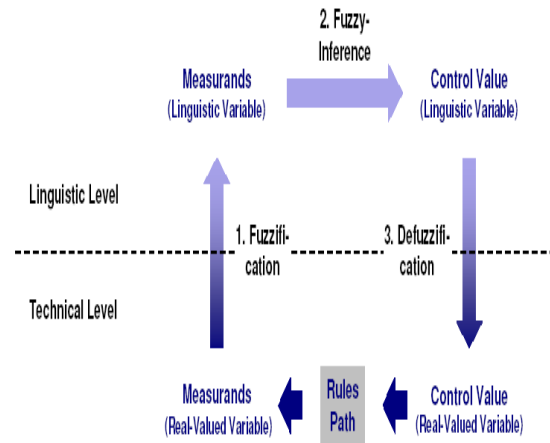


Membership Function for fuzzy set

IV. FUZZY SYSTEMS

A fuzzy system has a fixed set of input and output variables, whose respective terms are connected with fuzzy rules consisting of a condition and a conclusion part. For example “WHEN customer assessment = middle AND order value = very high THEN order assessment = high”. The value domains of the (linguistic) variables are partitioned by fuzzy sets, which serve the representation of the linguistic terms.

We can describe the fuzzy system as given below:-



Level or Working in Fuzzy System

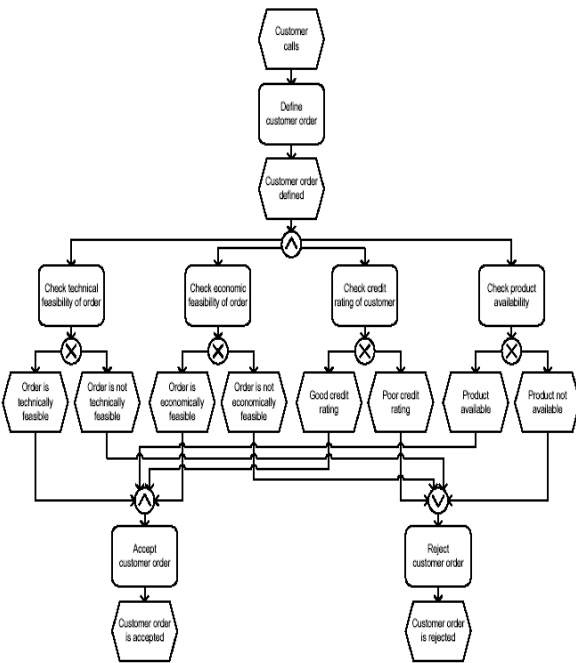
Fuzzification:- Representation of crisp values on fuzzy sets.
 Fuzzy-Inference: -A fuzzy system consists of input and output variables, whose respective attributes are connected by rules, consisting of a premises and a conclusion. The input and output variables are assigned to one another by way of an inference procedure.
 Defuzzification:- For an executable action, for example: “set priority”, a crisp value from the output variable is required. A defuzzification step provides this crisp value.

V. REFERENCE MODEL ADAPTATION WITH REGARD TO FUZZINESS

The fundamental idea followed here states that the systematic consideration of fuzzy data in the adaptation of reference models can only be successful when the models to be adapted themselves allow the consideration of fuzzy data. The fuzzy theory-based extension of information modeling is therefore the foundation for the development of a methodology, as well as for the prototypical realization of a tool for reference model adaptation under consideration of fuzzy data. The fuzzy theory makes the representation of the decision-logic based on the experience of those responsible for the business processes possible. By taking fuzzy conditions and vaguely formulated objectives into consideration, the user with technical process knowledge should himself be able to carry out the enterprise-specific adaptation of reference models using intuitive and simple linguistic evaluations. The adaptation-tool should like us

humans and make decisions on the basis of fuzzy terms. The following section justifies the consideration of fuzzy data in reference modeling and points

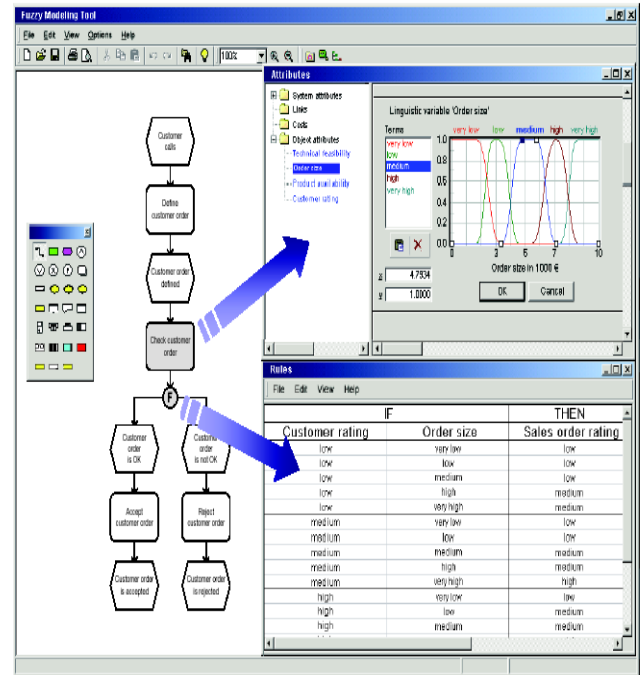
out its application potential using a simple example process for customer order processing. The following figure represents a part of a reference process for customer order processing in the form of an Event-driven Process Chain (EPC). The model describes the course of events for the definition and execution of the checking functions for a customer order. The decision regarding the acceptance or the refusal of the customer order is made by the parallel execution of various sub-functions. The customer order is checked for technical feasibility and in addition, the customer creditworthiness and the availability of the product are determined. Negative results, such as for example, "Order is not technically feasible" or "Poor credit rating", lead to the rejection of the customer order by way of the function "Reject customer order". The Reference Model for Customer Order processing is given below.



Reference Process Model for Customer Order Processing

A weak point in the modeled process immediately becomes obvious: each of the negative results leads to the immediate rejection of the customer order independent of the inspection results of the other functions. This is contradictory to business practice where such absolute elimination criteria are only rarely complied. In fact, through man as the decision-maker implicit compensation mechanisms are used, which counter-balance an exceedance of limiting values in one area with better values in another area. The rules for the interdependent impact are not documented here, but rather based upon the decision-makers know-how. Furthermore, it is usually a case of simple rules, which establish only scale-related combinations and which orient themselves on target systems with vague interdependences. In the present case, the decision as to

whether a product is available could be answered not only with a crisp "Yes" or "No", but rather also be characterized by the additional effort resulting from weighing things up, so that the product for example, could be requested from another warehouse, if all other inspections turned out to be positive. A corresponding decision orients itself on the trade-off between the goal to avoid additional costs and the focus on customer needs. This results in the challenge to represent fuzziness in reference and procedure models for their adaptation, in addition to the problem of the development of implicit knowledge



Reference Process Model for Customer Order Processing with Fuzzy System

VI. ADVANTAGE OF FUZZY SYSTEM

- 1) Only minor differences between output of simple fuzzy systems and complex crisp model.
- 2) Flexibility of Fuzzy-Systems by robustness and adaptability.
- 3) Similar to human way of thinking
- 4) If-Then-Rules easy to understand
- 5) Rule base can be maintained by user

A Fuzzy System can support a business process at least as good as a crisp system with less efforts and in a user centered way

VII. FUTURE CHALLENGES.

- 1) Extension of process and organisational aspects through fuzzy technology.
- 2) Formalisation of the fuzzy extension.
- 3) Establishing standards to include fuzzy-values in process description.
- 4) Implementation of a tool for fuzzy enterprise modeling.

VIII. CONCLUSION

The manageability of the adaptation of reference models finds itself in the tug-of-war between theoretical foundation and pragmatic simplicity and displays a high degree of complexity in practice. To reduce this complexity a modeling approach, allowing the consideration of fuzzy data and its possible usage has been outlined in this article. Business process models are limited to the content required by the end-user for the comprehension of the logic of business processes, while the technical knowledge necessary for the decision support of individual model-elements is deposited elsewhere. The fundamental idea followed states that the systematic consideration of fuzzy data in the adaptation of reference models can only be successful when the models to be adapted themselves allow the consideration of fuzzy data.

IX. REFERENCES

- 1) [1] Adam, Thomas O.: A Fuzzy Based Approach to the Improvement of Business Processes.
- 2) Adam , Thomas O. & Loos P.: Soft Business Process Intelligence .
- 3) [3] Benedicenti, . Succi L, Vernazza G., Valerio : Object Oriented Process Modeling with Fuzzy Logic.
- 4) Bellman, R. E.; Zadeh, L. A.: Decision-Making in a Fuzzy Environment. In: Management Science 17 (1970), no. 4, pp. B141-B164.
- 5) Blechar, M. J. & Sinur, J. (2006). Magic Quadrant for Business Process Analysis Tools, 2006.
- 6) [6] Dehnert J. & Rittgen P. : Relaxed Soundness of Business Processes.
- 7) Scheer: Business Process Engineering : A Reference Models for Industrial Enterprises. (7th ed.).
- 8) Zadeh, L. A. : Fuzzy sets. In: Information and
- 9) Control 8, vol. no. 3, pp. 338–353.
Zadeh, L. A. : Outline of a new approach to the analysis of complex systems and decision processes.

Impact of Aspect Oriented Programming on Software Development Quality Metrics

Kotrappa Sirbi¹ Prakash Jayanth Kulkarni²

GJCST Computing Classification
D 2.3, D 2.m, K 6.4, K 6.3

Abstract-The aspect-oriented programming (AOP) is a new paradigm for improving the system's features such as modularity, readability and maintainability. Owing to a better modularisation of cross-cutting concerns, the developed system implementation would be less complex, and more readable. Thus, software development efficiency would increase, so the system would be created faster than its object-oriented programming (OOP) equivalent. In this paper, we provide some insight into the OO software development quality metrics were significantly associated with using AOP. The method that we are currently studying is based on a popular C & K metrics suite that extends the metrics traditionally used with the OO paradigm and also extend to AO paradigm. We argue that a shift similar to the one leading to the Chidamber and Kemerer's metrics is necessary when moving from OO to AOP software.

Keywords- Aspect Oriented Programming (AOP), Aspect Oriented (AO) system, AO metrics, AspectJ.

I. INTRODUCTION

THE past decade has seen the increased use of Aspect Oriented Programming (AOP) based software development techniques as a means to modularize crosscutting concerns in software systems, thereby improving a development organization's working practices and return on investment (ROI). Numerous industrial-strength aspect-oriented (AO) programming frameworks exist, including AspectJ, JBoss, and Spring, as do various aspect-oriented analysis and design techniques. The "Major Industrial Projects Using AOP" are many notable applications, of which the most prominent is the IBM WebSphere Application Server. Developers considering AOP techniques must ask three fundamental questions:

- How is AOP being used in industrial projects today?

Developers must determine whether AOP techniques are suited to the problem at hand and the particular project context.

Does the improved modularity yield real benefits when engineering and evolving software?

Developers must understand whether the potential benefits outweigh the costs of introducing a new technology and, if so, be able to convince management of its long-term profitability.

- What do developers need to be aware of when using AOP techniques?

Developers must avoid known pitfalls and deploy design strategies and tools to help counter their potential threat to product quality.

Answers to these questions are not readily available, and narrowing knowledge from existing literature on the topic is difficult, but there is some insight by working with many several medium and large-scale open source projects employing AOP techniques. Much AO adoption shows that software development projects mainly rely on basic features of AO languages to modularize well-known crosscutting problems; developers introduce AOP concepts incrementally, initially addressing developmental concerns and not core product features. In addition, AOP techniques improve design stability over a system's evolution and can substantially reduce design model size [1].

The aspect oriented programming (AOP) is a relatively recent approach that has been argued to better enable modularization of crosscutting concerns [2] and consequently accelerate the development process. The hypotheses are that well separated concerns are more easily maintained, changed and developed, so the total programmer's working time should be shorter than the development time of analogous system, realized without mechanisms offered by AOP. The validation of these hypotheses requires empirical studies. Many researchers in literature present results of preliminary empirical evaluation of the impact of AOP on software development efficiency and design quality. This paper includes a comparison of developed AOP and OOP systems, based on software metrics proposed by Chidamber and Kemerer (hereafter CK) [3], Distance from the Main Sequence metric proposed by Martin [4], external code quality metric (defined as a number of acceptance tests passed) [5, 6, 7], and programmers' productivity metric. CK software metrics [3] were adapted to new properties of aspect-oriented software [8, 9].

Subramanyam and Krishnan state that research on metrics for object oriented software development is limited, and empirical evidence, linking the object-oriented methodology and project outcomes, is scarce [10]. Even more scarce is empirical evidence of the effect of aspect-oriented programming on software design quality, or development efficiency metrics. Therefore, the aim of this paper is to fill this gap and provide empirical evidence of the impact of aspect-oriented programming on software development efficiency and design quality metrics, as design aspects are extremely important to produce high quality software [10]. The hypothesis that design quality metrics are good predictors of the fault proneness is supported in [11] and [12].

About-¹Kotrappa Sirbi, Department of Computer Science & Engineering, K L E's Dr.M.S.Sheshagiri College of Engineering & Technology, Belgaum, India.

About-² Prakash Jayanth Kulkarni, Department of Computer Science & Engineering, Walchand College of Engineering, Sangli, India.

The rest of the paper is organized as follows, Section II provides a related work in the field of OO and AO metrics, Section III give a brief overview of Aspect-Oriented Programming(AOP), Section IV explaining importance of OO software metrics in OOD, Section V explaining requirements of AO software metrics, Section VI explains the potential effect of AO on the C&K metrics and Section VII implementation of case study AJHotDraw and Section VIII shows the impact of AO metrics on AJHotDraw and includes the results of the experiment. The conclusion of the paper is presented in Section IX.

II. RELATED WORK

The literature available on the quantitative assessment of aspect-oriented solutions [14]. Kersten and Murphy [15] described the effect of aspects on object-oriented development practices, as well as some rules and policies that were employed to achieve maintainability and modifiability. Walker et al. [16] provided initial insights into the usefulness and usability of aspect-oriented programming. Soares et al. [17] reported that the AspectJ implementation of the Web-based information system has significant advantages over the corresponding pure Java implementation. Garcia et al. [18] presented a quantitative study, designed to compare the maintenance and reuse support of a pattern-oriented approach, and an aspect-oriented approach for a multi-agent system. It turned out that the aspect-oriented approach allowed the construction of the investigated system with improved modularization of the crosscutting agent-specific concerns. The use of aspects resulted in superior separation of the agent-related concerns, lower coupling (although less cohesive) and fewer lines of code. Tsang et al. [19] evaluated the effectiveness of AOP for separation of concerns. They applied the CK metrics suite to assess and compare an aspect-oriented and object-oriented real-time system in terms of system properties. They found improved modularity of aspect-oriented system over object-oriented system, indicated by the reduction in coupling and lack of cohesion values of the CK metrics. Hannemann and Kiczales [20], as well as Garcia et al. [21], have developed systematic studies that investigated the use of aspect-oriented programming to implement classical design patterns. It is worth mentioning that Tonella and Ceccato [22] performed an empirical assessment of refactoring the aspectizable interfaces. This study indicates that migration of the aspectizable interfaces has a limited impact on the principal decomposition size, but, at the same time, it produces an improvement of the code modularity. From the point of view of the external quality attributes, modularization of the implementation of the crosscutting interfaces clearly simplifies the comprehension of the source code. Unfortunately, most empirical studies involving aspects have been based on subjective criteria and qualitative investigation [14].

III. ASPECT ORIENTATION PROGRAMMING(AOP)

Aspect Oriented Programming (AOP) is a novel software development paradigm that aims at modularizing *aspects*,

which are defined as well-modularized crosscutting concerns [23][24]. This type of concerns cuts across traditional module boundaries such as classes and interfaces, and their implementation is scattered and tangled with the implementation of other concerns. *AspectJ* is the popular Java extension language of AOP [23]. This basic constructs of the language are

Join point: A join point is a well-defined point in the execution of a component. It can be a method call or execution, an access to an attribute, or the execution of a constructor.

Pointcut: A pointcut is the mechanism that encapsulates join points. It can be composed of one or more join points.

Advice: An advice specifies the action (i.e., code) that must take place at a certain pointcut (i.e., a group of join points). With both abstractions mentioned above, advice gives developer the ability to implement crosscutting concerns.

There are three types of advice:

–before: The code declared is executed before the join point.

–after: The code declared is executed after the Join point.

–around: The code declared is executed instead of the one in the join point.

Inter-type declaration: This mechanism allows the developer to crosscut concerns in a static way. It permits alterations to classes and inheritance hierarchies from outside the original class definition. We enumerate below the types of possible changes through Inter-type declaration:

–Add members (methods, constructors, fields) to types (including other aspects).

–Add concrete implementation to interfaces.

–Declare that types extend new types or implement new interfaces.

–Declare aspect precedence.

–Declare custom compilation errors or warnings.

–Convert checked exceptions to unchecked.

Aspect: An aspect is the container for the encapsulation of pointcuts, advice code, and inter-type declaration. Acting like a Java classes, it can contain its own attributes and methods.

In AspectJ, an application consists of two parts: *base code* which corresponds to standard Java classes and interfaces, and *aspect code* which contains the crosscutting code. Next we describe the two types of crosscuts that AspectJ provides.

Static Crosscuts

Static crosscuts affect the static structure of a program [25,33]. We consider *Inter-Type Declarations (ITDs)*, also known as *introductions*, that add fields, methods, and constructors to existing classes and interfaces [25, 33].

Dynamic Crosscuts

Dynamic crosscuts run additional code when certain events occur during program execution. The semantics of dynamic crosscuts are commonly described and defined in terms of an event-based model [26][27]. As a program executes, different events fire. These events are called *join points*. Examples of join points are: variable reference, variable assignment, execution of a method body, method call, etc. A

pointcut is a predicate that selects a set of join points. *Advice* is code executed before, after, or around each join point matched by a *pointcut*[26].

IV. OBJECT ORIENTATION (OO) SOFTWARE METRICS

The inadequacy of the metrics in use with procedural code (size, complexity, etc.), when applied to OO systems, led to the investigation and definition of several metrics suites accounting for the specific features of OO software. However, among the available proposals, the one that is most commonly adopted and referenced is that by Chidamber and Kemerer [3]. Some notions used in the Chidamber and Kemerer's suite can be easily adapted to AOP software, by unifying classes and aspects, as well as methods and advices. Aspect introductions and static crosscutting require minor adaptations. However, novel kinds of coupling are introduced by AOP, demanding for specific measurements. For example, the possibility that a method execution is intercepted by an aspect *pointcut*, triggering the execution of an advice, makes the intercepted method coupled with the advice, in that its behavior is possibly altered by the advice. In the reverse direction, the aspect is affecting the module containing the intercepted operation, thus it depends on its internal properties (method names, control flow, etc.) in order to successfully redirect the operation's execution and produce the desired effects.

V. ASPECT ORIENTED(AO) SOFTWARE METRICS

In this section, the Chidamber and Kemerer's metrics suite is revised. Some of the metrics are adapted or extended, in order to make them applicable to the AOP software. Since the proposed metrics apply both to classes and aspects, in the following the term *module* will be used to indicate either of the two modularization units. Similarly, the term *operation* subsumes class methods and aspect advices/introductions.

- **WOM (Weighted Operations in Module):** *Number of operations in a given module.*
Similarly to the related OO metric, WOM captures the internal complexity of a module in terms of the number of implemented functions. A more refined version of this metric can be obtained by giving different weights to operations with different internal complexity.
- **DIT (Depth of Inheritance Tree):** *Length of the longest path from a given module to the class/aspect hierarchy root.*
Similarly to the related OO metric, DIT measures the scope of the properties. The deeper a class/aspect is in the hierarchy, the greater the number of operations it might inherit, thus making it more complex to understand and change. Since aspects can alter the inheritance relationship by means of static crosscutting, such effects of aspectization must be taken into account when computing this metric.
- **NOC (Number Of Children):** *Number of immediate sub-classes or sub-aspects of a given module.*
Similarly to DIT, NOC measures the scope of the properties, but in the reverse direction with respect to

DIT. The number of children of a module indicates the proportion of modules potentially dependent on properties inherited from the given one.

- **CAE (Coupling on Advice Execution):** *Number of aspects containing advices possibly triggered by the execution of operations in a given module.*
If the behavior of an operation can be altered by an aspect advice, due to a *pointcut* intercepting it, there is an (implicit) dependence of the operation from the advice. Thus, the given module is coupled with the aspect containing the advice and a change of the latter might impact the former. Such kind of coupling is absent in OO systems.
- **CIM (Coupling on Intercepted Modules):** *Number of modules or interfaces explicitly named in the pointcuts belonging to a given aspect.*
This metric is the dual of CAE, being focused on the aspect that intercepts the operations of another module. However, CIM takes into account only those modules and interfaces an aspect is aware of – those that are explicitly mentioned in the *pointcuts*. Submodules, modules implementing named interfaces or modules referenced through wild-cards are not counted in this metric, while they are in the metric CDA (see below), the rationale being that CIM (differently from CDA) captures the *direct* knowledge an aspect has of the rest of the system. High values of CIM indicate high coupling of the aspect with the given application and low generality/reusability.
- **CMC (Coupling on Method Call):** *Number of modules or interfaces declaring methods that are possibly called by a given module.*
This metric descends from the OO metric CBO (Coupling Between Objects), which was split into two (CMC and CFA) to distinguish coupling on operations from coupling on attributes. Aspect introductions must be taken into account when the possibly invoked methods are determined. Usage of a high number of methods from many different modules indicates that the function of the given module cannot be easily isolated from the others. High coupling is associated with a high dependence from the functions in other modules.
- **CFA (Coupling on Field Access):** *Number of modules or interfaces declaring fields that are accessed by a given module.*
Similarly to CMC, CFA measures the dependences of a given module on other modules, but in terms of accessed fields, instead of methods. In OO systems this metric is usually close to zero, but in AOP, aspects might access class fields to perform their function, so observing the new value in aspectized software may be important to assess the coupling of an aspect with other classes/aspects.
- **RFM (Response For a Module):** *Methods and advices potentially executed in response to a message received by a given module.*

Similarly to the related OO metric, RFM measures the potential communication between the given module and the other ones. The main adaptation necessary to apply it to AOP software is associated with the *implicit* responses that are triggered whenever a pointcut intercepts an operation of the given module.

- **LCO (Lack of Cohesion in Operations):** *Pairs of operations working on different class fields minus pairs of operations working on common fields (zero if negative).*

Similarly to the LCOM (Lack of Cohesion in Methods) OO metric, LCO is associated with the pairwise dissimilarity between different operations belonging to the same module. Operations working on separate subsets of the module fields are considered dissimilar and contribute to the increase of the metric's value. LCO will be low if all operations in a class or an aspect share a common data structure being manipulated or accessed.

- **CDA (Crosscutting Degree of an Aspect):** *Number of modules affected by the pointcuts and by the introductions in a given aspect.*

This is a brand new metric, specific to AOP software, that must be introduced as a completion of the CIM metric. While CIM considers only explicitly named modules, CDA measures all modules possibly affected by an aspect. This gives an idea of the overall impact an aspect has on the other modules. Moreover, the difference between CDA and CIM gives the number of modules that are affected by an aspect without being referenced explicitly by the aspect, which might indicate the degree of generality of an aspect, in terms of its independence from specific classes/aspects. High values of CDA and low values of CIM are usually desirable.

- **Weighted Methods per Class (WMC):** WMC is a measure of the number of methods implemented within a class. This metric measures understandability, maintainability, and reusability as follows:

- The number of methods in a class reflects the time and effort required to develop and maintain the class.
- The larger the number of methods, the greater the potential impact on children, since children inherit all of the methods defined in a class.

A class with a large number of methods is more application-specific, and therefore is not likely to be reused.

- **Lack of Cohesion in Methods (LCOM):** LCOM is the degree to which methods within a class are related to one another and work together to provide well-bounded behavior. Well-designed systems should designing some existing software systems to incorporate the aspect-oriented paradigm.

- **Weighted Methods per Class:** aspects might help reduce the number of methods per class as

maximize cohesion, since it promotes encapsulation. LCOM measures the degree of similarity of methods by data input variables or class attributes. In [28], two ways of measuring LCOM are described:

- Calculate for each data field in a class what percentage of the methods use that data field. Average the percentages then subtract from 100%. Lower percentages mean greater cohesion of data and methods in the class.
- Methods are more similar if they operate on the same attributes. Count the number of disjoint sets produced from the intersection of the sets of attributes used by the methods.

This metric evaluates efficiency and reusability. High cohesion indicates good class subdivision. Low cohesion increases complexity, thereby increasing the likelihood of errors during the development process. Classes with low cohesion could probably be subdivided into two or more subclasses with increased cohesion.

- **Coupling Between Objects (CBO):** CBO is a count of the number of other classes to which a class is coupled. CBO is measured by counting the number of distinct non-inheritance related class hierarchies on which a class depends. Excessive coupling prevents reuse. The more independent a class is, the more likely it can be reused. The higher the coupling the more sensitive the system is to changes in other parts of the design, and therefore maintenance is more difficult. High coupling also reduces the system's understandability because it makes the module harder to understand, change, or correct by itself if it is interrelated with other modules.
- **Response For a Class (RFC):** RFC is the number of all methods that can be invoked in response to a message to an object of the class or by some method in the class. This measures the amount of communication with other classes. The larger the number of methods that can be invoked from a class through messages, the greater the complexity of the class. If a large number of methods can be invoked in response to a message, the testing and debugging of the class becomes complicated as it requires a greater level of understanding on the part of the developer. This metric evaluates understandability, maintainability, and testability.

VI. THE EFFECT OF AO ON THE C&K SUITE

In this section it provides an analysis of the effect of aspect-orientation on the C&K metrics suite. It is based on the case studies found in the literature about re-

follows:

- Aspects combine crosscutting functionalities in modular, encapsulated units. Without aspect-oriented design, these crosscutting functionalities

would be tangled in the core class.

- In some cases, a sub- class might have to over- ride a function in its parent class in order to define its own aspectual behavior (not a core behavior). Let’s take exception handling as an example. A function in the subclass might have to override a super- class function just to implement the sub-class’s method of handling a certain exception. If exception handling was implemented as an aspect, the subclass will not have to add a function to implement its own exception handling technique. This reduces the WMC factor. Dealing with exception handling as aspects is discussed in more detail in [29].
- **Depth of Inheritance Tree:** subclasses that might be defined only for the purpose of applying their own implementation of aspectual behavior will not exist in systems designed using the AO Paradigm, because aspects will be responsible for that. This helps in reducing the depth of inheritance tree.
- **Number Of Children:** the same argument of “Depth of Inheritance Tree” is valid for this metric.
- **Lack of Cohesion in Methods:** aspects filter out crosscutting behavior, and therefore increases cohesion. Figure 1 is an example of this. The function Movable() is likely to contain synchronization checking that determine if the function Move can be invoked on an object of type

Shape. This can be seen as a synchronization aspect, which uses its own flags to determine synchronization. Such a crosscutting function reduces the cohesion of the class Shape.

- **Coupling Between Objects:** the presence of aspects is likely to decrease the coupling between core classes, yet increase the coupling between core classes and aspect classes. This is because aspects are new entities on which core classes depend. It should be noted, however, that, unlike aspects, core classes are more likely to be reused. Decreasing the coupling between core -classes is a beneficial issue, and increasing coupling between aspects and core classes in return can be seen as a good trade-off. Given that a design might involve coupling between c lasses, it would be better to have this coupling occur between core and aspect classes, rather than having it happen between core classes.
- **Response For a Class:** RFC is likely to increase in the presence of aspects. This is because the number of entities that a class communicates with increases, and classes have to communicate with aspects. The positive point with using aspects is that they can be designed in a way that encapsulates the logic and the objects with which a class communicates in a modular way.

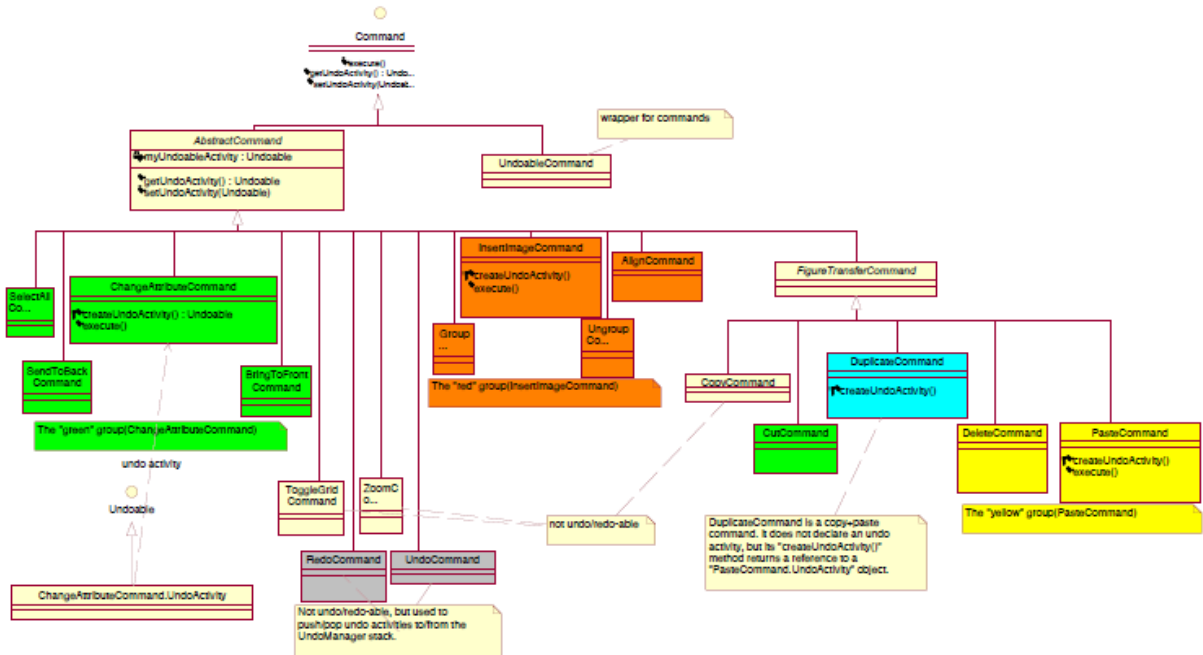


Figure 1:JHotDraw Command Hierarchy

VII. CASE STUDY : AJHOTDRAW

The case study selected is AJHotDraw [30], an AspectJ implementation of JHotDraw [31]. The original JHotDraw project was developed by Erich Gamma and Thomas Eggenschwiler. It is a Java GUI frame- work for technical

and structured graphics. It has been developed as a design exercise but it is quite powerful. Its design relies heavily on some well-known design pat- terns. The AJHotDraw program contains more than 400 elements (classes, interfaces and aspects). To our best knowledge, there is no application of that size that has been carefully studied in the past regarding aspect- oriented quality. The *Command*

hierarchy in JHOTDRAW, shown in figure 1, implements the design pattern bearing the same name. The (12) undoable commands store a reference to their associated undo activity. These command's execution through dedicated factory methods. AJHotDraw is an open source software project that provides numerous features for drawing and manipulating graphical and planar objects [1]. It consists of 13 features for a total of ~ 50KLOC. It is implemented with 279 classes and interfaces and only 31 aspects. Not surprisingly approximately 99% per-cent of the code is standard Java and only 1% of aspect code, of which almost all comes from ITDs. The modularized crosscutting concerns are persistence, design policies, contract enforcement, Undo command.

VIII. IMPACT OF AOP METRICS ON AJHOTDRAW

The proposed metrics have been computed on an open source project AJHotDraw, taken from the implementation of some design patterns [32] provided by Jan Hannemann both in Java and in AspectJ (appropriate AO Metrics are shown in Figure 2)

The practical implementation is based on *Observer* design pattern [32], in which there are two distinct roles, the *Subject* and the *Observer*. The Subject is an entity that can be in several different states. Some of the state changes are of interest to the Observer, which may take some actions in response to the change.

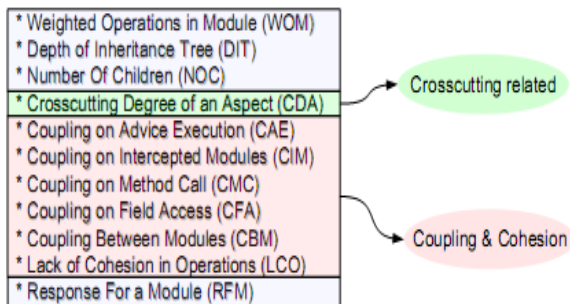


Figure 2: AOP C & K Metrics

The Observer pattern requires that the Observer registers itself on those Subjects it intends to observe. The Subject maintains a list of the Observers registered so far. When the Subject changes its state, it notifies the Observers of the change, so that the Observers can take the appropriate actions. In the OO implementation by Jan Hannemann, this design pattern consists of two interfaces, ChangeSubject and ChangeObserver, with the abstract definitions of the Subject and Observer roles. Moreover, the implementation contains the Point and the Screen classes, the first playing the role of Subject whereas the second plays both roles in two different instances of the pattern. The Main class contains the code to set up the two different pattern instances and run them. In the first pattern instance Point acts as the Subject and Screen as the Observer. In the second case, an instance of the class Screen is the Subject, while other instances of the same classes are its Observers. The AOP implementation contains a different version of the classes Point and Screen, with no

code regarding the Subject/Observer roles. *ObserverProtocol* is an abstract aspect defining the general structure of the aspects that implement the Observer pattern. This abstract aspect is extended by *ScreenObserver*, *ColorObserver* and *CoordinateObserver*. These concrete aspects contain the actual implementation of the protocol. By means of inter-type declarations, they impose roles onto the involved classes and by means of appropriate pointcuts they specify the Subject actions to be observed. Moreover, these aspects contain the mapping that connects a Subject to its Observers. The class Main runs the code for the initialization of the patterns for their execution. The output of the metric suite to the two implementations of the Observer pattern and the median values produced by the tool are shown in Table 1. The value of LCO for the OO code is indicated as 1-12, since these two values are adjacent to the median point. We observe that the improvement in some metrics (WOM, LCO, CMC and RFM), no change in other metrics (NOC and CFA) and a worse value of DIT (due to the superaspect *ObserverProtocol*). But the general values change only a little bit, for RFM the change is relatively high, passing from 7 to 2. LCO is also affected positively, going from 1-12 to 0. The cost to be paid for such improvements is an increase of the CIM metric, due to the aspects intercepting method executions (AOP coupling).

version	WOM	DIT	NOC	CAE	CIM
java	3	1	0	0	0
aspectj	1	2	0	0	2
version	CMC	CFA	RFM	LCO	CDA
java	2	0	7	1-12	0
aspectj	1	0	2	0	3

Table 1: AOP Metrics for AJHOTDRAW

IX. CONCLUSIONS AND FUTURE WORK

Assessing the quality of software has been the preoccupation of software engineers for two decades. The problem of separation of concerns led to the apparition of the aspect-oriented paradigm. This new paradigm raises questions about quality, due to its close relations with object-oriented programming. In this paper, we argue that the impact of AOP on software development quality metrics is significant. The proposed work shall be validated through empirical studies. In fact, case study used here shall enable us to appraise the quality of an aspect-oriented system over object oriented system.

Some of the issues that require more research and metrics are:

Aspect Granularity: how many crosscutting functionalities should an aspect encapsulate. Dependency between aspect

and class: how aspects can be designed such that the dependency of core classes on them is minimal.

Understandability: how aspects affect the system's understandability.

Depth of aspect inheritance tree: are there limitations for aspect-inheritance? And how far does it affect the design understandability.

X. REFERENCES

- 1) Awais Rashid et al "Aspect-Oriented Software Development in Practice: Tales from AOSD-Europe", Lancaster University, UK. Published by the IEEE Computer Society, February, 2010.
- 2) Kiczales, G., Lamping, J., Mendhekar, A., Maeda, C., Lopes, C. V., Loingtier, J.-M., and Irwin, J.: 'Aspect-Oriented Programming', Proc. European Conf. Object-Oriented Programming (ECOOP 1997), vol. 1241 of Lecture Notes in Computer Science, Jyvaskylä, Finland, June 1997, pp. 220–242
- 3) Chidamber, S. R., and Kemerer, C. F.: 'A Metrics Suite for Object Oriented Design', IEEE Trans. Softw. Eng., 1994, 20, (6), pp. 476–493
- 4) Martin, R. C.: 'OO Design Quality Metrics: An Analysis of Dependencies', <http://www.objectmentor.com/resources/articles/oodmetric.pdf>, accessed September 2006
- 5) George, B., and Williams, L. A.: 'An Initial Investigation of Test Driven Development in Industry', Proc. ACM Symposium on Applied Computing (SAC 2003), Melbourne, USA, March 2003, pp. 1135–1139
- 6) George, B., and Williams, L. A.: 'A structured experiment of test driven development', Inf. Softw. Tech., 2004, 46, (5), pp. 337–342
- 7) Madeyski, L.: 'Preliminary Analysis of the Effects of Pair Programming and Test-Driven Development on the External Code Quality', in Zieliński, K., and Szmuc, T. (Ed.): 'Software Engineering: Evolution and Emerging Technologies', vol. 130 of Frontiers in Artificial Intelligence and Applications, (IOS Press, 2005), pp. 113–123
- 8) Ceccato, M., and Tonella, P.: 'Measuring the Effects of Software Aspectization', Proc. Workshop on Aspect Reverse Engineering (WARE 2004), Delft, The Netherlands, November 2004 (Cd-rom)
- 9) Aopmetrics project, <http://www.e-informatyka.pl/sens/Wiki.jsp?page=Projects.AOPMetrics>, accessed September 2006
- 10) Subramanyam, R., and Krishnan, M. S.: 'Empirical Analysis of CK Metrics for Object-Oriented Design Complexity: Implications for Software Defects', IEEE Trans. Softw. Eng., 2003, 29, (4), pp. 297–310
- 11) Briand, L. C., Wüst, J., Ikonovskii, S. V., and Lounis, H.: 'Investigating Quality Factors in Object-Oriented Designs: an Industrial Case Study', Proc. Int. Conf. on Software Engineering (ICSE 1999), Los Alamitos, USA, May 1999, pp. 345–354
- 12) Emam, K. E., Melo, W. L., and Machado, J. C.: 'The Prediction of Faulty Classes Using Object-Oriented Design Metrics', J. Syst. Softw., 2001, 56, (1), pp. 63–75
- 13) Basili, V. R., Caldiera, G., and Rombach, H. D.: 'The Goal Question Metric Approach', in Marciniak J.J. (Ed.): Encyclopedia of Software Engineering, (Wiley, 1994), pp. 528–532
- 14) Garcia, A. F., Sant'Anna, C., Figueiredo, E., Kulesza, U., Lucena, C. J. P.de, and Staa, A.von: 'Modularizing Design Patterns with Aspects: A Quantitative Study', in Rashid, A., and Aksit, M. (Ed.): T. Aspect-Oriented Software Development I, 2006, vol. 3880 of Lecture Notes in Computer Science, pp. 36–74
- 15) Kersten, M., and Murphy, G. C.: 'Atlas: A Case Study in Building a Web-Based Learning Environment using Aspect-oriented Programming', Proc. ACM SIGPLAN Conf. on Object-Oriented Programming, Systems, Languages, and Applications (OOPSLA 1999), New York, USA, November 1999, pp. 340–352
- 16) Walker, R. J., Baniassad, E. L. A., and Murphy, G. C.: 'An Initial Assessment of Aspect-oriented Programming', Proc. Int. Conf. on Software Engineering (ICSE 1999), Los Alamitos, USA, May 1999, pp. 120–130
- 17) Soares, S., Laureano, E., and Borba, P.: 'Implementing Distribution and Persistence Aspects with AspectJ', Proc. ACM SIGPLAN Conf. on Object-Oriented Programming, Systems, Languages, and Applications (OOPSLA 2002), New York, USA, November 2002, pp. 174–190
- 18) Garcia, A. F., Sant'Anna, C., Chavez, C., Silva, V. T.da, Lucena, C. J. P.de, and Staa, A.von: 'Separation of Concerns in Multi-agent Systems: An Empirical Study', Proc. Int. Workshop on Software Engineering for Large-Scale Multi-Agent Systems (SELMAS 2003), vol. 2940 of Lecture Notes in Computer Science, pp. 49–72
- 19) Tsang, S. L., Clarke, S., and Baniassad, E. L. A.: 'An Evaluation of Aspect-Oriented Programming for Java-Based Real-Time Systems Development', Proc. Int. Symposium on Object-Oriented Real-Time Distributed Computing (ISORC 2004), Vienna, Austria, May 2004, pp. 291–300
- 20) Hannemann, J., and Kiczales, G.: 'Design Pattern Implementation in Java and AspectJ', Proc. ACM SIGPLAN Conf. on Object-Oriented Programming, Systems, Languages, and Applications (OOPSLA 2002), New York, USA, November 2002, pp. 161–173
- 21) Garcia, F., Bertoa, M. F., Calero, C., Vallecillo, A., Ruiz-Sánchez, F., Piattini, M., and Genero, M.: 'Towards a consistent terminology for software

- measurement', *Inf. Softw. Tech.*, 2006, 48, (8), pp. 631–644
- 22) Tonella, P., and Ceccato, M.: 'Refactoring the aspectizable interfaces: An empirical assessment.', *IEEE Trans. Softw. Eng.*, 2005, 31, (10), pp. 819–832
 - 23) AspectJ, <http://eclipse.org/aspectj/>
 - 24) Kiczales, G., Hilsdale, E., Hugunin, J., Kirsten, M., Palm, J., Griswold, W.G.: 'An overview of AspectJ'. ECOOP (2001)
 - 25) Laddad, R.: 'AspectJ in Action. Practical Aspect-Oriented Programming'. Manning (2003)
 - 26) Lämmel, R.: 'Declarative Aspect-Oriented Programming'. PEPM (1999),
 - 27) Wand, M., Kiczales, G., Dutchyn, C.: 'A Semantics for Advice and Dynamic Join Points in Aspect Oriented Programming'. TOPLAS (2004)
 - 28) Rosenberg, Linda H. and Lawrence E. Hyatt. "Software Quality Metrics for Object-Oriented Environments". NASA, SATC. http://www.satc.gsfc.nasa.gov/support/CROSS_AP_R97/oocross.html
 - 29) Lippert, Martin, Cristina Videira Lopes. 'A Study on Exception Detection and Handling Using Aspect-Oriented Programming'. Xerox Palo Alto Research Center. Technical Report, Dec. 99.
 - 30) Ajhotdraw project. <http://sourceforge.net/projects/ajhotdraw/>.
 - 31) Jhotdraw project. <http://www.jhotdraw.org/>.
 - 32) E. Gamma, R. Helm, R. Johnson, and J. Vlissides. 'Design Patterns: Elements of Reusable Object Oriented Software'. Addison-Wesley Publishing Company, Reading, MA, 1995.
 - 33) LADDAD, R. 'Enterprise AOP with Spring Applications: AspectJ in Action' 2nd Edition, Manning Publications, 2010.

HE-Advisor: A Multidisciplinary Web-Based Higher Education Advisory System

Fawzi Albaloooshi¹ Safwan Shatnawi²

{ GJCST Computing Classification }
J 1, L 2.1

Abstract-The paper addresses academic advising-an important issue that is not often given enough attention by students and advisors alike. A web-based multidisciplinary advising system is presented that can be utilized by students, advisors, course timetable planners, and heads of departments. Students are given informative advice through web-based services to help them make best decisions towards a successful degree of their choice. Services, such as registering for courses to stay on the right degree path; a dependency graph showing their progress in their degree plan; a GPA simulator to help students on probation determine the grades they must obtain in the newly registered semester; information about their graduation requirements; their expected graduation semester; and other services. Advisors and heads of departments are able to see students' progress towards their graduation and are able to generate a variety of useful statistics, charts, and reports. Timetable planners are given statistics on courses and their sections' requirements for the coming semester.

Keywords-Academic Advising, Automated Advising, Web-Based Advising, Proactive Advising.

I. INTRODUCTION

Student advising is an important issue that is not regularly given enough attention by instructors and students alike. Many students do not take the time and effort to see their advisors to plan their timetable before registration, resulting in many registration issues and long queues for advising at registration time. Advisors have difficulties obtaining accurate and detailed academic information on students to assess their situation. Academic departments face difficulties with timetabling to plan ahead for the courses to be offered for the coming semester, and to determine the number of sections for each course.

In this paper we present a web-based advising system that accesses academic information on students, such as degree information, transcript records, and existing registration plans. It also offers students, advisors, and heads of departments an easily accessed set of services that will enable them to be better informed and therefore act effectively. The system is multidisciplinary in that it can be used by different departments and can host different degree plans that are introduced to the system through a special web page with administrative privilege. Students are then linked to their registered degree plans and academic departments. It is web-based, enabling students and advisors easy access anywhere, anytime, and thus overcoming place and time barriers, which are the main limitations of

traditional advising. Students have access to their existing timetable; transcript; suggested courses to be registered in the coming semester; can track their graduation progress using either tabular listing, or in a form of a dependency graph; simulate their new GPA by inputting expected grades for newly registered courses; and update their personal information. Advisors have controlled access to their advisees' accounts, thus enjoying all the services available to students; are able to retrieve a list of students expected to graduate; a list of graduated students; and update their profiles. Heads of departments have controlled access to all department advisors' accounts, thus enjoying all services provided to advisors and their students; can view historical records and charts of courses' grades; can view students' performance statistics and charts; timetable statistics of courses and their sections that need to be offered in the coming semester; and lists of students expected to graduate and graduated students, and their records.

Section II presents a literature survey of published research on automated student advising. The students' academic information available in the existing registration system is extracted, processed, and stored in a new suitable format as explained in section III. In section IV the HE-Advisor is presented in detail with sample screen shots and reports. The system is evaluated from two different users' perspectives: students and advisors. The details and analysis of this evaluation are presented in section V. In section VI we compare our system with a number of systems reported in the published literature as outlined in section II, in terms of the services they offer to the different stakeholders. Our concluding remarks and envisaged future work are presented in section VII.

II. LITERATURE SURVEY

Although the system described in this paper was primarily motivated by our need to optimize advising, to determine which courses should be offered, and to utilize existing information available in the university registration system, there is no doubt that academics worldwide agree that proper advising is an important factor for students' successful progress in higher education. Many studies have been conducted to confirm this matter as reported by Bailes et al. (2002) and Siegfried (2003). As a result, many academic institutes have investigated the use of computer technologies in academic advising to overcome the difficulties experienced with traditional methods. A sample of such studies is presented in the remainder of this section.

Bailes et al. (2002) proposed systemized academic advising is made of key subsystems grouped under basic study planning and high-level planning that can be automated to

About-^{1,2}College of Applied Studies University of Bahrain Sakhir Campus, Kingdom of Bahrain
(e-mail-¹fawzi@itc.uob.bh, 2sshatnawi@asc.uob.bh)

the benefit of advisers and advisees. Under basic study planning, the authors suggest that a number of basic queries be available, such as course availability, prerequisites, degree completion requirements, degree credit transfer possibilities, and time constraints. Under high-level planning it is suggested to offer services such as course availability, course of student interest, popular plans, and degree change implications. The proposed system was later on implemented as an honors degree thesis by Ganatra (2002) under Bailes supervision. Marques et al. (2001) report a system that offers advisees up-to-date online advising and related information including a recommended list of courses in which a student must register in the next semester in order to complete his degree requirements. The system has a web-based main page through which system users such as students, faculty, and administrative staff are allowed access to their respective sites. Siegfried et al. (2003) present the motivations to develop FROSH (Siegfried 1993), the automated advisor for freshmen. O'Mahony and Smyth (2007) present a collaborative course recommender system that recommends elective modules to students, based on the core modules they have selected. A "More Like This" recommender offers students similar modules to their first choice of electives in cases where no places are available or where timetable clashes occur. Pokraja and Rasamny (2006) present InVEStA, an expert system for advising students on the courses in which they must register next. The recommender system also generates a semester schedule based on the courses offered for the semester and the student curricula. Naini et al. (2008) present a web-based interactive student advising system for course plan selection and registration using Java framework. Patanker (1998) presents the use of an expert system shell called VP-Expert to develop an advising system called Academic Counseling Expert (ACE) with three major objectives: present the student with suggested courses to register for based on his major and completed courses; present the student with equivalent courses from other universities; and prepare a suitable student timetable avoiding time conflicts. Grupe (2002) presents a web-based expert system that assesses a student's capabilities and advises him on the best major he should consider. Zucker (2009) introduces ViCurriAS a visual tool for advising that is composed of two main modules. The first is used to easily register new curriculum plans (such as course details, their semesterwise arrangements, and interrelationships) and the second is used to track the progress of enrolled students. Bansal et al. (2003) describe KRAK, a web-based advising system primarily developed to help students tailor design their college study path, while advisors play a major role as mentors. It allows the students to plan their entire degree, do semester scheduling, and provides course, faculty, and university information.

The works cited in this section are a selection of many systems developed by academic institutes for their specific programs that utilize the power of computer technologies to make student advising easier, accurate, and available to all.

III. REGISTRATION SYSTEM AND DATA PREPROCESSING

Our university registration system keeps track of students' personal information, transcripts, enrollment, and other course related details. All of this information is stored in the registration database and can be accessed as HTML pages by department managers, advisors, and students. As far as we know, not only does such information lack future indications, but it is also static and cannot be used either for educational data mining or for database processing.

A. Student information preprocessing.

In order to utilize the available student and course information for further processing and educational benefits, we convert the HTML pages to database records. This process is achieved by using customized web components and web semantic techniques elaborated in a supporting module created for the system. The goal behind this process is to create the students database.

The supporting module scans the HTML tags inside students' transcripts to identify the required data. After that, these data are organized into a records data structure and then inserted into the database tables. By scanning all students' transcripts and applying the semantic processing to these transcripts we come up with the students' database that includes information about courses registered and grades obtained, students' major, and students' status. The preprocessing step can be omitted in case of direct database access to the students' registration information.

B. Additional information

Information about students' degree plans, prerequisites, staff information, and department information are input to the students' database. Such information is essential for the system to be able to provide useful services for its potential users such as students, advisors, and educational managers.

IV. HE-ADVISOR SERVICES

The system reported in this paper offers standard, advanced, and configuration management services that can easily be accessed by students, advisors, educational managers, and alumni. Each user has a separate menu of services that is particular to his interest. For example, figure 1 presents a snapshot of a typical page accessible by students, showing the menu of services on the left. Figure 3 presents a typical page accessible by advisors showing the menu of services. Similarly, heads of departments have access to a special page of services, as shown in figure 8. The system can easily be configured to support new degree plans and update existing ones. The group of offered services benefits all stake holders i.e. students, advisors, and educational managers, therefore helping to improve the learning process at various levels and from different perspectives. More details of services offered are described in the subsections that follow and are categorized as standard, advanced, and configuration management services.

A. Standard Services

These services are based on retrieving and organizing students' information in direct queries that require retrieval, formatting, and organizing of students' data. Basically we have five services in this category:

1) Student transcript and registration information

The system stores information about students' current semester registration and transcripts. Students are able to obtain their current enrollment information and transcript after logging into their respective accounts. A sample student transcript is shown in figure 1.

2) Courses to be registered next semester

Courses that can be registered by the student in the following semester can be decided by the system. These courses are obtained by comparing the student's degree plan against the courses that the student has successfully passed and generating a list of all uncompleted courses. For each course in this list, the student's transcript is scanned to check whether the student successfully passed the prerequisite for that course, if any, or not. Then for each course the student successfully passed, its prerequisite is marked as able to be registered next semester and is displayed to the student. The courses are prioritized to ensure appropriate course registration according to the degree plan. If this service is accessed while the student is studying for a list of courses, the system assumes that the student passes these courses and generates a list of new courses based on the already completed and currently registered courses. Figure 2 shows a sample list of courses that can be registered in the coming semester for a student. Similar information can be retrieved by advisors, but for a range of students (his advisees) as shown in figure 3.

3) Student's graduation progress.

The student degree plan is displayed showing course status whether successfully passed or uncompleted. Courses in the degree plan are displayed in chronological order starting from first year toward graduation courses, providing the student with a comprehensive view of his/her study progress. This service offers the student five categories of courses. Each is labeled with a different icon as indicated in the key beside the table so that they can easily be recognized. The categories are: system suggested next semester courses; courses that cannot be registered; completed courses; completed courses that can be repeated (this is a special university rule for courses with a grade of C- and below); and currently registered courses. Figure 4 shows a snapshot for the graduation progress table for the same student shown in figures 1 and 2.

4) Student Progress Dependency Graph

This graph presents the student with the same information to the graduation progress table explained above and shown in figure 4, and using the same color coding, except it shows it

in the form of a graph. An example of such a graph is shown in figure 5 for the same student shown in figure 4.

5) Alumni services

Alumni are able to update their profiles, and stay in touch with their academic departments, enabling the academic departments to follow up their alumni and get feedback regarding degree plans, and how they can enhance the quality of curriculums and pursue the working market requirements.

B. Advanced Services

In this category not only does the system retrieve students' information but also it allows students to interact with the information such as shown in 1 below. Educational managers such as heads of departments and deans of colleges can retrieve useful information as explained in 2 and 3 below.

1) GPA simulation

A student can assign grades to courses that he/she is currently studying to simulate his/her expected GPA in advance. Our university regulations allow students to repeat courses with a grade less than or equal to C-. The system allows such students to check how their GPA would be affected when repeating one or more courses and by setting an expected grade for each. Figure 6 shows an example for a GPA simulation operation for the same student shown in figure 4, but expected grades for the currently registered courses are input to find out the change in the GPA.

2) Students' Statistical Information

In addition the system introduces extra services such as a student's rank among his/her colleagues registered for the same degree plan. This is achieved by calculating the number of students whose GPA is greater than the student's GPA. The student can also be ranked among all students in the same faculty. Figure 7 shows a sample snapshot of a student's ranking among others in the same program and college.

3) Statistical Information for Managers

For educational managers such as heads of departments and deans of colleges the system provides statistics and reports showing students' performance such as, students distribution over GPA ranges represented as numbers in a table, and as a bar chart as shown in the sample snapshot in figure 8. The system also provides educational managers statistics of students' performance in a given course over a given period of time as shown in figure 9 for the course ECONA131. Statistics for courses to be offered next semester can also be obtained by grouping information available in students' next semester registration tables and can be of great help to set the coming semester timetable. A sample snapshot is presented in figure 10 showing the courses that the department needs to offer and the expected number of students to register for each. Other useful reports include the

list of students expected to graduate by the end of semester and the list of graduated students.

C. Configuration Management Services.

These services enable the system manager to configure the business rules and system parameters, or, on the other hand, to update, delete, or insert new records to the database, i.e. if a new degree program is introduced, the system administrator can easily add it and load the degree plan and other required information to the database. Also, in case of updating an existing degree plan, the privileged users can update the database records. This part of the system is also responsible for managing users' accounts.

V. HE-ADVISOR EVALUATION

To evaluate the system we prepared a modified version of the original questionnaire proposed by Cafferella (1987). Our evaluators were a group of 54 students belonging to four different degree programs and a group of 16 advisors from different degree programs. The aim was to allow the two most concerned parties that benefit from the system services to first use the system and then answer the questionnaire shown in appendix A. The evaluators were shortly briefed about the system and then were allowed to logon and use the system. Each student had access to a student case enrolled in the same degree he is enrolled in. The advisors had access to the records of their advisees. All rated questions start with 1 as the lowest (negative) rating and end with 5 as the highest (positive) rating. Next we summarize the questionnaire results, broken down to its eight main sections as also shown in table 1, but a more detailed summary is available in Appendix B

- A. Program Content: this section is made-up of two questions judging the system content. The average rating for this part was 4.63 out of 5 by the students and 4.53 out of 5 by the advisors.
- B. Audience to Program: this part is made-up of five questions to measure the system's suitability for the targeted audience. 94% of the students respondents agreed that the system would benefit students and only 37% thought that advisors could benefit from the system, and only 20% thought that parents could benefit from the system services. As for advisors, 100% thought that students and advisors would benefit from the system and 63% thought that parents could also benefit from it. The average response for the three rated questions in this part was 4.37 by the students and 4.28 by the advisors. Instructional Strategies: this part measures the appropriateness of the system for advising. The students' average rating for all questions was 4.2 and advisors' rating 4.4.
- C. Program Design: this section included questions about system feedback, screen displays, ease of use, and user

friendliness. The students gave an average rating of 4.46 and the advisors rated it with an average of 4.59.

- D. Appropriate Use of Computers: this part measured the suitability of computers and the internet for advising. The students and advisors gave an average rating of 4.23 and 4.50 respectively. In response to the third open question related the suitability of other mediums for advising, few responses were received suggesting that direct consultation between student and advisor would be an alternative.
- E. Programming Techniques: this part included rated questions on the system's performance and operation. The average rating for the students' and advisors' were 4.23 and 4.58 respectively.
- F. Cost/Benefits Analysis: in this part question were asked to obtain feedback related to the system benefits. On average the students and advisors gave a rating of 4.39 and 4.67 respectively. We also obtained feedback of the expected time spent using the system for an advising session. When reading the responses to the expected usage times for the system we eliminated the unreasonable ones. Times such as below 4 and above 35 minutes for average usage; below 2 and above 30 minutes for minimum; and below 5 and above 50 minutes for maximum usage we found such values odd and did not consider them. The average for students' responses for the average time was 12 minutes and the advisors' was 12. The average for the students' responses for the minimum time was 8.5 and the advisors' was 10. The average for the students' for the maximum time was 18.6 and the advisors' was 15.
- G. Overall Evaluation: an overall evaluation was obtained from the two main users. The students rate it 4.5 out of 5 and the advisors 4.77 out of 5. With regards to the question related to the system strengths many messages of merit were received and can be summarized by stating that the system provides the user with valuable, accurate information; can solve many advising issues; is easy to follow, understandable, efficient, and time saving. Few comments on weakness were received and are mostly attributed to interface design issues such as the use of colors and fonts.

As developers we were happy with the evaluation results as students' and advisors' feedback was satisfactorily positive. All average ratings were above 83% reaching up to 95% in the average overall rating given by the advisors. We also received constructive feedback especially from advisors with regard to improvements to the interface design. They were looking forward to using the system as soon as possible for their regular student advising sessions, and were happy to learn that it is available online for further feedback.

Table 1. HE-Advisor Evaluation Brief Summary

	Students	Advisors
Evaluation Part	Avg. Rating / 5	Avg. Rating / 5
Program Content	4.63	4.53
Audience for CBI Program	4.37	4.28
Instructional Strategies	4.19	4.41
Program Design	4.46	4.59
Appropriate Use of Computers	4.23	4.50
Program Techniques	4.23	4.58
Cost/Benefit Analysis	4.39	4.67
Overall Evaluation	4.50	4.77

VI. COMPARISON WITH OTHER SYSTEMS

To benchmark our system we compared it with a number of similar systems in terms of functionality and services they provide to different users, as shown in table 2. At first we tried to classify the characteristics for each system, such as ease of use, accessibility, and configurability; whether the system is multidisciplinary or not in that it can be used for multiple disciplines, or specially designed for a specific degree program; and whether the system can show some intelligence towards its users. We then compared the systems towards the services they provide to the students, such as advising and degree planning; other related services offered, if any; and alumni records. Advisors are key players in the advising process so we looked at the services, reports and statistics the systems offers them, to improve the whole process. Important decision makers at a higher level are heads of departments and deans, so we looked at the reports and statistics the system offers them, to gain insight into the learning process. The first column in table 2 shows the systems and their references in brackets. The remaining columns show the system characteristics, benefits to the student, advisor, and management respectively. The values in those four columns are explained in the key below the table. As it is apparent from the table, the system presented in this paper offers more functionality and services than any other system reported in the surveyed literature.

Table 2. Comparison Between a Number of Advising Systems

System	Characteristics	Student	Advisor	Management
Frosh (12)	1, 2	1	1	
WISRAS (7)	1, 2	1	1	
InVEStA (10)	1, 2	1		
ACE (9)	1, 2	1	1	
ViCurriAS (13)	1, 2	1	1	
KRAK (2)	1, 3	1, 2		
HE-Advisor	1, 2, 3	1, 2, 3	1, 2, 3	1

Key:

Characteristics:

1. Ease of use and access, Configurable,
2. Multidisciplinary,
3. Intelligent Services.

Student: 1. Advising and Planning,

2. Services,
3. Alumni.

Advisor: 1. Access,

2. Services,
3. Reports & Statistics

Management: 1. Reports & Statistics

VII. CONCLUSIONS AND FUTURE WORK

The paper presents a multidisciplinary, web-based, higher education advisory system that offers students, advisors, and heads of departments anytime/anywhere easily accessed academic information in a user friendly form. Students have access to their timetables, transcripts, graduation progress, future courses to register for, and other services. Thus they are given all the information they need at their fingertips to properly plan their academic careers. Advisors have access to their advisees' records and graduation reports that can help them actively advise and assist their students to plan their future semester. Heads of departments have useful access to all advisors' and students' records; course statistics; students' statistics; and timetable statistics to assist them in better planning, decision making, and continuous improvement in the provision of learning.

We hope in the future to be able to develop a timetabling system that shares information with the advising system reported here. The new system would be able to prepare the new semester's timetable, keeping in mind the hard and soft constraints associated with timetabling. The new system will also be responsible for suitable course load distribution, based on the available instructors, their teaching history, course preferences, and timing preferences. Such sharing of students and timetable information between the two systems may trigger the possibility of developing an interactive registration system in which students and instructors would have a positive and active degree of contribution to timetable planning.

VIII. ACKNOWLEDGMENT


Thanks goes to deanship scientific research at the University of Bahrain for supporting research published in this paper

IX. REFERENCES

- 1) (2002) The Academic Advisor's Apprentice – An Automated Academic Advising System. CSEE Strategic Development Proposal, Department of Information Technology and Electrical Engineering, University of Queensland.
- 2) Bansal A., Cline K., Rangel K., and Zunke R. (2003). KRAK: knowledge-based resource advising kit. *Journal of Computing Sciences in Colleges*, vol. 18, issue 3, pp. 245-253, February 2003.
- 3) Caffarella, E. P. (1987). Evaluating the New Generation of Computer-Based Instructional Software. *Educational Technology*, 27(4), pp.19-24, 1987.
- 4) Grupe Fritz H. (2002). Student Advisement: Applying a Web-based expert system to the selection of an Academic Major. *College Student Journal*, December 2002.
- 5) Ganatra C. (2002). Academic Advisor's Apprentice. Honors Degree Thesis, October 2002. Department of Information Technology and Electrical Engineering. University of Queensland, Australia.
- 6) Marques O., Ding X., and Hsu S. (2001). Design and Development of A Web-based Academic Advising System. 31st ASEE/IEEE Frontiers in Education Conference, October 10-13, 2001 Reno, NV.
- 7) Naini V. R., Sadasivam R. S., and Tanik M. M. (2008). A Web-based Interactive Student Advising System using Java Frameworks. *Proceedings of the SoutheasternCon*, pp. 172-177, IEEE 2008.
- 8) O'Mahony M. P., and Smyth B. (2007). A Recommender System for On-line Course Enrolment: An Initial Study. *RecSys'07*, Minneapolis, Minnesota, USA, October 19-20, 2007, published by ACM.
- 9) Patankar Manoj(1998). ARule-Based Expert System Approach to Academic Advising. *Innovations in Education and Training International*, v35 n1 p49-58 Feb 1998.
- 10) Pokraja D., and Rasamny M. (2006). Inetractive Virtual Expert System for Advising (InVESTA). 36th ASEE/IEEE Frontiers in Education Conference, San Diego, CA, October 28-31, 2006.
- 11) Siegfried R. M.(1993). FROSH: An Expert System for Freshman Advisement. *Journal of Computing in Small Colleges*, vol. 9, no. 2, pp. 88-98, November 1993.
- 12) Siegfried R. M., Wittenstein A. M., and Sharma T. (2003). An Automated Advising System for

Course Selection and Scheduling. *Journal of Computing Sciences in Colleges*, vol. 18, no. 3, pp. 17-25, February 2003.

- 13) Zucker Ron(2009). ViCurriAs: a curriculum visualization tool for faculty, advisors, and students. *Journal of Computing Sciences in Colleges*, vol. 25, issue 2, pp. 138-145, December 2009.



University of Bahrain Higher Education Advisory System

Sunday, 30, May 2010 05:11
Welcome PROF. I

Name: AQEELA SAYED AHMED ALI AHMED(20071152)
Associate Diploma in Information Systems
GPA: 1.49 Cr. Passed: 42 **First Probation**

[Advisor Information](#) [Logout](#)

- [Transcript](#)
- [Timetable](#)
- [Next semester Courses](#)
- [Graduation Progress](#)
- [Profile](#)
- [Student's Rank](#)
- [Study Plan](#)
- [Registration system](#)

Course	Title	Grade	Year	Semester	Repeated
CSA111	INTRODUCTION TO IT AND COMPUTERS	C	2007	1	0
ENGLA111	ENGLISH LANGUAGE DEVELOPMENT I	D	2007	1	0
ENGLA112	ENGLISH COMMUNICATIONS SKILLS	C	2007	1	0
MATHA111	APPLIED MATHEMATICS	D	2007	1	0
CSA112	INTRODUCTION TO COMPUTER PROGRAMMING	F	2007	1	0

1 2 3 4


Fig. 1 A Student Transcript Offered by the System

- [Transcript](#)
- [Timetable](#)
- [Next semester Courses](#)
- [Graduation Progress](#)

You are advised to register the following courses

Priority	Course	Title	Lec	Lab	Credit
1	MISA290	Technical Project	0		3
2	MISA260	Applied Database Management Systems	2	3	3
3	MISA244	System Analysis & Design	2	3	3
4	STATA231	Applied Statistics	2	3	3

Fig. 2 Courses to be Registered Next Semester for a Student



University of Bahrain Higher Education Advisory System

Sunday, 30, May 2010 12:13

[Logout](#)

- [Students List](#)
- [Registration](#)
- [Expected Graduates](#)
- [Graduates](#)
- [Profile](#)
- [Main Menu](#)

Courses Can Be Register in Next Semester

S#	Name	courses									
20055326	QADSEYA SAYED HASHEM SHARAF HASHEM	--	--	--	--	--	--	--	--	--	--
20052232	NARJES ABDULREDHA HABIB JASIM	CSA299	--	--	--	--	--	--	--	--	--
20060035	NOOR MANSOOR MOHAMED RADHI	CEA112	CSA113	CSA114	CSA115	ENGLA210	--	--	--	--	--
20072853	NAJMA MOHAMED EBRAHIM ALSAKRAN	CSA299	--	--	--	--	--	--	--	--	--
20073441	MOHAMED ABDULRAB MOHAMED ALSAADI	CSA112	GSA111	ENGLA111	ENGLA112	CEA112	--	--	--	--	--
20073767	SALMAN HASAN SAIF ALMUTAWA	ARABA111	GSA111	ENGLA120	ACCA121	MGTA121	--	--	--	--	--
20082586	LAYLA AHMED MAHDI HASAN	CSA212	MATHA111	ENGLA210	CSA290	CSA223	--	--	--	--	--
20083174	LAYLA ALI ABDULLA SULAIMAN	MGTA121	ENGLA210	CSA214	CSA217	CSA218	--	--	--	--	--
20092023	MOHAMED A. RASOOL. ABDALI SHAMLAN	ENGLA120	CSA212	HIST121	ENGLA112	CSA211	--	--	--	--	--
20092188	ALI ISA ALI AL MARHOON	CEA112	ENGLA112	ENGLA120	HIST121	CSA211	--	--	--	--	--
20092262	MOHAMED ALI MOHAMED ALZAYED	CEA112	CSA115	ENGLA120	CSA114	CSA113	--	--	--	--	--
20092286	AHMED ALI YUSUF ALHAYKI	CSA114	MATHA111	CEA112	ENGLA112	CSA113	--	--	--	--	--
20092291	MAHDI HABIB HASAN ALI	CSA211	ENGLA210	ENGLA112	CSA212	CSA115	--	--	--	--	--
20092330	FATEMA ABDULLA MARHOON ALI	CSA212	ENGLA112	HIST121	ENGLA210	CSA211	--	--	--	--	--
20092334	WALAA MAKI AHMED QAMBER	HIST121	ENGLA210	CSA211	CSA212	ENGLA112	--	--	--	--	--
20092422	MOHAMED HASAN MOHAMED ALMALOOD	ENGLA112	CSA114	CSA115	ENGLA210	CSA211	--	--	--	--	--
20092493	FATEMA ABDULLA ALI ALHAMMADI	HIST121	ENGLA112	CSA211	CSA115	ENGLA210	--	--	--	--	--
20092631	WADEEA MAHDI ALI ALESKAFI	ENGLA112	CSA113	MATHA111	ENGLA111	CEA112	--	--	--	--	--
20092758	FATEMA MOHAMED JASIM FARHAN	CSA211	CSA212	ENGLA112	ENGLA210	HIST121	--	--	--	--	--

Fig. 3 Courses to be Registered Next Semester for a Group of Advisees

Transcript	key	Course	Title	Grade
Timetable		MATHA111	Applied Mathematics	D
Next semester Courses		ENGLA112	English Communications Skills	C
Graduation Progress		CSA111	Introduction to IT & Computers	C
Profile		CSA112	Introduction to Computer Programming	D
Student's Rank		GSA111	Work Ethics & Values	B
Study Plan		ENGLA111	English Language Development I	D
Registration system		ECONA131	Microeconomics	C-
Main Menu		MGTA121	Introduction to Business Administration	C+
		ENGLA120	English Language Development II	D
		MISA121	Management Information Systems	C-
		MISA123	Information Systems Architecture	D+
		ACCA121	Basic Accounting I	C-
		STATA231	Applied Statistics	F
		MISA233	Applied Business Data Communications	--
		MISA240	Info. Systems Applications in Business	D
		MKTA221	Marketing Management	C-
		MISA210	Electronic Business	--
		ENGLA210	Technical Report Writing	--
		MISA244	System Analysis & Design	--
		MISA260	Applied Database Management Systems	--
		MGTA264	Organization Management	--
		MISA290	Technical Project	--
		MISA299	Professional internship	--
	GPA Simulation	Reset	1.49	

Icon's key

- Can be registered next semester
- Can not be registered
- Passed but can not be repeated
- Can be repeated
- Currently registered

- Courses passed with grade >= C can not be repeated
 -Course marked as "--" means you didn't study these courses

To transfer to bachelor's your GPA must be minimum :
 A- 3 if you completed the first semester.
 B- 2.50 if you completed the second semester.
 C- 2.33 if you completed the graduation requirement.

Graduation Requirements
 1. Successfully passing all courses in your study plan
 2. Your GPA must be 1.5 or more.

Fig. 4 A Student's Study Progress Table

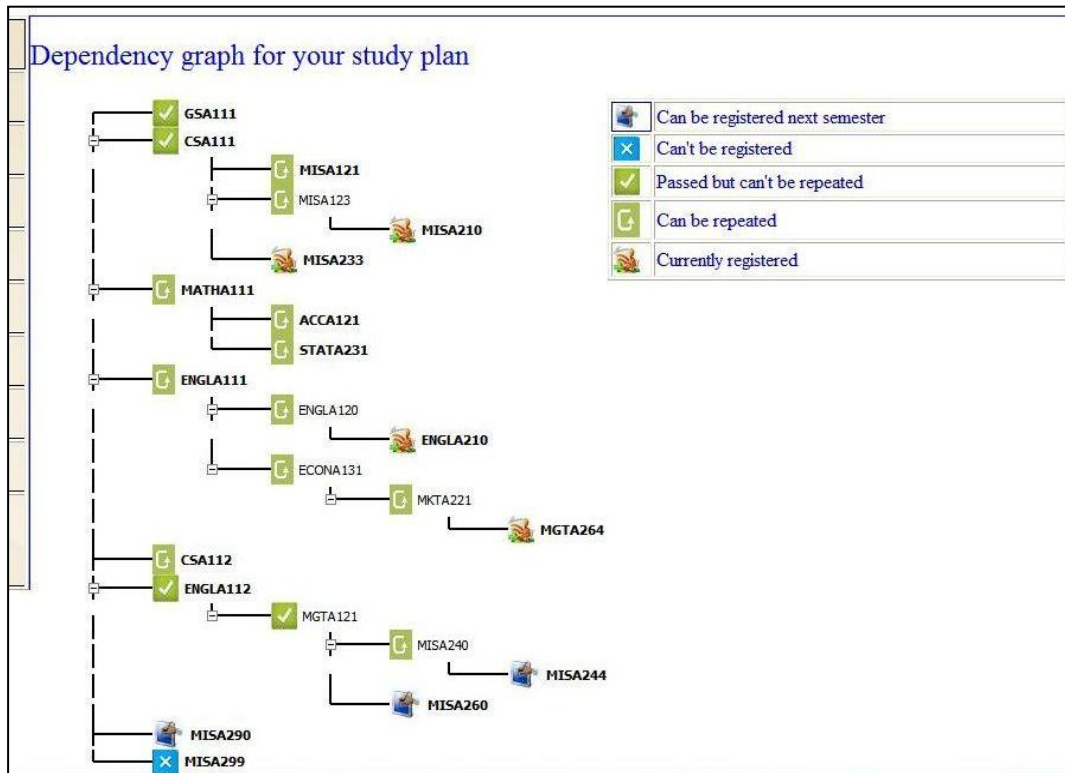


Fig. 5 Dependency Graph for a Student's Study Progress

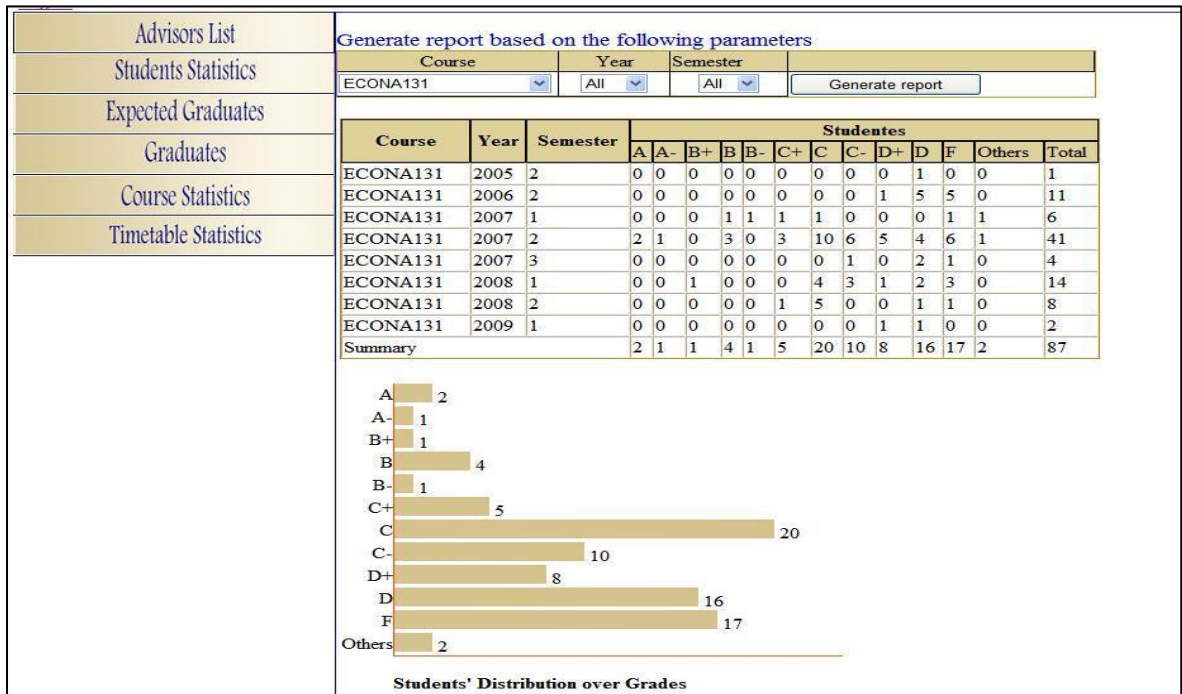


Fig. 9 Five Year Grades Statistics for ECONA131

Advisors List	Time table statistics for next semester		
Students Statistics	Major	Course	Student
Expected Graduates	Associate Diploma in Commercial Studies	All	
Graduates	Sort	Sort	Sort
Course Statistics	Associate Diploma in Commercial Studies	ACCA121	53
Timetable Statistics	Associate Diploma in Commercial Studies	ACCA221	53
	Associate Diploma in Commercial Studies	CSA111	24
	Associate Diploma in Commercial Studies	CSA112	11
	Associate Diploma in Commercial Studies	ECONA121	86
	Associate Diploma in Commercial Studies	ENGLA111	29
	Associate Diploma in Commercial Studies	ENGLA112	32
	Associate Diploma in Commercial Studies	ENGLA120	50
	Associate Diploma in Commercial Studies	ENGLA210	66
	Associate Diploma in Commercial Studies	FINA200	63
	Associate Diploma in Commercial Studies	GSA111	33
	Associate Diploma in Commercial Studies	HIST121	68
	Associate Diploma in Commercial Studies	MATHA111	37
	Associate Diploma in Commercial Studies	MGTA121	58
	Associate Diploma in Commercial Studies	MGTA251	11
	Associate Diploma in Commercial Studies	MGTA290	20
	Associate Diploma in Commercial Studies	MGTA299	32
	Associate Diploma in Commercial Studies	MISA138	16
	Associate Diploma in Commercial Studies	MISA210	28
	Associate Diploma in Commercial Studies	MISA238	23
	Associate Diploma in Commercial Studies	MISA260	21
	Associate Diploma in Commercial Studies	MKTA120	88
	Associate Diploma in Commercial Studies	MKTA221	27
	Associate Diploma in Commercial Studies	STATA231	20
	Use the table's header to filter data		

Fig. 10 Timetable Statistics for Next Semester Courses for Commercial Studies Program

Appendix A: Evaluation Questionnaire

University Of Bahrain
College of Applied Studies
Business and IT Programs

Multidisciplinary Web-based Higher Education Advisory (HE-Advisor) System Evaluation Form

Most questions are followed by a scale such as –SD 1 2 3 4 5 SA”. The scale ranges from strongly disagree (SD) through strongly agree (SA). The midpoint on the scale represents a neutral position of neither disagreement nor agreement. The respondent is expected to circle his answer. Some questions require a written answer, a few others require a tick (✓).

I. PROGRAM GOALS

Describe briefly the purpose of this system:

What does the system actually offer?

II. PROGRAM CONTENT

The content of the system reflects your plan of study. SD 1 2 3 4 5 SA
The system content (such as study plan, courses, student info .. etc.) is accurate. SD 1 2 3 4 5 SA

III. AUDIENCE FOR CBI PROGRAM

Who are the author’s intended users for the system (you may make more than one tick)?

Students Educators (Instructors/Advisors/Chair persons .. etc) Parents
 Others: _____

Who could benefit from this system (you may make more than one tick)?

Students Educators (Instructors/Advisors/Chairpersons .. etc) Parents
 Others: _____

The level of difficulty is appropriate for the system users. SD 1 2 3 4 5 SA

The system users have the necessary prerequisites SD 1 2 3 4 5 SA

The readability is appropriate for the users. SD 1 2 3 4 5 SA

IV. INSTRUCTIONAL STRATEGIES

The system uses appropriate advising strategies. SD 1 2 3 4 5 SA

The use of graphics, sound, and color contributes to the user’s achievement of the objectives. SD 1 2 3 4 5 SA

The system provides interesting and valuable advice. SD 1 2 3 4 5 SA

The system has intelligent capabilities. SD 1 2 3 4 5 SA

V. PROGRAM DESIGN

The system makes effective use of feedback to the user. SD 1 2 3 4 5 SA

The screen displays are readable, logically arranged, and pleasing to look at. SD 1 2 3 4 5 SA

The user can go directly to any part within the system. SD 1 2 3 4 5 SA

The system can be stopped and restarted at any desired place. SD 1 2 3 4 5 SA

The user can easily start the system. SD 1 2 3 4 5 SA

VI. APPROPRIATE USE OF COMPUTERS

The system takes advantage of the interactive capability of the computer. SD 1 2 3 4 5 SA

This system is a reasonable use of computers in education. SD 1 2 3 4 5 SA

In your opinion what other mediums could be used for student advising?

VII. PROGRAM TECHNIQUES

The system runs properly. SD 1 2 3 4 5 SA

The system directions are clear. SD 1 2 3 4 5 SA

The system uses consistent commands and directions throughout. SD 1 2 3 4 5 SA

VIII. COST/BENEFIT ANALYSIS

The students will benefit from this system. SD 1 2 3 4 5 SA

The educators (Instructors/Advisors/Chairpersons .. etc) will benefit from the system. SD 1 2 3 4 5 SA

What is your expected usage time for an advising session?

Average time Minimum Time Maximum Time

The system is worth using and will improve advising in general. SD 1 2 3 4 5 SA

IX. OVERALL EVALUATION

Identify the system weaknesses:

Identify the system strengths:

What is your overall evaluation of the system? Bad 1 2 3 4 5 Excellent

This system should be adopted by our college. SD 1 2 3 4 5 SA

Your Comments:

Reviewer's Details

Student ID:

Name:

Contact Number(s):

Email(s):

Design and Implementation of RADIUS – An Network Security Protocol

GJCST Computing Classification
C 2.0, D 4.6

Md. Hashmathur Rehman¹ Dr.A. Govardhan² T. Venkat Narayana Rao³

Abstract-RADIUS (Remote Authentication Dial in User Service) is a protocol used for authentication, authorization and accounting of network objects in networking environment. The protocol has set of weaknesses due to its implementation. First the overview presents the basic operation and functioning of RADIUS protocol. Then analysis part focuses on Vulnerability issues such as Security, transport and implementation. Finally, how to minimize or resolve various issues of the RADIUS protocol using deployment best practices and extensions are discussed.

Keywords-Radius; authentication; authorization; accounting; security; extension, Network Access Server.

I. INTRODUCTION

Remote Authentication Dial in User Service (RADIUS) is an AAA (authentication, authorization and accounting) protocol for applications such as network access or IP mobility. It is commonly used for network devices such as routers, modem servers, switches, etc. RADIUS is currently the de-facto standard for remote authentication and accounting. There are many Vulnerability issues with RADIUS; these issues can be viewed as security issues, transport issues and implementation issues. RADIUS consistently provides some level of protection against a sniffing, active attacker, but it indicates that the RADIUS protocol still exist several leaks. It makes up the security shortcomings of RADIUS protocol to a certain extent and makes the RADIUS system meet the requirement of application. The rest of the paper is presented as follows. Section II gives the overview of RADIUS protocol and presents basic security mechanism used by the protocol. Section III analyses the RADIUS protocol, focusing in security, transport and implementation issues. However it does not cover RADIUS protocol's accounting functionality. Section IV presents the concrete implementation of the extended RADIUS. Finally, the paper concludes in section V

II. PROTOCOL OVERVIEW

A. Basic Information

The newest RADIUS protocol is described in RFC 2865[1], "Remote Authentication Dial-in User Service (RADIUS)," and RFC 2866[2], "RADIUS Accounting". RADIUS

protocol is used between two servers. RADIUS server is a shared authentication server that has a list of valid clients. There is a shared secret between the RADIUS server and these clients. This secret cannot be empty, but otherwise it is not defined by the protocol standard how strong it must be. It is only recommended that it is 16 octets minimum and unguessable. This secret is used for to authenticate the RADIUS server to the NAS and to hide the user password. For these purposes the secret is part of value that is hashed and the hash value is sent. [5]. RADIUS server also has a database of users containing their passwords, possible other requirements for these users to gain access and configuration data. According to information in this database the RADIUS server accepts or rejects the request or sends a challenge to user. RADIUS server can also act as a proxy relaying requests to other RADIUS server and to NAS [5], [15]. When acting as proxy RADIUS server relays messages between the NAS and other RADIUS server. There can be many RADIUS servers as proxies between the NAS and the RADIUS server that finally handles the authentication and authorization of the request. Network Access Server (NAS) acts as a client to the RADIUS server. Users call in and NAS prompts for needed authentication information, for example user name and password. The NAS then can use RADIUS server for user authentication. When doing so the NAS sends request to the RADIUS server containing attributes that have information about user that the RADIUS server needs. When sending request containing user password, the password is not sent as clear-text, instead it is encrypted as described in section 4. NAS then waits for reply from the RADIUS server. Server can accept or reject the request or present a challenge for the user to respond. If request is accepted the server can also provide the NAS with configuration data and type of service granted for the user. If RADIUS server does not response in given time, NAS can retransmit the request or it can also use possible alternate RADIUS servers. In figure 1 basic RADIUS system architecture operation is shown in Figure 1 [5], [6], and [9].

RADIUS protocol uses UDP as its means of transport. UDP port assigned for RADIUS protocol is 1812. Previously RADIUS used UDP port 1645, but usage of this port conflicted with datametrics service. Choice for using UDP instead of TCP is mainly for the reason that UDP is lighter protocol than more reliable TCP. RADIUS is a stateless protocol that does not carry much data as maximum size for UDP packet 4096 octets. As RADIUS is used for user authentication, few seconds delay is acceptable. In addition to complete the authentication and authorization does not

About-¹Cisco Systems(i) Pvt Ltd, Hyderabad A P,India
(e-mail: hasmoham@cisco.com)

About-²Jawaharlal Nehru Technological University, Hyderabad, A P, India, (e-mail: govardhan_cse@yahoo.co.in)

About-³Hyderabad Institute of Technology and Management (HITAM),R.R District, Hyderabad, A P, India (e-mail: tvnrbbobby@yahoo.com)

database of authentication information that it can use to validate the client's authentication request.

B. Authentication and Authorization

When NAS wishes to authenticate user via RADIUS, the NAS sends Access-Request packet to RADIUS server. To this packet the NAS set appropriate attributes that describe the needed information about the user and the service required to the RADIUS server. User password in User-Password attribute is sent encrypted and not in clear-text. The NAS also generates unique Request Authenticator for this request and sets Identifier so that the NAS can connect the reply to this request. [10], [13].

Upon receiving this request the RADIUS server checks it's list of valid clients that it has a shared secret with. If the request does not come from a client in this list, the request is not handled and no error message is sent. If the client is valid, the RADIUS server decrypts the user password (if present) and checks it's user database for entry for requesting user and checks whether user passwords match.[6], [5].

If user is not found, passwords do not match or the user is not allowed to specific clients or ports that may be defined listed in user entry, the RADIUS server send Access-Reject packet to the client. If user is found, passwords are equal, user is allowed to access and no challenge/response is needed, then Access-Accept packet is sent to the client. [3], [4], [5].

For any response Response Authenticator is calculated for this packet and Identifier is identical to that of the request. The Access-Accept packet can have additional information about configuration values in attributes. Access-Reject packet in the other hand can only have attribute hat contains a text message to be shown to the user. Figure 3 shows RADIUS authentication and authorization procedure [11].

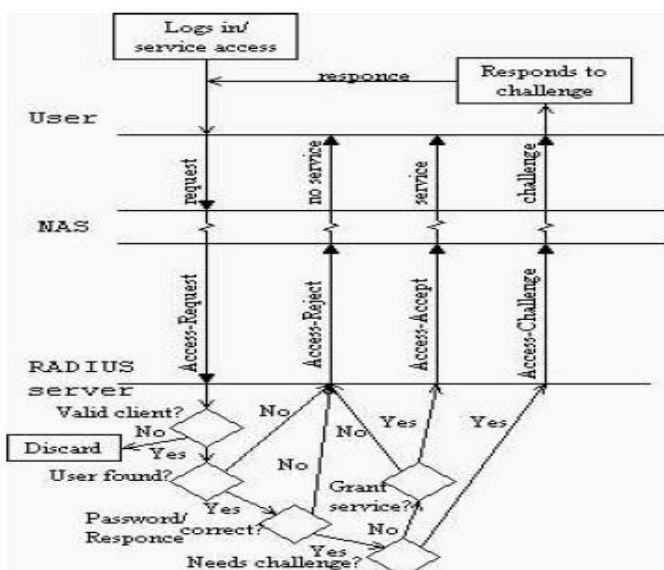


Figure 3: RADIUS authentication procedure.

When the NAS receives the reply, it matches the reply to the request using the Identifier. Then the NAS calculates Response Authenticator for received reply the same way the RADIUS server did and compares this value to Response Authenticator in the message. If these match the the RADIUS server is authenticated and the integrity of the reply is verified [4]. All the fields of the reply (with Request Authenticator in place of Response Authenticator that is being calculated) and attributes of the reply are concatenated with the shared secret and hash value of this concatenation is the Response Authenticator. The Response Authenticator can therefore be used for checking the integrity and to authenticate the RADIUS server, as any change in the message or mismatched shared secret, and the values would not match [5]. RADIUS protocol support also additional challenge/response authentication. In this method the RADIUS server, after receiving Access-Request and having checked the user information from user database, sends Access-Challenge packet to the client. This packet may have an attribute that is message to be displayed to the user. [5] When the NAS receives Access-Challenge packet, it displays the message to the user (if message is present in the attributes) and waits for user's response to this challenge. After user has responded, the NAS resends the original Access-Request packet with new identifier and the users response encrypted in User-Password attribute. If the NAS does not support this challenge/response scheme, it will regard Access-Challenge as Access-Reject [4], [5].

RADIUS server then again checks from it's user database if the response to the challenge was correct. If not, then Access-Reject packet is send. If response was correct, the RADIUS server can send Access-Accept or new Access-Challenge packet. In figure 3 can RADIUS challenge/response operation be seen. [5]. With this challenge/response method RADIUS protocol can use special devices such as one-time-password generators or smart cards to enforce stronger authentication for dial-in users. This enhances RADIUS authentication strength because new innovation in this field can be added as part of RADIUS user authentication process [4].

III. ACCOUNTING

RADIUS accounting is done almost the same way as RADIUS authentication and authorization.

There are some differences. RADIUS accounting uses the UDP port 1813. There are also two RADIUS message codes and 12 attributes for RADIUS accounting. In addition Request Authenticator is calculated differently when using RADIUS accounting [6]. Accounting starts with the NAS sends RADIUS packet with code Accounting-Request having Acct-Status-Type attribute for Start to the RADIUS server. In the accounting start request, the attributes containing information about the user and service being used. All attributes that can be used in Access-Request can also be used in Accounting-Request with five exceptions. These are User-Password, CHAP-Password, Reply-Message, State and CHAP-Challenge attributes [6]. When the NAS wishes to stop the accounting, it sends RADIUS

packet with code Accounting-Request having Acct-Status-Type attribute for Stop to the RADIUS server. This packet can have attributes containing information about the service that was used and statistics of the use [6], [3]. Having received a request packet the RADIUS server then records the Accounting-Request and after successfully recording the packet acknowledges it by sending Accounting-Response packet to the NAS. If the request is not successfully recorded, then no acknowledgment is sent. If the NAS does not get acknowledgment for its request, it will retransmit the request or transmit the request to other RADIUS server. In Accounting-Response packet there are no attributes, except for possible Proxy-State and Vendor-Specific. Figure 4 shows operation of RADIUS accounting [6].

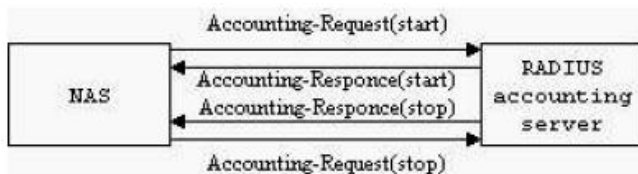


Figure 4: RADIUS accounting.

In Accounting-Request packet the generation of Request Authenticator is different from the generation of Request Authenticator in Access-Request packet. In Access-Request the Request Authenticator is random number, but in Accounting-Request packet the Request Authenticator a hash value so that it will protect the integrity of the request. Accounting-Request Request Authenticator is MD5 hash over concatenation of Code, Identifier, Length, 16 zero octets, attributes in the request and secret shared between the NAS and the RADIUS server[3], [6]. Response Authenticator in Accounting-Response packet is calculated the same way as described for Response Authenticator in section 2.2 [6].

A. Security Measures

Firstly, to provide security for RADIUS messages, the RADIUS client and the RADIUS server are configured with a common shared-secret, which is never sent over the network...

Secondly, the RADIUS protocol adopts Authenticator mechanism. The Authenticator authenticates the reply from the RADIUS server to the NAS and is also used in encryption of User-Password attribute. Two different kinds of Authenticator fields are defined. Request Authenticator is the name of the Authenticator field in Access-Request type packets. It is a random number that the NAS generates in order to be able to authenticate that the reply is intended exactly for the request that the Request Authenticator was generated for. Therefore it must be unique and unpredictable. NAS also uses Request Authenticator when encrypting User-Password attribute. Response Authenticator is the name of the Authenticator field in Access-Accept, Access-Reject and Access-Challenge type packets. Its value is calculated by the RADIUS server. Equation (1) shows formula for Response Authenticator.

Response

Authenticator=MD5(Code+Identifier+Length+Request Authenticator+Attributes+Shared Secret)

$$\dots\dots\dots (1)$$

Thirdly, user password in User-Password attribute is encrypted by stream-cipher. Encryption is done as follows. First password is divided to 16 octet segments, padded is not multiple of 16. So a number of 16 octet segments are gained, now denoted as p_1, p_2, \dots . Then MD5 [13], [14] hash is calculated over concatenation the secret shared between the NAS and the RADIUS server and the Request Authenticator result now denoted as b_1 . Then p_1 and b_1 are XORed, the result denoted as $c(1)$. $c(1)$ is the placed as the first 16 octets of the User-Password attribute. For all values of p_i , i greater than 2, MD5 hash is calculated over concatenation the secret shared between the NAS and the RADIUS server and $c(i-1)$ and p_i is gained. Then p_i and b_i are XORed and $c(i)$ is the result. Finally the User-Password attribute contains concatenation of all $c(1)$ to $c(i)$ values.

$$\begin{aligned}
 b_1 &= \text{MD5}(S+RA) \quad c(1) = p_1 \text{ xor } b_1 \\
 b_2 &= \text{MD5}(S+c(1)) \quad c(2) = p_2 \text{ xor } b_2 \\
 &\dots \\
 b_i &= \text{MD5}(S+c(i-1)) \quad c(i) = p_i \text{ xor } b_i \\
 c &= c(1)+c(2)+c(i) \quad (2).
 \end{aligned}$$

IV. RADIUS ISSUES

The RADIUS protocol has a set of vulnerabilities that are either caused by the protocol or caused by poor implementation and exacerbated by the protocol [7].

A. Protocol Dependent Issues

- According to the protocol, a RADIUS server will not validate Access-Request packet really originated by RADIUS client before (and even after, if packet has no User-Password attribute) decoding all attributes, that is to say, RADIUS access requests need not be authenticated and integrity protected. It opens a packets. And it will make User-Password based password attack possible.
- The RADIUS hiding mechanism uses the RADIUS shared-secret, the Request Authenticator, and the MD5 hashing algorithm to encrypt the User-Password and other sensitive attributes. This is a well-known issue stated in [1]. MD5 is not designed to be a stream cipher primitive; it is designed to be a cryptographic hash [13], [14]. This sort of misuse of cryptographic primitives often leads to subtly flawed systems. It makes User-Password attribute based shared-secret attack and user-password based password attack easy.
- RADIUS encrypts only the password in the access-request packet, from the client to the server. The remainder of the packet is in the clear. Other information, such as username, authorized services, and accounting, could be captured by a third party.

The matter will be even worse when dealing with plaintext password authentication (such as Password Authentication Protocol, PAP).

- There is also problem with some vendor-specific RADIUS authentication implementation. For example Microsoft has its specific attributes defined in RFC 2548[8]. These attributes allow MS-CHAP and MS-CHAPv2 authentication via RADIUS. Microsoft doesn't use some cryptographic schema for its MS-CHAP-Challenge, MS-CHAP-Response and MS-CHAP2-Response attribute [9]. This opens possibility of both replay and spoof attack against MS-CHAP and MS-CHAPv2 authentication. There is design flaw in this scenario which makes it vulnerable against Man-in-the-Middle attack.
- The RADIUS protocol does not offer replay attack prevention. An old packet can be replayed without detection by a malicious NAS impersonator. This can result in denial of service(DoS) if the server limits concurrent sessions for a user. Duplicate accounting messages can also create havoc. A malicious RADIUS server impersonator can replay response message to NAS too.
- The RADIUS protocol offers only hop-by-hop security and has no facility for securing Attribute-Value pairs between the NAS and the RADIUS server. This offers proxy servers the opportunity to collect confidential information or modify messages without detection by the endpoints.
- The RADIUS protocol does not allow a server to send unsolicited messages to the NAS. Where server initiated actions are needed, vendors are forced into solutions outside of the RADIUS protocol or solutions involving proprietary extensions to the RADIUS protocol in ways that often compromise interoperability.
- RADIUS runs on UDP, with no defined retransmission or accounting record retention policy, and according to the protocol, the NAS cannot distinguish the cause of the failure from RADIUS server. After several times (can be configured) retransmissions, the RADIUS protocol specifies that messages are silently discarded for a variety of error conditions. If the messages are Accounting-Request, silent discarding may cause RADIUS server lose user's accounting message. In addition, the connectionless nature of UDP means that one spoofed UDP packet is often enough for an attacker.

B. Implementation Dependent Issues

- Request Authenticator in Access-Request packet is a 128-bit quantity intended to be unpredictable and pseudo-random. But bad implementations do not create Request Authenticators that are sufficiently random. It opens possibility to spoof RADIUS

client Access-Request. It is possible for an attacker with the ability to capture traffic between the RADIUS client and server, and to attempt network access to create a dictionary of RADIUS Request Authenticators and the corresponding key stream used to encrypt the User-Password and other attributes. And it can lead to Access-Accept/Reject replay and bring DoS.

- In many RADIUS installations, the same shared secret is used to protect many RADIUS client-server pairs, and many implementations only allow shared secrets and user-passwords that are ASCII characters, and less than 16 characters resulting the RADIUS shared-secrets and user-passwords does not have sufficient randomness to prevent a successful offline dictionary attack. For a guess of the RADIUS shared-secret, the Response Authenticator field and the contents of the Message-Authenticator attribute are easily computed. Offline dictionary attack on RADIUS shared-secrets can be easy.
- In some cases RADIUS may allow privilege escalation: for example user can try to login to NAS via telnet and change Service-Type attribute of Access-Request packet from login to framed to be authenticated by RADIUS. Everything depends on RADIUS and NAS configuration (the good practice for RADIUS server is always send back Service-Type attribute). If Access-Accept doesn't contain Service-Type attribute or NAS doesn't check it, user can login via telnet.
- According to [1], each RADIUS packet can be up to 4096 bytes. It allows putting > 2000 attributes into a single packet. Most implementations of RADIUS servers allocate maximum attribute length for each attributes, it means for each attributes > 256 bytes of memory will be allocated. Therefore, it is possible to lock >512K of memory and amount of CPU time with a single 4K packet. It opens a possibility to spoof source IP for this kind of packets. This is a major weakness in RADIUS protocol rather than all hard-to-exploit cryptographic Man-in-the-Middle issues. Coupled with the spoofing of Access-Request packets with no Message-Authenticator attribute, this may be serious.
- Some of current RADIUS server implementations are derived from Cistron. And most of them have buffer overflow in digest calculation. Probably this overflow can only lead to DoS. Since overflow occurs before packet is checked, it can be exploited from spoofed IP.
- The RADIUS server has local root exploits if its configuration files had open write permissions. So the system is vulnerable to insider attacks. A common way to protect a static encryption key is to save it in a file with restricted access. However, it

is inadequate to prevent a super-user privilege from accessing the static key in its hosting file.

V. RADIUS IMPLEMENTATION AND EXTENSIONS

To address RADIUS issues when deploying a RADIUS solution, the following deployment can be used:

A. Deployment Best Practices

- Use cryptographically strong Request Authenticators. The Request Authenticator value must be changed each time a new Identifier is used.
- Allow the configuration and use of shared-secrets and user-passwords consisting of a random sequence at least 32 hexadecimal digits long or 22 keyboard characters long including a random sequence of upper and lower case letters, numbers, and punctuation. Ideally, the shared-secrets and userpasswords should be computer-generated. Use a different shared-secret for each server- client pair.
- Check the sizes of attributes or messages properly, and ignore inconsistent attributes result in the message. To solve the problem that a single RADIUS packet locks too much memory, have value-pair with variable size data.
- Use the Message-Authenticator attribute in all Access-Request messages. The access server must send Access-Request messages with the Message-Authenticator attribute and the RADIUS server must silently discard the message if the Message-Authenticator attribute is either not present or fails verification [11].
- Configure the NAS to send RADIUS Accounting-Interim-Update attribute. It can keep the accounting information up to the point of the last Accounting-Interim-Update. But this will bring periodic traffic between RADIUS client and RADIUS server [4].
- Set primary RADIUS server and secondary RADIUS server at the same time to realize disaster recovery. They can be all in one LAN or not.
- Limit the number of the same user attempt to login. If a user try to login too much times, use lockout mechanism to lock its accounting.
- Disable PAP authentication by default. Use a strong CHAP (Challenge Handshake Authentication Protocol) challenge when implement CHAP authentication. If implement MS-CHAP or MSCHAPv2 authentication, do not support LAN manager encoding of MS-CHAP challenge responses or password changes.
- Use dynamic user-passwords to gain access. RADIUS is capable of using a strong, two-factor form of authentication, in which users need to possess both a user ID and a hardware or software token to gain access. Token-based schemes use dynamic passwords. Every minute or so, the token generates a unique 4-, 6- or 8-digit access number that is synchronized with the security server. To

gain entry into the system, the user must generate both this one-time number and user ID and password.

B. Extensions

- Design a two server billing mechanism to record accounting messages on two RADIUS servers at the same time. This makes it easy for ISPs (Internet Server Provider) or users to check up if two servers have the same accounting records, to avoid some cheating. Notice that just authenticate a user on one RADIUS server.
- Add a message to traditional RADIUS protocol to achieve the synchronization between RADIUS server and client. The message can be serverinitiated or client-initiated. It will be used when server or NAS want to synchronize with another, for example when RADIUS server wants to restart.
- Extend some vendor-specific attributes. For example, add Error-Info attribute to record error information from RADIUS server, add No-Account-Info attribute to declare that need not to record a user's accounting message because it is in monthly or yearly packet, add VLAN (Virtual Local Area Network) attribute to declare a user in a VLAN. If there are new requirements, it's convenience to add new attribute to the system, but pay attention not to bring insecurity factors.
- The perfect solution is to use it in conjunction with IPSec, if RADIUS traffic cross untrusted network. If the NAS can support IPSec, then the best thing to do is to forsake RADIUS application-layer security entirely and to just run RADIUS over IPSec ESP with a non-null transform. This is described in [6]. Unfortunately, many embedded systems do not have the horsepower or headroom to run IPSec, so RADIUS/IPSec is not widely used today.

VI. WHY MODIFY RADIUS?

So, why attempt to modify RADIUS at all? Why not just go to another (presumably more modern, more secure) protocol? Well, for the most part, the answer is "Because such a protocol doesn't currently exist." In the near future, however, Diameter is likely to be released by the ETF. Diameter is the planned RADIUS replacement. The great majority of all the protocol work that has gone into Diameter has been directed to removing some of the functional limitations imposed by the RADIUS protocol. Effectively no work has been done as relates to the client/server security of the protocol [8], [9]. (CMS is defined, but this is a security layer for the proxy to proxy interaction, not the client to proxy/server interaction) So, does this mean that they continue to use even RADIUS's ad hoc system? No, they removed all security functionality from the protocol. They did the protocol designer's equivalent of punting. Section 2.2 of the current Diameter protocol spec says:

"Diameter clients, such as Network Access Servers (NASes) and Foreign Agents MUST support IP Security, and MAY support TLS. Diameter servers MUST support TLS, but the administrator MAY opt to configure IPsec instead of using TLS. Operating the Diameter protocol without any security mechanism is not recommended."

So, all security aspects of the protocol are handled by IPsec and/or TLS. From a security aspect, this strikes me as a very good idea. Both IPsec and TLS are fully featured (sometimes too fully featured) protocols that many people have reviewed. (That's already much better than RADIUS ever did).

Examining this from a slightly different angle gives me some cause for concern, however. It strikes me that the overhead imposed by a full TLS/IPsec implementation is very significant for many current-day embedded devices. This would seem to indicate that (at least in the near future) manufacturers are going to either continue to use RADIUS or ignore the Diameter standard and perform Diameter without TLS or IPsec.

VII. CONCLUSIONS

RADIUS is a widely used AAA protocol because it is simple, efficient and easy to implement. This paper provided an overview of RADIUS protocol and described how RADIUS security issues are addressed or minimized using implementation, deployment best practices and extensions. These include using strong shared-secrets, the Message-Authenticator attribute, cryptographic-quality values for the Request Authenticator, different shared-secrets for each RADIUS client/server pair, and IPsec to provide data confidentiality for RADIUS messages, and so on. At the same time, many new-world technologies are requiring a secure, peer-to-peer, and reliable framework that not only has the richness of RADIUS but also the flexibility and robustness of Diameter, the next-generation AAA protocol.

VIII. REFERENCES

- 1) [C. Rigney, A. Rubens, W. Simpson and S. Willens. RFC 2865: Remote Authentication Dial In User Service (RADIUS). <http://www.ietf.org/rfc/rfc2865.txt>
- 2) C. Rigney. RFC 2866: RADIUS Accounting. <http://www.ietf.org/rfc/rfc2866.txt>
- 3) G. Zorn, B. Aboba, D. Mitton. RFC 2867: RADIUS Accounting Modifications for Tunnel Protocol Support. <http://www.ietf.org/rfc/rfc2867.txt>
- 4) G. Zorn, D. Leifer, A. Rubens, J. Shriver, M. Holdrege, I. Goyret. RFC 2868: RADIUS Attributes for Tunnel Protocol Support. <http://www.ietf.org/rfc/rfc2868.txt>
- 5) C. Rigney, W. Willats, P. Calhoun. RFC 2869: RADIUS Extensions. <http://www.ietf.org/rfc/rfc2869.txt>
- 6) B. Aboba, G. Zorn, D. Mitton. RFC 3162: RADIUS and IPv6. <http://www.ietf.org/rfc/rfc3162.txt>
- 7) J. Hill. An Analysis of the RADIUS Authentication Protocol. On Bugtraq mailing list, 12 November 2001,
- 8) G. Zorn. RFC 2548: Microsoft Vendor-specific RADIUS Attributes. <http://www.ietf.org/rfc/rfc2548.txt>
- 9) S. Bruce. Cryptanalysis of Microsoft's PPTP Authentication Extensions (MS-CHAPv2). CQRE '99, Springer-Verlag, 1999, pp.192-203.
- 10) P. Calhoun, J. Loughney, E. Guttman, G. Zorn, J. Arkko. Diameter Base Protocol. <http://www.ietf.org/internet-drafts/draft-ietf-aaadiameter-17.txt>
- 11) Andrew S. Tanenbaum, Computer network (Fourth edition), 2007
The Handbook of Applied Cryptography, by Alfred J. Menezes, Paul C. van Oorschot, Scott A. Vanstone.
- 12) Hans Dobbertin. The Status of MD5 After a Recent Attack, CryptoBytes, volume 2, number 2, summer 1996. <ftp://ftp.rsasecurity.com/pub/cryptobytes/crypto2n2.pdf>
- 13) R. Rivest, S. Dusse. RFC 1321: The MD5 Message-Digest Algorithm. <http://www.ietf.org/rfc/rfc1321.txt>
- 14) B. Aboba et al. RFC 2989: Criteria for Evaluating AAA Protocols for Network Access. <http://www.ietf.org/rfc/rfc2989.txt>

DiRiboPred: A Web Tool for Classification and Prediction of Ribonucleases

Bhasker Pant¹ K.R. Pardasani²

{ GJCST Computing Classification }
J. 3

Abstract-Ribonuclease [commonly abbreviated RNase] is a type of nuclease that catalyzes the degradation of RNA into smaller components and can be divided into endoribonucleases and exoribonucleases. All organisms studied contain many RNases of many different classes, showing that RNA degradation is a very ancient and important process. They are shown to have an important role in cancer, tumor and many neuro degenerative disorders for controlling which, in-silico drug designing can be a valuable tool. In the past machine learning has been used to classify other proteins like GPCRs but no attempt has been made for classification of Ribonucleases. Realizing their importance here an attempt has been made to develop an SVM model to predict, classify and correlate all the major subclasses of ribonucleases with their dipeptide composition. The method was trained and tested on 1857 proteins of ribonucleases. The method discriminated Ribonucleases from other enzymes with Matthew's correlation coefficient of 1.00 and 100% accuracy. In classifying different subclasses of Ribonucleases with dipeptide composition, an overall accuracy of 94.534% was achieved. The performance of the method was evaluated using 5-fold cross-validation. A web server DiRiboPred has been developed for predicting Ribonucleases from its amino acid sequence <http://www.bif.manit.org/RiboPred2>.

Keywords-Classifier,Dipeptide,Composition,Ribonucleases, Support Vector Machine.

I. INTRODUCTION

All organism studied to date contain many kind of ribonucleases of different classes. They have a role to play not only in cleaning of cellular RNAs that is no longer required, but also in the maturation of all RNA molecules, both messenger RNAs and non-coding RNAs that function in varied cellular processes. Besides, an active RNA degradation system are a first defense against RNA viruses, and provide the underlying machinery for more advanced cellular immune strategies such as RNAi [1]. RNases play a critical role in many biological processes, including angiogenesis and self-incompatibility in flowering plants [angiosperms]. Also, RNases in prokaryotic toxin-antitoxin systems are proposed to function as plasmid stability loci, and as stress-response elements when present on the chromosome [1].

A Endoribonuclease is a ribonuclease endonuclease which cleaves RNA molecule by attacking the internal bonds. Major types of endoribonucleases are RnaseA, RnaseH, RnaseI, RnaseIII, RnaseL, RnaseP, RnasePhyM, RnaseT1, RnaseT2, RnaseU2, RnaseV1 and RnaseV. An

exoribonuclease is an exonuclease ribonuclease, which are enzymes that degrade RNA by removing terminal nucleotides from either the 5' end or 3' end of the RNA molecule. Enzymes that remove nucleotides from the 5' end are called 5'-3' exoribonucleases and enzymes that remove nucleotides from the 3' end are called 3'-5' exoribonucleases. The major types of exoribonucleases are PNPase, RnasePH, RnaseIII, RnaseR, RnaseD, RnaseT, Oligoribonucleases, ExoribonucleaseI and ExoribonucleaseII [2].

These enzymes have an important role to play not only in normal body physiology but also in diseased conditions. Majority of ribonucleases have been implicated in tumors and cancers like ovarian cancer, melanoma and non-Hodgkin lymphomas. They can directly and indirectly influence cancer causation and spread. For ex. Ribonucleases have been shown to contribute significantly to telomerase inhibitory activity detectable in foregut cancer specimens [1].

These molecules with their immense capability to degrade RNA, on the contrary can be used for curing, treating and preventing many fatal diseases like cancer. Many variants of natural ribonucleases have been developed with minor changes in dipeptide composition with properties included but not limited to stability, cytotoxicity towards pathogenic cells, efficacy of degradation of pathogenic RNA of any origin including viral RNA, evasion of binding by RNase inhibitors, resistance to degradation by proteases, delivery to target cells, efficiency of import into the cell, dose response properties, pharmacokinetic properties, and longevity within the human body [2]. Due to their cytotoxic properties ribonucleases [RNases] can be potential anti-tumor drugs. Particularly members from the RNase A and RNase T1 superfamilies have shown promising results. Among these enzymes, Onconase, an RNase from the Northern Leopard frog, is furthest along in clinical trials [4].

Everyday many new ribonucleases are discovered the annotation and functional assignment of classes to these through wet lab techniques involve time consuming, laborious experiments, hence machine learning techniques like Support Vector Machines can be effectively used to complement them saving time, money and labor.

In this paper an attempt has been made to predict, classify and correlate these enzymes with their dipeptide composition by implementing SVM using SVM Light, a freely down loadable software [11]. This is a novel step where all the major classes of ribonucleases have been taken into consideration.

The typical strategies for identifying Ribonucleases and their types include similarity search based tools, such as BLAST, FASTA and motif finding tools. Although these

About-¹Department of Bioinformatics, Manit, Bhopal, India
(telephone :+91-9425118244, email: pantbhaskar2@gmail.com)

About-²Department of Mathematics, Manit, Bhopal, India

tools are very successful in searching similar proteins, they fail when members of a subfamily are divergent in nature. Hence there arises a need for in-silico prediction which is not only quicker and economical but also accurate which has been proved in previous classifiers. Earlier SVM has been used for classification of biological data like GPCRs [5], Nuclear receptors [6] etc. but no attempt has been made for classification of these proteins.

On the basis of the above study, an online web tool DiRiboPred has been made available at <http://www.bifmanit.org/RiboPred2>. To the best of authors' knowledge, there is no web server that allows recognition and classification of Ribonucleases. Such kind of classifiers will help in annotation of piled up proteomic data and would complement the existing wet lab techniques.

II. MATERIALS AND METHODS

A. Data Repository

Only non fragmented entries were obtained from SwissProt/Uniprot database of ExPasy server [8]. The subclasses of Ribonuclease [Rnases] i.e. Endoribonucleases and Exoribonucleases have member families with much of the redundant data. A package of softwares called CD-HIT Suite [7] was used for removing the redundancy to 90% with sequence identity cutoff of 0.9. The final dataset has following number of proteins.

Endoribonucleases have following members with following number of instances after redundancy removal.

- 1] RNaseA [51]
- 2] RNaseH [682]
- 3] RNaseIII [230]
- 4] RNaseP [286]
- 5] RNaseT1 [15]
- 6] RNaseT2 [36]

Exoribonucleases have following

- 1] PNPase [338]
- 2] RNasePH [184]
- 3] ExoribonucleaseI [06]
- 4] ExoribonucleaseII [29]

For rest of the members the number of instances was very less hence they were not included in the classifier.

B. Proposed Methodology

The classifier, predictor works in 4 step methodology.

- 1] First checks whether the protein is ribonucleases or not.
- 2] If it is, then check whether it is exoribonuclease or endoribonuclease.
- 3] If endoribonucleases then to which subclass it belongs.
- 4] If exoribonucleases then to which subclass it belongs.

C. Recognition of Ribonucleases from rest

At first step, the main aim is the recognition of novel ribonucleases or discriminating ribonucleases from rest of the enzyme protein. A SVM was trained to discriminate the ribonucleases from other proteins. The training and testing was carried out on a dataset of 1857 proteins of ribonucleases. The training also required negative examples for discriminating ribonucleases from other proteins. Since

ribonucleases can be divided into endoribonucleases and exoribonucleases, and comprise several sub-classes within the EC 2.7 [for the phosphorolytic enzymes] and 3.1 [for the hydrolytic enzymes] classes of enzymes, the dataset was extended by including 1857 enzymes other than class 3 and 2 to which ribonucleases belong. The final dataset has equal number of positive and negative examples, so that the performance of the method can be evaluated using single parameter, such as accuracy.

D. Classification of Endoribonucleases

Endoribonucleases has major six subclasses [RNaseA, RNaseH, RNaseIII, RNaseP, RNaseT1 and RNaseT2]. The dataset for classifying this class [endoribonucleases] consisted of 1300 sequences, of which 51 were RNaseA, 682 RNaseH, 230 RNaseIII, 286 RNaseP, 15 RNaseT1 and 36 for RNaseT2 type of enzymes. To achieve second step in our proposed methodology one SVM module with all the 1300 instances belonging to endoribonucleases labeled as positive and 611 sequences belonging to exoribonucleases marked as negative, was constructed. In the third step for classifying an unknown protein into one of the six types of endoribonucleases which is a multiclass classification problem, a series of binary classifiers were developed. Here, six SVMs were developed, one each for a particular type of endoribonucleases. The *i*th SVM was trained with all samples of the *i*th type enzymes with positive label and samples of all other types of enzymes as negative label. The SVMs trained in this way were referred as 1-v-r SVMs [5, 6, 9]. Explaining more explicitly, 6 datasets were prepared. The first dataset for subclass RNaseA comprised of 51 instances labeled positive and rest of the instances belonging to the other 5 were marked as negative, similarly in the second dataset for RNaseH 682 instances were labeled as positive and rest belonging to others as negative. The methodology was repeated for all the 6 subclasses.

In such classification, each of the unknown protein achieved six scores. An unknown protein was classified into the endoribonucleases type that corresponds to the 1-v-r SVM with highest output score.

E. Classification of Exoribonucleases

The same strategy was followed for exoribonuclease class which comprised of four major subclasses [PNPase, RNasePH, ExoribonucleaseI and ExoribonucleaseII]. The dataset for this class consisted of 557 sequences with 338 members of PNPase subclasse, 184 members of RNasePH subclass, 06 members of ExoribonucleaseI and 29 members of ExoribonucleaseII subclasses. Here also one SVM module comprising 557 positive instances and 1300 negatively labelled instances was made. Like above, classification of exoribonucleases is also a multiclass classification problem where 4 SVMs were developed for distinguishing 4 subclasses of exoribonucleases as indicated in step 4 of proposed methodology. The *i*th SVM was trained with all samples of the *i*th type enzyme with positive label and samples of all other types of enzymes as negative label. The classification was achieved in the manner

indicated above. In the final dataset all the instances belonging to 10 subclasses were clubbed together i.e. $51+682+230+286+15+36+338+184++06+29=1857$, and equal amount of negatives were taken which were proteins other than ribonucleases.

F. Support Vector Machine

Kernel-based techniques [such as support vector machines, Bayes point machines, kernel principal component analysis, and Gaussian processes] represent a major development in machine learning algorithms. Support vector machines [SVM] are a group of supervised learning methods that can be applied to classification or regression. Support vector machines represent an extension to nonlinear models of the generalized portrait algorithm developed by Vladimir Vapnik. The SVM algorithm is based on the statistical learning theory and the Vapnik-Chervonenkis [VC] dimension introduced by Vladimir Vapnik and Alexey Chervonenkis. A Support Vector Machine [SVM] performs classification by constructing an N-dimensional hyperplane that optimally separates the data into two categories. SVM models are closely related to neural networks. In fact, a SVM model using a sigmoid kernel function is equivalent to a two-layer, perceptron neural network [10].

Support Vector Machine [SVM] models are a close cousin to classical multilayer perceptron neural networks. Using a kernel function, SVM's are an alternative training method for polynomial, radial basis function and multi-layer perceptron classifiers in which the weights of the network are found by solving a quadratic programming problem with linear constraints, rather than by solving a non-convex, unconstrained minimization problem as in standard neural network training [10].

G. SVM implementation: SVM Light and LibSVM

For implementing SVM, a software called SVMLight developed by Joachims et.al. has been used [11]. In this software there is inbuilt facility for choosing among many kernel types and their parameters. For our study RBF kernel was found to be the best. This kernel nonlinearly maps samples into a higher dimensional space so it, unlike the linear kernel, can handle the case when the relation between class labels and attributes is nonlinear. Furthermore, the linear kernel is a special case of RBF the linear kernel with a penalty parameter C has the same performance as the RBF kernel with some parameters [C, γ]. In addition, the sigmoid kernel behaves like RBF for certain parameters [12], [13]. The second reason is the number of hyperparameters which influences the complexity of model selection. The polynomial kernel has more hyperparameters than the RBF kernel. Finally, the RBF kernel has fewer numerical difficulties. There are two parameters for an RBF kernel: C and γ . It is not known beforehand which C and γ are best for a given problem; consequently some kind of model selection [parameter search] must be done. The goal is to identify good [C, γ] so that the classifier can accurately predict unknown data [i.e. testing data] [14]. For obtaining the value of these two parameters software called LIBSVM by Chih-

Chung Chang and Chih-Jen Lin was used. LIBSVM is an integrated software for support vector classification, [C-SVC, nu-SVC], regression [epsilon-SVR, nu-SVR] and distribution estimation [one-class SVM]. It supports multi-class classification [15]. The value of these two parameters was fed into SVMLight and analysis done.

H. The Attribute

For analyzing these protein enzymes attribute of dipeptide composition was used. Previously this characteristics has been used for solving other biological problems like classification of nuclear receptors, mycobacterial proteins [16], cytokines [17], GPCRs [5] and virulent proteins to name a few, but it has not been used till now for classifying and predicting ribonucleases.

One of the problems with protein classification is that proteins are variable in length. Dipeptide composition provides the information of protein in the form of a fixed vector of 400 dimensions. The dipeptide composition encapsulates the information about fraction of amino acids present in the protein but also local order in the proteins. Hence the information contained in the variable primary amino acid structure of the proteins is converted into a fixed length feature value vector, which is required by SVM and is used here for ribonucleases. The dipeptide composition is calculated by the following formula.

The dipeptide composition is calculated by the following formula.

$$\text{Dep}(i) = \frac{\text{total no. of dep}(i)}{\text{total number of all possible dipeptides}}$$

Where dep(i) is one out of 400 dipeptides.

I. PROCOS [protein composition server]

For feeding the 400 dimensions of dipeptide composition into SVM, it is required to be converted into feature value vectors format eg. 1: 2: -----400: etc. With servers like Copid, dipeptide composition can be easily found out but its manual conversion into feature vector form as indicated above is time consuming and laborious. ProCos is an integrated set of software with inbuilt facility of converting protein composition of any degree in the form required by the user. For our model, dipeptide composition of ribonucleases was calculated using the ProCos software. It is available at www.manit.ac.in/downloads/polycomp/ [18].

J. Evaluation of Performance

Here cross-validation was performed on the dataset. In limited cross-validation, a set of proteins is divided into M equally balanced subsets. The method was trained or developed on $[(M - 1) N]/M$ proteins and then tested on the remaining N/M proteins. This process is repeated M times, once for each subset. In this study, the performance of dipeptide composition based classifiers was evaluated through 5-fold cross-validation [5], [6], [9]. The performance of the classifier developed at the first level [for recognizing proteins of ribonuclease] was evaluated using the standard threshold-dependent parameters, such as sensitivity, specificity, accuracy and Matthew's correlation

coefficient [MCC]. The performance of classifiers for classifying subclasses of ribonucleases was evaluated by measuring accuracy and MCC as described by Hua and Sun [9].

The LibSVM provides a parameter selection tool using the RBF kernel: cross validation via grid search. A grid search was performed on C and Gamma using an inbuilt module of libSVM tools as shown in figure1. Here pairs of C and Gamma are tried and the one with the best cross validation accuracy is picked.

K. Prediction System Assessment

True positives [TP] and true negatives [TN] were identified as the positive and negative samples, respectively. False positives [FP] were negative samples identified as positive. False negatives [FN] were positive samples identified as negative. The prediction performance was tested with sensitivity [TP/ [TP+FN]], specificity [TN/ [TN+FP]], overall accuracy [Q2], and the Matthews correlation coefficient [MCC]. The accuracy and the MCC for each subfamily of ribonucleases, was calculated as described by Hua and Sun [9] and shown below in equation 1 and 2.

$$Accuracy(x) = \frac{tp+tn}{tp+tn+fp+f} \quad \text{EQ.1}$$

$$MCC = \frac{(tp)(tn)-(fp)(fn)}{\sqrt{(tp+fp)(tp+fn)(tn+fp)(tn+fn)}} \quad \text{EQ.2}$$

III. RESULT AND DISCUSSION

The performance of the method in distinguishing ribonucleases enzyme from other enzymes and also various subclasses is shown in Table 1 for training data set. The performance of the method is evaluated using a 5-fold cross-validation. This demonstrates that ribonucleases can be distinguished from other proteins on the basis of dipeptide composition with 100% accuracy. Prediction of endoribonucleases from exoribonucleases also reached higher accuracy of 99.89%. On using the RBF kernel with value of parameters [$\gamma = 0.0078125$ and $C = 0.03125$] an accuracy of 100% was obtained in distinguishing ribonucleases from rest of the proteins. The tabulated C and Gamma values for predicting various classes and subclasses of Ribonucleases for training dataset are given in Table1. The results are also consistent with our previous observation that dipeptide composition is better in classifying the proteins. The dipeptide composition is a better feature to encapsulate the global information about proteins as it provides information about fraction of amino acid contained in the protein as well as their local order. Furthermore, to classify different types of ribonucleases, a series of binary SVMs were constructed. The separate SVM modules have been developed for each type of nucleases of exoribonuclease and endoribonuclease family and the accuracy of classifying different subclasses for test dataset is

indicated in Table2. The average accuracy of dipeptide composition based classifier is 94.534% for the above. This proved that dipeptide composition is an important feature not only for recognizing but also for classifying different types of the ribonucleases. This observation can also be extended to other types of enzymes by establishing good training data.

These results suggest that types of Ribonucleases are predictable to a considerably accurate extent with dipeptide composition. The development of such accurate and fast methods will speed up the identification of drug targets for curing various cancers and also will be helpful in formulation of newer ribonucleases with cytotoxic properties. For many of the classes like RnaseA, RnaseT1, RnaseT2, ExoribonucleaseI and ExoribonucleaseII since the dataset was small, quiet variation was seen in the values of Precision and MCC which were lesser than others.

Fig.1. Coarse Grid Search on C = 2-5, 2-4 ... 210 and Gamma = 25, 24 ... 2 - 10 [Adapted from Xu et al. 2004].

Table1. C and Gamma values for training set of Ribonucleases with accuracies

Modules	C	Gamma	Accuracy
Ribonucleases	0.03125000	0.00781250	100.00%
Endoribonucleases	8.00000000	2.00000000	99.89%
Exoribonucleases	32.00000000	0.00781250	99.96%
ExoribonucleaseI	128.00000000	0.00781250	99.96%
ExoribonucleaseII	0.03125000	2.00000000	99.86%
PNPase	8.00000000	0.00781250	100.00%
RnaseIII	8.00000000	2.00000000	99.72%
RnaseA	512.00000000	0.03125000	99.72%
RnaseH	32	0.0078125	99.97%
RnaseP	32.00000000	0.0078125	99.89%
RnasePH	0.5	0.50000000	100.00%
RnaseT1	512	0.000122	97.89%
RnaseT2	32	0.0078125	98.99%

Table2. Statistical detail and testing accuracies for various classes and subclasses in the testing dataset.

S.No	Ribonuclease	Accuracy	Precision	Tp	Tn	Fp	Fn	Specificity	MCC
1	Pnpase	95.00%	97.00%	120	60	5	4	0.92	0.89
2	Ribonuclease 1	93.12%	75.00%	6	170	11	2	0.94	0.49
3	Ribonuclease 11	93.93%	96.66%	29	147	12	1	0.92	0.8
4	Rnase111	90.00%	99.00%	100	70	10	9	0.88	0.79
5	RnaseA	95.76%	92.72%	51	130	4	4	0.97	0.9
6	RnaseH	95.00%	96.00%	130	50	4	5	0.93	0.89
7	RnaseP	96.29%	96.94%	127	55	3	4	0.95	0.91
8	FinalrnasePH	99.47%	99.29%	140	44	3	2	0.94	0.93
9	FinalrnaseT1	93.65%	88.88%	16	161	10	2	0.94	0.71
10	FinalrnaseT2	93.12%	90.62%	29	147	10	3	0.94	0.78

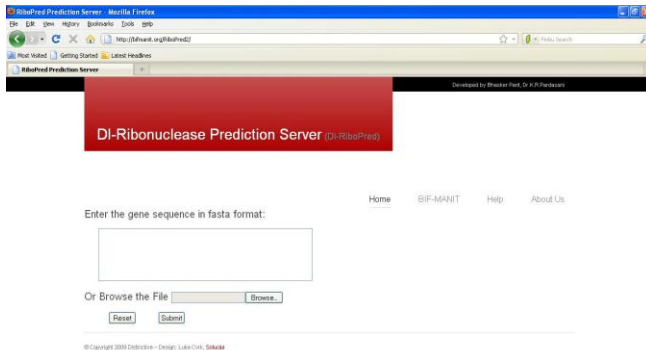


Fig.2. Coarse Grid Search on C = 2-5, 2-4 ... 210 and Gamma = 25, 24 ... 2 - 10 [Adapted from Xu et al. 2004].

IV. DESCRIPTION OF SERVER

RiboPred is freely available at www.bifmanit.org/RiboPred2/. RiboPred server is installed on a Windows Server environment. The user can provide the input sequence by cut-paste or directly uploading sequence file from disk. The server accepts the sequence in standard FASTA format. A snapshot sequence submission page of server is shown in Figure 2. User can predict the type of Ribonucleases based on dipeptide composition. On submission the server will give results in user-friendly format [Figure 2].

V. CONCLUSION

With dipeptide composition as evaluation parameter of ribonucleases, an overall average accuracy of 94.534% was obtained in classifying various subclasses. The tool

DiRiboPred developed at www.bifmanit.org/RiboPred2 can be an efficient and time saving. These kinds of web servers can be an economical and time saving approach for annotation of piled up genomic data. They can be used to effectively complement the existing wet lab techniques. The author awaits discovery of more of these proteins in the future so that more accurate classifiers and tools can be developed.

VI. ACKNOWLEDGEMENT

We are highly thankful to Madhya Pradesh Council of Science and Technology and Department of Biotechnology, New Delhi for providing support in the form of Bioinformatics Infrastructure Facility

VII. REFERENCES

- Holzmann, J., Frank, P., Löffler, E., Bennett, K., Gerner C., & Rossmann, W. (2008), RNase P without RNA: Identification and functional reconstitution of the human mitochondrial tRNA processing enzyme, *Cell*, 135(135), 462–474.
- Alessio, D.G., & Riordan JF. (1997), Ribonucleases: Structures and Functions, Academic Press.
- Gerdes, K., Christensen, S.K., & Lobner-Olesen, A. (2005), Prokaryotic toxin-antitoxin stress response loci, *Nat. Rev. Microbiol.*, (3), 371–382.
- Ardelta, B., & Darzynkiewicz, Z. (2009), Ribonucleases as potential modalities in anticancer therapy. *European Journal of Pharmacology*, (625), 1-3.
- Bhasin, M., & Raghava, G. P. S. (2004), GPCRpred: an SVM-based method for prediction of families and subfamilies of G-protein coupled receptors, *Nucleic Acids Research*, 32[Web Server Issue], W383-W389.

- 6) Bhasin, M., & Raghava, G. P. S. (2004), Classification of Nuclear Receptors Based on Dipeptide composition and Dipeptide Composition, *The Journal of Biological Chemistry*, (279), 23262-23266.
- 7) Huang, Y., Niu, B., Gao, Y., Limin, F., & Weizhong, Li. (2010), CD-HIT Suite: a web server for clustering and comparing biological sequences, *Bioinformatics*.
- 8) Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Ron, D., & Bairoch, Amos. (2003), ExPASy: the proteomics server for in-depth protein knowledge and analysis, *Nucleic Acids Research*, 31(13), 3784-3788.
- 9) Hua, S., & Sun, Z. (2001), Support vector machine approach for protein subcellular localization prediction, *Bioinformatics*, 17, 721-728.
- 10) Ivanciuc, O. Applications of Support Vector Machines in Chemistry. (2007), *Rev. Comput. Chem.*, 23, 291-400.
- 11) Joachims, T., (1999), Making large-scale SVM learning practical. In Scholkopf, B., Burges, C. and Smola, A. [eds], *Advances in Kernel Methods Support Vector Learning*. MIT Press, Cambridge, MA and London, 42-56. SVMlight; <http://svmlight.joachims.org/>
- 12) Keerthi, S. S., & Lin, S. S. (2003), Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Computation*, 15(7), 1667-1689.
- 13) Lin, H.-T., & Lin, C.-J., (2003), A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. Technical report, Department of Computer Science, National Taiwan University.
- 14) Wei, Hsu, C., Chung Chang, C., & Chih-Jen Lin, A., (2003), *Practical Guide to Support Vector Classification*.
- 15) Chang, C.-C., & Lin, C.-J., (2003), LIBSVM: a library for support vector machines.
- 16) Rashid, M., Saha, S., & Raghava, G.P.S. (2007), Support Vector Machine-based method for predicting subcellular localization of mycobacterial proteins using evolutionary information and motifs, *BMC Bioinformatics*, 8:337.
- 17) Lata, S., & Raghava, G.P.S. (2008), CytoPred: a server for prediction and classification of cytokines, *Protein Engineering Design and Selection*, 21(4), 279-282.
- 18) Protein composition server at MANIT, Bhopal, www.manit.ac.in/downloads/polycomp/.

Detecting Copy-Move Forgery in Digital Images: A Survey and Analysis of Current Methods

B.L.Shivakumar¹ Lt. Dr. S.Santhosh Baboo²

GJCST Computing Classification
14.1

Abstract-As one of the most successful applications of image analysis and understanding, digital image forgery detection has recently received significant attention, especially during the past few years. At least two trend account for this: the first accepting digital image as official document has become a common practice, and the second the availability of low cost technology in which the image could be easily manipulated. Even though there are many systems to detect the digital image forgery, their success is limited by the conditions imposed by many applications. For example, detecting duplicated region that have been rotated in different angles remains largely unsolved problem. In an attempt to assist these efforts, this paper surveys the recent development in the field of Copy-Move digital image forgery detection.

Keyword-Image forgeries, Digital forensics, Copy-Move forgery detection, block matching

I. INTRODUCTION

From the early days an image has generally been accepted as a proof of occurrence of the depicted event. Computer becoming more prevalent in business and other field, accepting digital image as official document has become a common practice. The availability of low-cost hardware and software tools, makes it easy to create, alter, and manipulated digital images with no obvious traces of having been subjected to any of these operations. As result we are rapidly reaching a situation where one can no longer take the integrity and authenticity of digital images for granted. This trend undermines the credibility of digital images presented as evidence in a court of law, as news items, as part of a medical records or as financial documents since it may no longer be possible to distinguish whether a given digital images is original or a modified version or even a depiction of a real-life occurrences and objects. Digital image forgery is a growing problem in criminal cases and in public course. Currently there are no established methodologies to verify the authenticity and integrity of digital images in an automatic manner. Detecting forgery in digital images is an emerging research field with important implications for ensuring the credibility of digital images [1]. In the recent past large amount of digital image manipulation could be seen in tabloid magazine, fashion Industry, Scientific Journals, Court rooms, main media outlet and photo hoaxes we receive in

our email. Digital image forgery detection techniques are classified into active and passive approach [3]. In active approach, the digital image requires some pre-processing such as watermark embedding or signature generation at the time of creating the image, which would limit their application in practice. Moreover, there are millions of digital images in internet without digital signature or watermark. In such scenario active approach could not be used to find the authentication of the image. Unlike the watermark-based and signature-based methods; the passive technology does not need any digital signature generated or watermark embedded in advance [4]. There are three techniques widely used to manipulate digital images [3]. 1) Tampering – tampering is manipulation of an image to achieve a specific result. 2) Splicing (Compositing) - A common form of photographic manipulation in which the digital splicing of two or more images into a single composite 3) Cloning (Copy-Move)

II. COPY-MOVE FORGERY

Copy-Move is a specific type of image manipulation, where a part of the image itself is copied and pasted into another part of the same image (Fig 1).

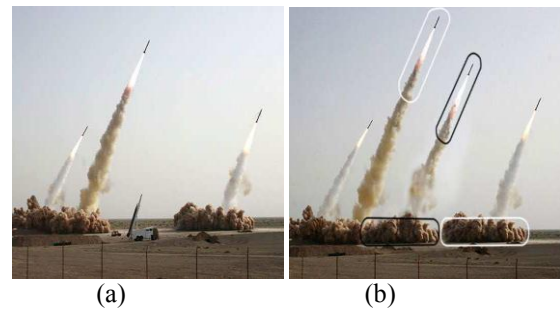


Fig 1. An example of copy-move forgery [5]: (a) the original image with three missiles (b) The forged image with four missiles

Copy-Move forgery is performed with the intention to make an object “disappear” from the image by covering it with a small block copied from another part of the same image. Since the copied segments come from the same image, the color palette, noise components, dynamic range and the other properties will be compatible with the rest of the image, thus it is very difficult for a human eye to detect. Sometimes, even it makes harder for technology to detect the forgery, if the image is retouched with the tools that are available.

¹Associate Professor, Department of Computer Applications, S.N.R. Sons College Coimbatore -641006, India

²Associate Professor, PG and Research Dept. of Computer Applications, DG Vaishnav College, Chennai - 600 106, India

III. COPY-MOVE FORGERY DETECTION TECHNIQUES

The simplest way to detect a Copy-Move forgery is to use an exhaustive search. In this approach, the image and its circularly shifted version are overlaid looking for closely matching image block. This approach is simple and effective for small-sized images. However, this method is computational expensive and even impractical for image of medium-sized. In this method for an image size $M \times N$ it would take $(MN)^2$ steps, since the comparison and image processing require the order of MN operations for one shift. Another technique for detecting forgery is based on autocorrelation. All Copy-Move forgery introduces a correlation between the original segment and the pasted one. However, this method does not have large computational complexity and often fail to detect forgery. However, in most other approaches the detected image is divided into overlapping blocks. The idea here is to detect connected blocks that are copied and moved. The copied region would consist many overlapping blocks. The distance between each duplicated block pair would be same since each block are moved with same amount of shift. The next challenge would be extracting features form these blocks, which would yield to very similar or same values for duplicated block. Several authors presented to use different features to represent the image block. These blocks are vectorized and inserted into a matrix and the vectors are lexicographically sorted for later detection. The computational time depends upon factor such as number of blocks, sorting techniques and the number of feature. Suppose an image size is $N \times N$, it is divided into $(N - b + 1)^2$ overlapping blocks of size $b \times b$. The blocks are represented as vectors of b^2 dimensions, and sorted in a lexicographical order (Fig 2). Vectors corresponding to blocks of similar content would be close to each other in the list, so that identical regions could be easily detected.

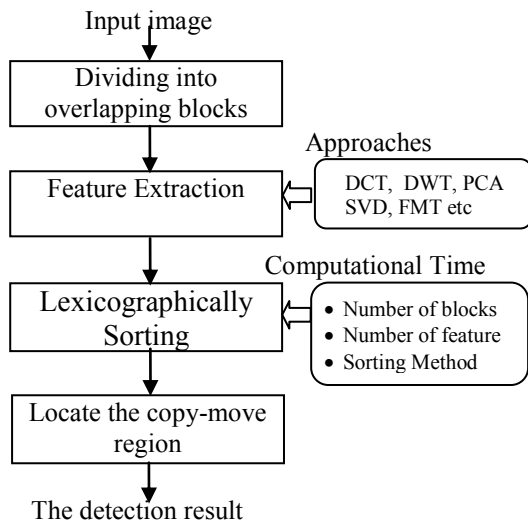
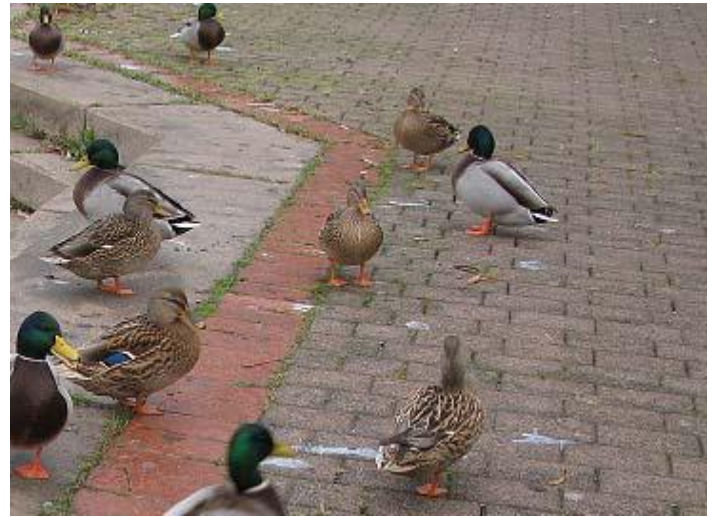


Fig. 2. Configuration of a block Copy-Move Digital Image Forgery Detection System

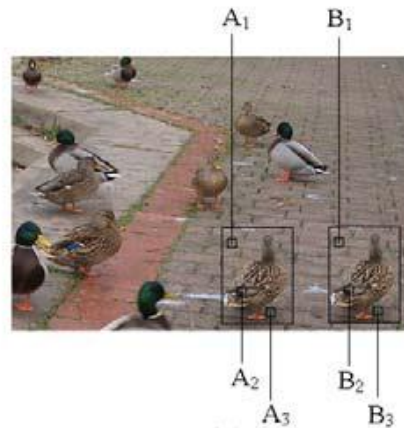
The image given in Figure 3(a) is the original image and Figure 3(b) is the tampered image by Copy-Move Forgery. As shown in Figure 3(c), the block B1, B2, and block B3 which are copies of blocks A1, A2, and block A3, respectively. Therefore, $V_{A1} = V_{B1}$, $V_{A2} = V_{B2}$, and $V_{A3} = V_{B3}$, where V_X denotes the vector corresponding to block X. As shown in sorted list, Figure 3(d), identical vectors are adjacent each other.



(a)



(b)



(c)

Matrix

V_{A1}
V_{B1}
V_{A2}
V_{B2}
V_{A3}
V_{B3}

(d)

Fig. 3 (a). An original image, (b). Forged image (c) Three pairs of identical blocks are marked by squares, (d). Feature

vectors corresponding to the divided blocks are sorting in a list [13]

Over the past 10 years, research has focused on how to make Copy-Move forgery detection system fully automatic. Meanwhile, some significant advances have been made in this field. Nevertheless, many of the finding have important consequences for engineers who design algorithms and system for Copy-Move forgery detection. In the following part of the paper we survey and highlight the summary of research on Copy-Move forgery detection.

A. Region duplication detection: without Scaling and Rotation.

Fridrich et al. [6] suggested the first method for detecting the copy-move forgery detection. In their method, first the image is segmented into overlapping small blocks followed by feature extraction. They employed discrete cosine transform (DCT) coefficients for this purpose. The DCT coefficients of the small blocks were lexicographically sorted to check whether the adjusted blocks are similar or not. In their paper, the method shown was robust to the retouching operations. However, the authors did not employ any other robustness tests.

On the other hand, A.C.Popescu et. al. [7] applied a principle component analysis (PCA) on small fixed-size image to yield a reduced dimension DCT block representation. Each block was represented as 16x16 and the coefficients in each block were vectorized and inserted in a matrix and the corresponding covariance matrix was constructed. The matrix constructed stores floating numbers. By finding the eigenvectors of the covariance matrix, a new linear basis was obtained. Duplicated regions are then detected by lexicographically sorting all of the image blocks. Their method was robust to compression up to JPEG quality level 50 and the time complexity of sorting was $O(32x k lg k)$ time.

Subsequently, G.Li et. al. [8] proposed a method which reduced the time complexity for sorting was reduced to $O(8k lg k)$. The given image was decomposed into four sub-bands by applying discrete wavelet transform (DWT). The singular value decomposition (SVD) was then applied on these blocks of low-frequency component in wavelet sub-band to yield a reduced dimension representation. The SV vector was lexicographically sorted to detect duplicated region. Their method was robust to compression up to JPEG quality level 70. Later on W. Luo et al. [9] suggested a new method based on the pixel block characteristics. The image was first divided into small overlapped blocks and measured block characteristics vector form each block. Then the possible duplicate region was detected by comparing the similarity of the block. In this approach the time complexity for sorting was further reduced to $O(7k lgk)$. Their method was robust to compression up to JPEG quality level 30 and against Gaussian blurring and additive noise with SNR 24 dB.

Myna et al. [10] proposed an approach based on the application of wavelet transform that detects and performed exhaustive search to identify the similar blocks in the image

by mapping them to log-polar coordinates and using phase correlation as the similarity criterion.

Recently, Jing Zhang et al. [12] proposed a new approach based on the idea of pixel-matching to locate copy-move regions. In this approach, DWT (Discrete Wavelet Transform) applied to the input image to yield a reduced dimension representation. Then the phase correlation is computed to estimate the spatial offset between the copied region and the pasted region. The task is to locate the Copy-Move region by the idea of pixel-matching, which is shifting the input image according to the spatial offset and calculating the difference between the image and its shifted version. At the end, the MMO (Mathematical Morphological Operations) are used to remove isolated points so as to improve the location. The proposed technique has lower computational complexity and it is reasonably robust to various types of Copy-Move post processing. However, the performance of this method relies on the location of Copy-Move regions.

Ye et. al.[20] described a passive approach to detect digital forgeries by checking the inconsistencies based on JPEG blocking artifacts. There approach consists of three main steps: i) Collection of DCT statistics ii) Analyses of statistics for quantization tables estimation and iii) Assessment of DCT blocks errors with respect to the estimated quantization tables. The experimental result in their paper shows that the blocking artefact measure of JPEG compression version is 97.1. In this paper, the authors failed to mention how to remove the suspicious tampered regions for estimating quantization table. However, Battiato et. al [21], suggests that such techniques are strictly related with the amount of forged blocks in comparison with the total number of blocks.

All the above copy-move methods are most effective for detection when the region is pasted without any change (scaling or rotation) to another location in the image. However, in practice, the duplicated region is often scaled or rotated to better fit it into the surroundings at the target location. Since, scaling or rotation change the pixel values, a direct matching of pixel is unlikely to be more effective for the detection.

B. Region duplication detection: with Scaling and Rotation.

Recently, Bayram et. al [19] suggested a method by applying Fourier Mellin Transform (FMT) on the image block. They first obtained the Fourier transform representation of each block, re-sampled the resulting magnitude values into log-polar coordinates. Then they obtained a vector representation by projecting log-polar values onto 1-D and used these representations as our features. In their paper, the authors showed that their technique was robust to compression up to JPEG quality level 20 and rotation with 10 degree and scaling by 10%.

Hwei-Jen Lin et. al. [13] proposed a method in which each block B of size $b \times b$ ($=16 \times 16$) by a 9-dimensional feature vector. Unlike other techniques, where the feature vector

extracted stored floating numbers, this method stored them as integer value. The feature vectors extracted are then sorted using the radix sort, which makes the detection more efficient without degradation of detection quality. The difference (shift vector) of the positions of every pair of adjacent feature vectors in the sorting list was computed. The accumulated number of each of the shift vectors was then evaluated and the large accumulated number was considered as possible presence of a duplicated region. The feature vectors corresponding to the shift vectors with large accumulated numbers were detected, whose corresponding blocks are then marked to form a tentative detected result. The final result was obtained by performing connected component analysis and medium filtering on the tentative detected result. Even though, the proposed technique reduced the time complexity to $O(9k)$ with help of radix sort, the method failed to detect all copied region of smaller size. According to their experimental results, the scheme performed well when the degree of rotation was 90, 180 and 270 degree. The figure 3 [13] shows duplicated region with and without rotation.

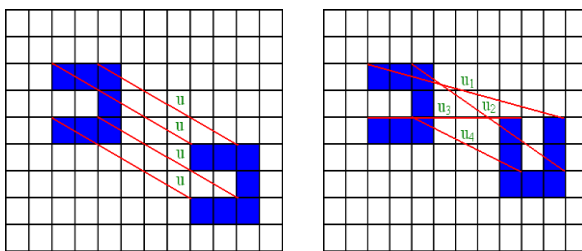


Fig. 3 (a) (b)

- a) Duplicated regions form several identical shift vector u .
- b) Duplicated region from several (different) shift vector (u_1-u_4) , rotated through 90 degree.

H. Huang et al. [11] presented a method to detect region duplication based on local image statistical features known as scale invariant features transform (SIFT). SIFT descriptors of an image are invariant to changes in illumination, rotation, scaling etc. First the SIFT descriptors of the image is extracted, and descriptors are then matched between each other to seek for any possible forgery in images. Even though this method enables to detect duplication, this scheme still have a limitation on detection performance since it is only possible to extract the keypoints from peculiar points of the image

More challenging situation for detection of copy-move forgery is to detect the duplicated region which is rotated some angle before it is pasted. The method presented by [13] to detect duplicated regions in limited rotation angles. More recently Xunyu Pan et. al[14] suggested a method to detect duplicated regions with continuous rotation regions. As described in [14] the new method was based on the image SIFT features

First the SIFT features are collected from the image, and the image is segmented into non-overlapping examination

blocks. The matches of SIFT keypoints in each non-overlapping pixel blocks are computed. After which the potential transform between the original and duplicated regions are estimated and the duplicated regions are identified using correlation map. Even though using SIFT keypoints guarantee geometric invariance and their method enables to detect rotated duplication, these methods still have a limitation on detection performance since it is only possible to extract the keypoints from peculiar points of the image.

Recently, Seung Jin Ryu et. al[15] suggested a method to detect duplicated region using Zernike moments. The authors proposed to use Zernike moments over other technique since they found it to be superior to the others in terms of their insensitivity to image noise, information content, and ability to provide faithful image representation. A detailed review of relevant studies in Zernike moments is beyond the scope of this paper. For details the readers are referred to the papers [16-18]. In their experiment, 12 different images were used to detect Copy-Move forgery with various manipulations such as rotation etc. In the proposed method the image was divided into $M \times N$ overlapped sub-blocks of $L \times L$ and calculated the magnitude of Zernike moments to extract vectors of each sub-block. The vectors were then sorted in lexicographically order. Finally, the suspected region is measured by *Precision*, *Recall*, and *F1-measure* which are often-used measures in the field of information retrieval. The experimental result in their paper show that their system could detect duplicated region rotated some angle before it is pasted, the system is weak against scaling or the other tempering based on affine transform.

IV. CONCLUSION

As Copy-Move forgeries have become popular, the importance of forgery detection is much increased. Although many Copy-Move Forgery detection techniques have been proposed and have shown significant promise, robust forgery detection is still difficult. There are at least three major challenges: tampered images with compression, tampered images with noise, and tampered images with rotation. In this paper we reviewed several papers to know the recent development in the field of Copy-Move digital image forgery detection. Sophisticated tools and advanced manipulation techniques have made forgery detection a challenging one. Digital image forensic is still a growing area and lot of research needed to be done.

V. REFERENCES

- 1) H.T. Sencar, and N.Memon, "Overview of State-of-the Art in Digital image Forensics", World Scientific Press, 2008
- 2) H. Farid, "A Survey of image forgery detection", IEEE Signal Processing Magazine, Vol. pp. 16-25, 2009
- 3) B.L.Shivakumar and S.Santhosh Baboo, "Digital Image Forgery Detection", SAJOSPS, Vol. 10(2), pp. 116-119, 2010

- 4) Lou Weigi, Qu Zhenhua, Pan Feng, and Herang Jiwu, —Survey of Passive Technology for Digital Image Forensics”, *Frontiers of Computer Science in China*, Vol. 1(2), pp. 166-179, May 2007
- 5) Nizza, M., Lyons, P.J.: In an iranian image, a missile too many. In: *The Lede*, The New York Times News Blog(2008) <http://thelede.blogs.nytimes.com/2008/07/10/in-an-iranian-image-a-missile-too-many/>.
- 6) Fridrich, D. Soukal, and J. Lukás, —Detection of copy move forgery in digital images,” in *Proc. Digital Forensic Research Workshop*, Aug. 2003.
- 7) A. C. Popescu and H. Farid, —Exposing Digital Forgeries by Detecting Duplicated Image Regions,” *Technical Report, TR2004-515*, Department of Computer Science, Dartmouth
- 8) G. Li, Q. Wu, D. Tu, and S. Sun, —ASorted Neighborhood Approach for Detecting Duplicated Regions in Image Forgeries based on DWT and SVD,” in *Proceedings of IEEE International Conference on Multimedia and Expo*, Beijing China, July 2-5, 2007, pp. 1750-1753.
- 9) W .Luo, J. Huang, and G. Qiu, —Rbust Detection of Region Duplication Forgery in Digital Image,” in *Proceedings of the 18th International Conference on Pattern Recognition*, Vol. 4, 2006, pp. 746-749.
- 10) A. N. Myna, M. G. Venkateshmurthy, and C. G. Patil, —Detection of Region Duplication Forgery in Digital Images Using Wavelets and Log-Polar Mapping,” in *Proceedings of the International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007)*, Vol. 3, pp. 371-377, 2007.
- 11) H . Huang, W. Guo, and Y. Zhang, —Detection of Copy-Move Forgery in Digital Images Using SIFT Algorithm,” in *Proceedings of IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application*, Vol. 2, pp. 272-276, 2008.
- 12) Jing Zhang, Zhanlei Feng and Yuting Su, —A New Approach for Detecting Copy-Move Forgery in Digital Images”, in: *IEEE Singapore International Conference on Communication Systems*, Guangzhou, China, pp. 362-366, 2008
- 13) Hwei-Jen Lin, Chun-Wei Wang, Yang-Ta Kao, —Fast Copy-Move Forgery Detection”, in *WSEAS Transaction on Signal Processing*, Vol 5(5), pp. 188-197, May 2009.
- 14) Xunyu Pan and Siwei Lyu, —Detecting Image Region Duplication Using SIFT Features”, in: *International Conference on Acoustics, Speech, and Signal Processing*, Dallas, TX, 2010
- 15) Seung-Jin Ryu, Min-Jeong Lee and Heung-Kyu Lee, —Detection of Copy-Rotate-Move Forgery using Zernike Moments”, in: *12th International Workshop on Information Hiding*, Calgary, Alberta, Candada, 2010
- 16) Kim, H.S., Lee, H.K., —Invariant image watermark using Zernike moments”, *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 13(8), pp. 766-775, 2003.
- 17) Khotanzad, A., Hong, Y.H., —Invariant image recognition by Zernike moments”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 12(5), pp. 489-497, 1990.
- 18) Teh, C.H., Chin, R.T., —On image analysis by the methods of moments”, *IEEE Trans. Pattern Analysis and Machine Intelligence* Vol.10(4) , pp. 496–513,1988.
- 19) Sevinc Bayram, Taha Sencar, and Nasir Memon, —Anefficient and robust method for detecting copy-move forgery,” in *Proceedings of ICASSP 2009*, 2009.
- 20) S.M. Ye, Q.B. Sun, and E.C. Chang, —Detecting digital image forgeries by measuring inconsistencies of blocking artifact”, in *Proc. IEEE International Conference on Multimedia and Expo 2007*, Beijing, China, pp.12-15, July 2007.
- 21) Sebastiano Battiato AND Giuseppe Messina, —Digital Forgery Estimation into DCT Domain – A Critical Analysis”, in *Proc of the First ACM workshop on Multimedia in forensics*, China, pp. 37-42 , 2009.

Configuration Complexity: A Layered based Configuration Repository Architecture for conflicts identification

GJCST Computing Classification
F 2.2, I 4.7

Uzma Afzal¹ Dr. Syed Irfan Hyder²

Abstract-An unstable product configures if conflicts (features\ requirements\decision) identify in late phases of software product configuration. This late discovery of conflicts makes the configuration process more complex. We proposed layered base complexity by capturing conflicts at the time of their generation.

Keywords-Configuration, Feature Conflicts, Configuration complexity

I. INTRODUCTION

Today, configuration management (CM) is more important than ever because customers want new designs of products of higher quality at lower prices. Efficient CM can shorten the product life cycle, minimize production cost, and guarantee product quality [1]. Market competitiveness forces product vendors to build flexible products that not only support a specific customer's need but also a group of customers having similar requirements domain. A software product line is a set of software-intensive systems sharing a common, managed set of features that satisfy specific needs of a particular market or mission, and that are developed from a common set of core assets in a prescribed way, according to the definition used by the Software Engineering Institute (SEI) [2]. Product configuration has proven to be an effective means to implement mass customization [3]. Through a configuration process, product modules or components are selected and assembled according to customer requirements [u2] into 4. Product configuration is a collaborative process and Deriving a product from a product line is a complex task requiring the involvement of many heterogeneous stakeholders. Taking their different roles and needs into account is essential to exploit the possible benefits of product lines. Numerous stakeholders need to be supported in understanding the variability provided by the product line. Integration of processes and people is critical. Many critical failures of today's major systems are the consequence of inadequate management and control over an integrated set of components [5]. Abstraction and instantiation are two steps to realize product configuration. So-called abstraction

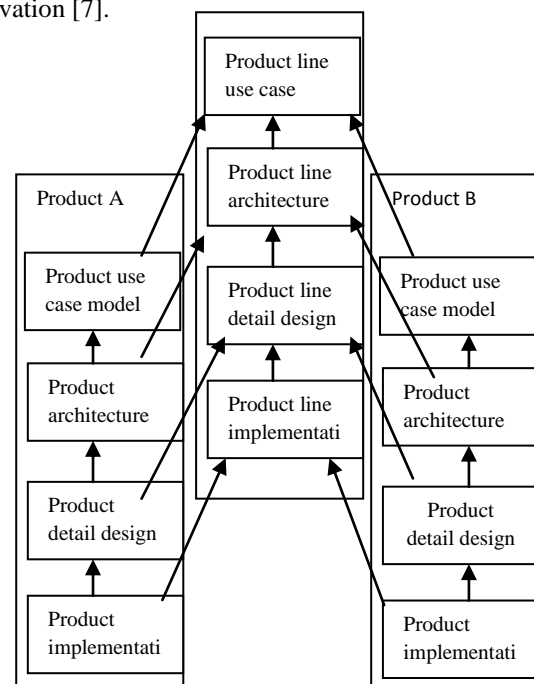
is to elicit a product model from all products, and use a product model, a configuration rule base, and a part instance base to represent all products. Instantiation is according to a customer's demands to confirm the value of every component in the product model tree, and the process of confirmation is based on the product model, configuration rules, and part instance base [6].

II. SPL CONFIGURATION

Software product lines, typically separating two key areas:

- Domain engineering
- Application engineering.

Figure 1 shows the relationship between domain engineering and application engineering. During domain engineering, the variability and commonalities of the product line's reusable core assets such as requirements, architectural elements, or solution components are captured in variability models. A significant body of research is available on modeling approaches and notations for this purpose. During application engineering, concrete products are derived from the product line by selecting, configuring, integrating, and deploying the core assets. Compared to the vast amount of research results on building product lines, few approaches and tools are available for product derivation [7].



About-¹ Department of computer science, Federal Urdu University of Arts Science & Technology (e-mail-u.afzalkhan@gmail.com)

About-² PAF-Karachi Institute of Engineering & Technology

Figure 1 : Modeling Dependencies [8]

Configuration of SPL is a collaborative process and its usual steps are:

- Organization selects the product that meets its business objectives.
- Configuration team starts working.
- Configuration manger splits product into configuration units (configuration repository is single and shared).
- Each configuration unit is assigned to single/multiple developer(s).
- Configuration units are re-assembled into a single product.

III. SPL CONFIGURATION MANAGEMENT

Software product configuration is the process of selecting components from the existing repository and their assembling with the objective of timely, cost effective product delivery. It is, an integral part of any software development activity, takes on a special significance in software product line context. This is due to the special property of software product line, in which the core assets are shared by all products. There are more member products in one product family than in conventional software systems. Hence, in product line, there are much more number of products, assets, and components that needs to be configuration managed. To reduce the working load and the complication of configuration management, it is important to select the right artifacts under configuration management [9].

It also involves identifying the configuration of software (i.e. selected software work products and their description) at given point in time, systematically controlling changes to the configuration through out the software development life cycle [10]. As a result of configuration process, configuration model are produced containing a list of desired product feature [11].

IV. SPL FEATURE MODEL

Features are key distinctive characteristics of a product [12]. A feature design provides a graphical tree like notation that shows the hierarchical organization of features [12]. A feature model is commonly used to guide the configuration process since it breaks down the variabilities and commonalities of product line into a hierarchy of feature as shown in figure 2. Additionally feature model encompass constraints that prevents the derivation of inconsisted product specification i.e. product containing incompatible feature [13].

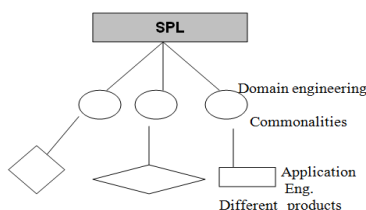


Figure 2 : SPL from domain to application engineering

A feature model allows for inclusion and exclusion of various features and variants so that a valid feature configuration is produced. A feature model also guides product configuration and can be used to validate a particular configuration for conformance [14]. Feature model provide the base for the configuration of whole system. Normally feature model develops in beginning of the development / configuration process. However, Change pervades the entire software life cycle. Requirements change when developers improve their understanding of the application domain. [9]. These changes affect the feature model and its consistency. An invalid feature model leads to an invalid product configuration or it can be said that only consistent and valid feature model gives a successfully configured product Additionally, In global environment, the software configuration becomes critical due to the characteristic of distributed development (physical distance, cultural differences, trust, communication and other factors [10].

V. SPL CONFIGURATION ISSUES

As shown in figure 3, a large-scale product configures from a centralized, shared repository and divides into different modules to make configuration process less complex. Enabling collaborative product configuration brings new and challenging problems such as the proper coordination of configuration decision [13].because a typical software development team consists of multiple developers who work together on closely related sets of common artifacts [15].

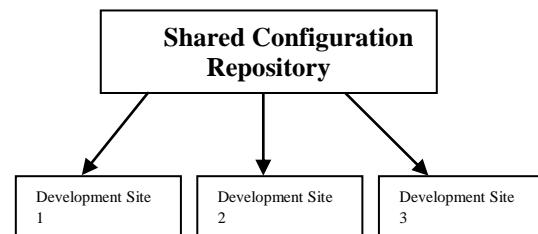


Figure 3 : Configuration from Multiple Sites

The main cause of the system design problems lay with the adhoc way in which large and distributed systems are built, where individual make their own decisions about configuration and life cycle[16] from a configuration and life cycle management perspective failure and recovery was usually inconsistently detected and handled[16].

In an ideal scenario either configuration is collaborative or not, feature model plays an important role in configuration and provides a base and work like a blue print for whole configuration process, only modeled features are configured in final product. Unfortunately, we are not living in an ideal environment in which every thing is according to our desire. Real /practical environment is quite different and it is very clear that the root cause of major configuration issues is the configuration of the products in an ad hoc way where each individual take his own configuration decisions. This late discovery of conflicts makes the configuration process more

complex and strongly affects the cost, efforts and schedule of the product.

A. Is It Really A Problem?

Distributed configuration management is intended to support the activities of project that is configured from multiple sites [16]. Multiple developers from multiple sites configure a product from a product family. SPL variant can not be constructed arbitrarily e.g. a car can not have both ABS and Standard braking software controller. A key step in building a SPL is therefore creating a model of the SPL variability and the constraints on variant configuration [18] however a model is an abstract representation of actual implementation.

In a distributed configuration environment there must be some collaborative mechanism to keep configuration synchronized. For software product configuration management tool support for collaboration on model is therefore crucial [19]. Traditional SCM have support this task for textual artifacts such as source code on the granularity of files and textual lines. They do not work well for graph like models [19]. However SPL product configuration is a decision making process in which group of stake holders chose features for a product [20] and in our collaborative scenario involvement of multiple stakeholders is a basement of product configuration, different configuration units are assigned to different developers that create problem when each individual takes his own configuration decisions (for e.g. feature selection) without going in detail. Integration of the asynchronous efforts of engineers who may be adhering to different configuration management procedures and practice is one of the critical issues [17]. There is a lot of techniques to describe features are existed but common to all of these notation is that they still require maintainers to identify and understand the interaction among features in systems [21].

VI. PROPOSED SOLUTION

An unstable product configures if conflicts are not captured or captured in the late phases of software product configuration so an approach is required to capture these conflicts in earlier stage.

To solve the problem we proposed a Layered based configuration repository (shared) architecture to reduce the configuration complexity by capturing conflicts (Requirements conflicts, features conflicts, decision conflicts) at earlier stage.

We separate the features from the usual configuration repository and proposed a layered based architecture for feature repository and provide facility to exchange information between layers on a common infrastructure to avoid feature/requirement/decision conflicts of collaborative configuration. The service of proposed shared repository does not merely concern storing data but the mechanism for conflicts detection.

A. Architecture Of Proposed Repository

We proposed architecture of the configuration repository that is shared between multiple developers and suggest the

storage of configuration data in layer format. Our repository consists on two main layers and one intermediate communication layer.

Layers are listed below.

- Product domain layer [PDL]
- Intermediate control layer [ICL]
- Product Application layer [PAL]

PDL and PAL will communicate via ICL. Product domain layer is also divided into two parts that are features layer and constraints layer. ICL plays an important role in conflicts identification because no feature will be added to the application layer until or unless Product Application Layer talk to Product domain Layer through Intermediate control Layer.

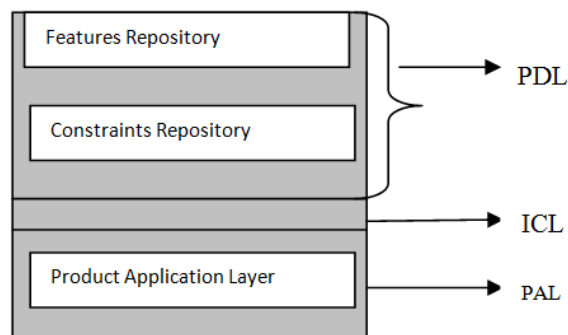


Figure 4: shared repository used by multiple developers from multiple sites

B. Product domain layer

It is the very first layer of Configuration repository and store features and constraints, related to the Product domain.

Features Repository sub layer: Features repository is the base of the product domain layer. Features are key distinctive characteristics of a product [12]. A feature design provides a graphical tree like notation that shows the hierarchical organization of features [12]. A unique identifier is assigned to each feature (naming convention can be used for ease). All features that stored here are the part of the domain of product line or they can be said the core features of product. Different types of features are stored in the repository figure 5 describes the two classifications that are: Independent/dependent and mandatory/variable [22].

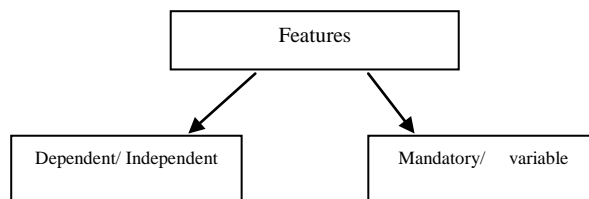


Figure 5: features classification

Independent Features: Because they are not depended on any other feature for their configuration and will not affect the any other component configuration and do not evolve any type of conflicts so only independent feature constraints that apply on them with feature identification tag are stored.

Dependent Features: because they are dependent on other features for their configuration or the configuration of any other dependent feature can affect them so applied constraint with feature identification tag and dependent feature tag are stored in the repository.

Mandatory Features: must be presented in all member products of Software Product Line. Mandatory features illustrate product family commonality [22]. They are stored with a mandatory tag and part of the all variants of any SPL product.

Variable Features: not necessarily appear in all member products in a SPL. Variable features illustrate product family variability [22].

Constraints Repository: It is the second sub layer of the product domain layer that contains all the constraints apply on features. How they stored in repository is dependent on their nature (Uni feature Constraint and multi feature constraint).

These listed constraints are taken from [23] and modified accordingly but it is not the limit other constraints can also be added to the repository.

Mandatory: A feature or a product P requires a feature F.

Optional: The existence of F in P is optional.

Or: In a feature or a product P, there is F1 or F2 or F3... or Fn.

Alternative: if $(P > 0)$ then $\text{sum}(F1, F2, Fn)$ in $\{1..1\}$ else $F1 = 0, F2 = 0, ..Fn = 0$.

Implies: if $(P > 0)$ then $f > 0$. That is, if there is a Feature P in a product, then there must be at least a feature F there.

Excludes: if $(P > 0)$ then $F = 0$. C cannot exist in a product P.

C. Product Application Layer

It is the second layer of proposed layered repository. This layer contains a reference tag for each derived product of the product family, uniquely identified by a Product identifier. As the configuration is moved on and features are configured their unique ids are linked with the product identifier tag by exchanging information from the product domain layer via intermediate layer.

D. Intermediate control layer

It is a middle layer that is used for communication between the two main layers. Both layers talk to each other or exchange information via this communication layer. At the time of product derivation no feature will be added to the PAL until or unless PDL communicate to PAL and find a positive response that the feature addition will not create any feature conflict.

VII. LAYERS COMMUNICATION MECHANISM

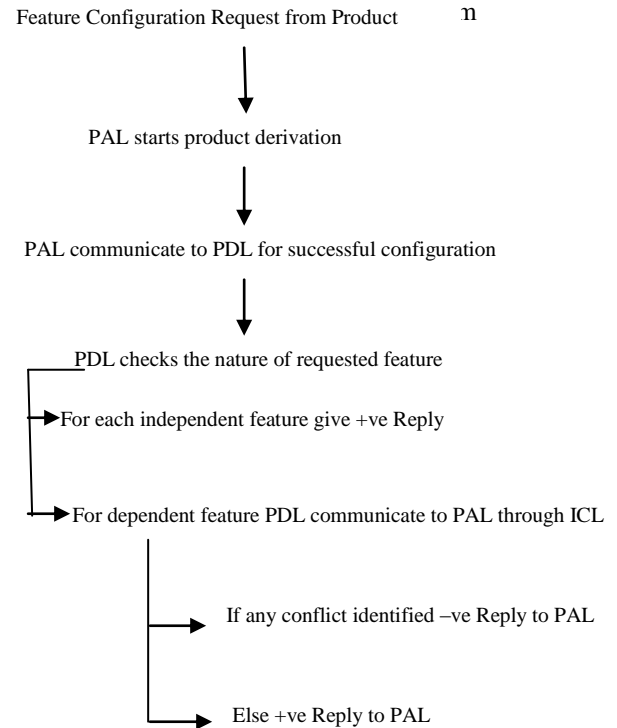


Figure 6 :communication mechanism

VIII. PROTOTYPE & RESULTS

A tool named “Product Configuration Tool” is developed to support the proposed architecture (conflict identification interface is shown in figure 7). An interface is related to each layer of the configuration repository.

Tool has two views.

- For the population of configuration repository
- For the product derivation

Business pattern data of an ERP system is used to validate the repository architecture and its supportive tool. We mapped the business pattern to our proposed schema and then plugged it to the Product Configuration Tool and setup a test environment figure 9 shows a sample of test case. Figure 8 shows the graphical representation of obtained results that proves our thesis statement.

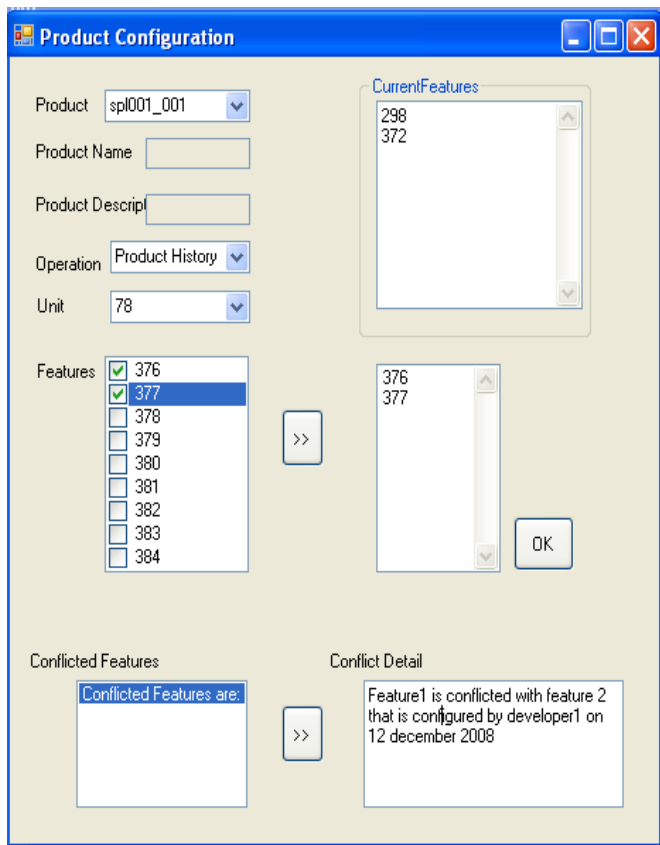


Figure 7 conflict identification view

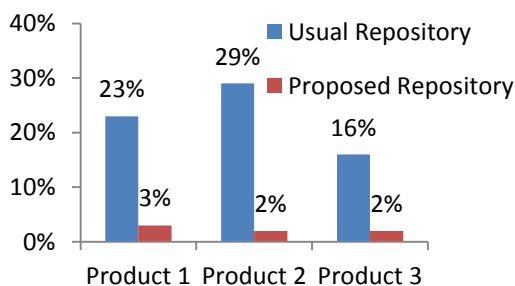


Figure 8: Graphical comparison of existing and proposed repository architecture

Test Case No.	TC_01	Test Status	Pass
Purpose & Scope			
The purpose of this test is to make sure that conflicted are identified to maintain product stability.			
Test strategy: Black Box Testing		Testing Methodology: Manual	
Test Script & Results			
Testing data: Configuration Repository: BP_2_For_Development [ERP Repository] Software Product Line: SPL_01 [Pharma] Derived Product: SPL_001_001 Product History: 298,15425,15463,17279,624 Request of Feature: 502 Constraint Apply: 298 OR 502			
Test Script: <ol style="list-style-type: none"> 1. Launch PCT application 2. From the Product derivation menu click "product derivation" 3. Select Product Id from testing data". 4. Check the history of product to make sure that 298 is added. 5. Select "Add Feature from combo". 6. Select the feature 502 from the automatic generated feature list box. 7. Press OK. 			
Expected Result: Product will not add to the product and a message is displayed to show the conflict detail.			
Exception & Corrective Action			
No.			
Comments & Conclusion			
Actual Result: As per ER Conflicts are identified.			

Figure 9: Sample test case

IX. CONCLUSION & FUTURE WORK

An unstable product configures if conflicts are not identified or identified in the late phases of software product configuration so an approach is required to capture these conflicts in earlier stage. We proposed a Layered based configuration repository (shared) architecture that reduces the configuration complexity by capturing conflicts (Requirements conflicts, features conflicts, decision conflicts) at earlier stage to reduce the configuration complexity.

A “Product Configuration Tool” (PCT) is developed to support the proposed architecture. PCT has two views one for the population of configuration repository and other for the product derivation. Business pattern data of an ERP system is used to validate the repository architecture and its supportive tool.

Future directions include the integration of architecture with existing feature analysis tools and Extend the interface to visualize the model and Enable Architecture to support distributed repository.

X. REFERENCES

- 1) I. Choi and S. Bae: An Architecture for Active Product Configuration Management in Industrial Virtual Enterprises, *Int J Adv Manuf Technol* (2001) 18:133–139 2001 Springer-Verlag London Limited
- 2) Paul Clements and Linda Northrop: *Software Product Lines: Practices and Patterns*, Addison-Wesley, 2001.
- 3) Chunjing Zhou & Zhihang Lin & Chuntao Liu: Customer-driven product configuration optimization
- 4) for assemble-to-order manufacturing enterprises, *Int J Adv Manuf Technol* (2008) 38:185–194 DOI 10.1007/s00170-007-1089-6
Zhang Jinsong · Wang Qifu · Wan Li · Zhong Yifang: Configuration-oriented product modelling and knowledge management formmade-to-order manufacturing enterprises, DOI 10.1007/s00170-003-1871-z *Int J Adv Manuf Technol* (2005) 25: 41–52
- 6) S R Moor, J Gunne-Braden and K J Gleen: Enterprise configuration management — controlling integration complexity, *BT Technol J* Vol 15 No 3 July 1997
- 7) F. S. Zeng . Y. Jin: Study on product configuration based on product model, *Int J Adv Manuf Technol* ,DOI 10.1007/s00170-006-0500-z
- 8) Mag. Rick Rabiser: A User-Centered Approach to Product Configuration in Software Product Line Engineering. Christian Doppler Laboratory for Automated Software Engineering.
- 9) John D. McGregor:Software Product Lines, *JOURNAL OF OBJECT TECHNOLOGY*
- 10) Ligu Yu1 and Srimi Ramaswamy :” A Configuration Management Model for Software Product Line”
- 11) Leonardo Pilatti,, Jorge, Luis Nicolas Audy, Rafael Prikladnicki: Software Configuration Management over a Global Software Development Environment: Lessons Learned from a Case Study, *GSD’06*, May 23, 2006, Shanghai, China.
- 12) [Marcilio Mendonca1, Thiago Tonelli Bartolomei2, Donald Cowan: Decision-Making Coordination in Collaborative Product Configuration, *SAC’08*, March 16-20, 2008, Fortaleza, Ceará, Brazil.
- 13) LUO Daizhong, DIAO Shanhui: Feature Dependency Modeling for Software Product Line, *International Conference on Computer Engineering and Technology*,2009
- 14) Marcilio Mendonca, Krzysztof Czarnecki : Towards a Framework for Collaborative and Coordinated Product Configuration, *OOPSLA’06* October 22-26, 2006, Portland, Oregon, USA.
- 15) Goetz Botterweck1, Steffen Thiel1, Ciarán Cawley1, Daren Nestor1, André Preußner2: Visual Configuration in Automotive Software Product Lines, *Annual IEEE International Computer Software and Applications Conference*,2008
- 16) Anita Sarma, David Redmiles and André van der Hoek :”Empirical Evidence of the Benefits of Workspace Awareness in Software Configuration Management, *SIGSOFT 2008/FSE-16*, November 9--15, 2008. Atlanta, Georgia, USA
- 17) Patrick Goldsack, Julio Guijarro, Steve Loughran, Alistair Coles, Andrew Farrell, Antonio Lain, Paul Murray, Peter Toft: The SmartFrog Configuration Management Framework, *HP Laboratories*
- 18) Andre van der hoek, Dennis Heimbigner and Alexenser I.Wolf: A generic peer to peer repository for distributed configuration management, ,*Proceeding of ICSE-18*
- 19) J.White and D. C. Schmidt , D. Benavides and P. Trinidad and A. Ruiz–Cortés: Automated Diagnosis of Product-line Configuration Errors in Feature Models, *12th International Software Product Line Conference*,IEEE 2008.
- 20) Maximilian Koegel, Jonas Helming, Stephan Seyboth: Operation-based conflict detection and resolution, *CVSM’09*, May 17, 2009, Vancouver, Canada
- 21) Marcilio Mendonca1, Thiago Tonelli Bartolomei2, Donald Cowan1: Decision-Making Coordination in Collaborative Product Configuration , *SAC’08*, March 16-20, 2008, Fortaleza, Ceará, Brazil.
- 22) Maryam Shiri, Jameledine Hassine, Juergen Rilling : Feature Interaction Analysis: A Maintenance Perspective, , *ASE’07*, November 5-9, 2007, Atlanta, Georgia, USA.
- 23) Yuqin Lee and Wenyun Zhao : A feature oriented approach to managing domain requirements dependencies in software product lines, *proceeding of the first international multi symposium on computer and computational science(IMSICC06)*
- 24) Cheng Thao1 Ethan V. Munson1 Tien N. Nguyen2,,:Software Configuration Management for Product Derivation in Software Product Families, *15th Annual IEEE International Conference and Workshop on the Engineering*

VANET Parameters and Applications: A Review

Kamini¹ Rakesh Kumar²

{ GJCST Computing Classification
C 2.1, C 2.m }

Abstract-Vehicular Ad-hoc Network (VANET) represents a challenging class of mobile ad-hoc networks that enables vehicles to intelligently communicate with each other and with roadside infrastructure. VANET poses number of challenges in terms of Quality of Service (QoS) and its performance. Quality of Service depends on numerous parameters such as bandwidth, packet delivery ratio, data latency, delay variance etc. In this paper we have discussed various issues associated with data latency, efficient bandwidth utilization and packet delivery ratio in VANETs. Moreover, challenges in providing security, reliability and confidentiality of the disseminated data are elaborated. Finally, various applications of VANETs in current computing scenario are also presented.

Keywords-Quality of Service (QoS), VANETs, Packet Delivery Ratio (PDR), DSRC (Dedicated Short Range Communication), Orthogonal Frequency Division Multiplexing (OFDM), Direct Sequence Spread Spectrum (DSSS).

I. INTRODUCTION

Work on the ad hoc network begins from 1970s when network were originally called packet radio networks. Inter-Vehicle Communications (IVC) and Roadside-to-Vehicle Communication (RVC) are becoming one of the most popular research topics in wireless communications. Capability of VANET has to provide safety and traffic management: vehicles can notify other vehicles of hazardous road conditions, traffic jamming, or rapid stops. In 1999, the Federal Communication Commission allocated a frequency spectrum for IVC and RVC. Studies in [1, 3] have demonstrated that communications among vehicles can exploit the short-range IEEE 802.11 based radio interface technology. IEEE, 802.11p group specifying the new physical layer and MAC (Medium access control) layer for inter- vehicular communication [2, 3]. Table 1. shows the comparisons between IEEE standards 802.11a, 802.11b and 802.11p.

In 2003, the commission then established the service and license rules for Dedicated Short Range Communications (DSRC) service, which uses the 5.850 to 5.925 GHz bandwidth (75 MHz) for the use of public safety and private applications. Vehicles and roadside base station use the allocated frequency and service to communicate with each other without central access point.

One of the most challenging tasks in VANET is quality of service (QoS) parameters. In wired networks, the QoS parameters are generally described in delay and throughput.

The quality-of-service (QoS) parameter in vehicular ad-hoc network is difficult because the network topology changes with high mobility and the available state information for routing is inherently imprecise. In this paper we have discussed the packet delivery ratio, data latency, efficient bandwidth utilization in data dissemination. The main objective of VANET is to provide safety to vehicles. Applications like collision alert, road surroundings warning, etc. will be classified under safety associated applications where the main accent is on timely broadcasting of safety critical alerts to nearby vehicles. Some challenges of VANET are security, reliability, confidentiality in data transmission that also affects the QoS. Security is provided by different ways like by authentication, encryption etc.

Table 1. Comparison of IEEE 802.11p with 802.11a, 802.11b.

Standard	IEEE 802.11 a	IEEE 802.11 b	IEEE 802.11p (DSRC)
Modulation	OFDM	DSSS	OFDM
Frequency [GHz]	5.725-5.850	2.400-2.485	5.850-5.925
Bandwidth [MHz]	20	22	10/(20)
No. of Channels/ non-overlapping	12/8	14/3	7/7
Max Rate [M Bit/s]	54	11	27/(54)

Rest of the paper is organized as follows: In section 2 we discuss QoS parameters such as data latency, packet delivery ratio, and bandwidth utilization. Section 3 presents the applications of VANET. Section 4 summarizes the challenges of VANET. And finally conclusion is made in Section 5.

II. QoS PARAMETERS FOR VANETS

Dedicated Short Range Communications (DSRC), specified under IEEE standard 802.11p. The IEEE 802.11 standard places the specifications for both the Physical layer (PHY) and for the Medium Access Control layer (MAC) [4]. The MAC extensions are mainly attention to get better security and QoS. The physical layer extensions mostly redefine the way in which the physical layer works. PHY and MAC layers of the VANET planned communication, Wireless Access in Vehicular Environment (WAVE) [5], defined in IEEE 1609.x family of standards. The transmission technology for Intelligent Transportation System (ITS) can be typically classified into two categories, i.e. Vehicle-to-Infrastructure communications (V2I) and Vehicle-to-Vehicle communications (V2V). V2V are achieved by using

About-¹Department of Computer Science and Engineering, M.M. University, Mullana, Ambala (e-mail-kamini.chawla.26@gmail.com)

About-²Department of Information Technology, M.M. University, Mullana, Ambala (e-mail-raakeshdhiman@gmail.com)

effective routing protocol that considers the specific characteristic of the road information, relative car movements and application restriction. Qi's can use to collect the most accurate information, to route the packet from sender to receiver. Qi's is usually defined as a set of service requirements in terms of data latency, bandwidth utilization, and probability of packet delivery ratio.

A. Data Latency

Data Latency means time duration between issuing a message from sender until it is received by receiver vehicles. An important parameter to be considered in sending and receiving a data packet is transmission time delay, through which the throughput rate can be calculated. In order to calculate transmission time delay, the following steps are utilized.

Bit-Rate = Data Size/ Transmission Time Delay

Transmission Time Delay=Data Size/Bit-Rate

Data size = User Data + Header

Authors [6] stress on finding the routing path that has maximum link reliability and a link delay less than an embarrassed bound. For the Qi's features of Deer, link reliability is of higher priority than link delay. This algorithm finds a path with maximum reliability and minimum data latency by NP -complete problem. This algorithm may be most useful in choosing a route for delivering multimedia content or other real-time data that depends on a reliable and minimal delay link.

In [7], they proposed an algorithm that minimizes the number of transmissions while forwarding a message to an access point within the message-specific delay threshold. They compare multi hop data forwarding strategy with Data Mulling strategy to achieve a good tradeoff between communication cost and delay. Data mulling strategy uses message buffer in local memory moving them at the vehicle's speed. Here *ditto*ing be the remaining length, until the next intersection, of the current street segment. *Distain* denotes the current shortest-path distance from the closest access point and *u* the average speed of the vehicle. Algorithm calculates the available delay budget *Del* for data broadcasting from current point to next intersection point as follows:

$$Del = TTL * distToInt / distToAp$$

It also calculates the expected delay by using Data Muling strategy for message dissemination to the next access point as follows:

$$Del_{dm} = distToInt / u$$

Moreover, the high-speed moving vehicles rapidly change the topology of network, and this might result in the potential link breakage of the delivering routes. So as probability of link breakage is high, the value of data latency is also high.

B. Efficient Bandwidth Utilization

The utilization of bandwidth estimation has a tremendous impact on system performance. If the bandwidth estimation is lower than that of network capacity, then the available bandwidth is under-estimated and if the estimation

bandwidth is higher than that of network capacity, then the available bandwidth is more-estimated. In both, systems performance decreases due to inaccurate estimation. In VANET, bandwidth utilization is more as compare to other wireless network due to high mobility in nodes.

One important factor in designing a VANET would be the ability to accommodate vehicles with equipment's of different network characteristics. Range and bandwidth of vehicle equipment may vary. There are number of protocols that assume homogeneous nodes may suffer due to the different properties of each protocol. Also vehicles that have velocity and GPS information will consume less bandwidth than others.

In [8], author estimated the bandwidth consumption from the interference range of the nodes. If the nodes are in their interference range they can easily communicate without any congestion. For sending information, sender checks its neighbor's bandwidth with its own bandwidth with in an interference range. If neighbor node has less bandwidth, then sender shares its own. Estimation of sender local bandwidth by, a node listens to transmission channel and the ratio of idle time and busy time for a predefined interval.

$$b_{local} = b_{channel} * t_{idl} / t_{channel}$$

Where b_{local} is sender's bandwidth, $b_{channel}$ is capacity of channel, t_{idle} denotes the idle time in a predefined interval $t_{channel}$.

In [9], in AODV [10] routing protocol to assisted with the roadside base station. AODV exactly matches the proposal for bandwidth calculation and check bandwidth is apposite for routing. Bandwidth utilization is precisely dependent on the traffic transmitted. They classified traffic as either real-time traffic or non-real-time traffic. The free bandwidth at base station for the request of real-time traffic can be expressed by

$$b_f = b_{unused} + \sum_i (b_{i, curr} - b_{i, min}) - b_e.$$

where *unused* is the unused bandwidth at the base station, *incur* is the bandwidth currently allocated for the non-real-time traffic with index *i*, *brimming* is the minimum bandwidth required for the non-real-time traffic with index *i*, and *be* is the bandwidth reserved for transmission of emergency events.

In VANET, roadside base station consumes more bandwidth, because each base station has more overhead and all time associated to every vehicle. So, if a base station has scarce bandwidth that base station informs to other base stations that it is unable to receive routing information.

C. Packet Delivery Ratio (PDR)

Packet delivery ratio is the ratio of the number of packet received by the destination to the number of packet sent by the sender. It is most significant metric that we should consider in packet forwarding. It may affect by different crucial factor such as packet size, group size, action range and mobility of nodes. The robust message transmission is defined as the 100% packet delivery. Here 100% delivery

means receiver receive all the packets send by sender node before time period expires. The time of the packet delivery for various VANET applications is defined in [11].

The basic idea for PDR is that choose reliable routes. Reliable route need longer predictable lifetime and less number of hops. If the sender have prior information about routes should be chosen instead of the shortest paths which may probably break soon and introduce high maintenance overhead. How to define Routing Overhead? The number of packet transmitted on a route, no matter broadcast or unicast per node. There are some options

- i. The total number of routing packets receives at per node.
- ii. The total number of routing bytes receives, at per node.
- iii. The number of routing packets, count with sequence number, this means end-to-end, not calculated by per node basis.

The link availability prediction [12] requires two nodes maintain their movement patterns during the prediction time. Normally the availability of route depends on the routing overhead. Also each forwarded packet is counted as one transmission. This metric is also highly correlated with the number of route changes occurred in the simulation. A realistic mobility model is not only very important for getting accurate results in routing performance evaluation but also a necessary component to predict the next positions of vehicles and make smarter route decisions in many VANET routing protocols. In [13] authors balances hop minimization with the ability to provide robust routes. From the global perspective of connectivity a new metric called the "expected disconnection degree" (EDD), is introduced to estimate the quality of a route based on factors such as speed, vehicle position and trajectory. It is an estimation of the probability that a given path would be broken during a given time interval. Thus, low EDD route is chosen. Prior the knowledge of vehicle positions, speeds, and trajectories, make some guesses about the stability of a route along a sequence of nodes. Intuitively, route along nodes moving in similar directions at similar speeds are more likely to be more stable.

In [14] solves the problem of path detachment by providing the safe guard which mechanically adjust the connectivity route when sender and receiver nodes change their direction and/or speed.

With a highly dynamic nature of nodes [15, 16], it is not possible to sustain multicast/unicast connections. And packet delivery is dependent on the connection between two nodes. So by using different intelligent techniques such as clustering [17], location aware broadcasting and aggregation [18] performance of packet delivery ratio can be increased

III. APPLICATION OF VANET

VANET communications (IVC and RVC) can be used for number of potential applications with highly diverse requirements. The three major classes of applications possible in VANET are safety oriented, convenience oriented and commercial oriented. Safety applications will

monitor the surrounding road, approaching vehicles, surface and curves of the road. Convenience application will be mainly of traffic management type. Commercial applications will provide the driver with the entertainment and services as web access, streaming audio and video. Below we identify the most representative VANET applications and analyze their requirements through use-cases.

A. Traffic Signal

Communication from the traffic light can be created with the technologies of VANET. Safety applications would be Slow/Stop Vehicle Advisor (SVA) in which a slow or motionless vehicle will broadcast alert message to its neighborhood. Congested Road Notification (CRN) detects and notifies about road congestions which can be used for route and journey planning. The toll collection [19] is yet another application for vehicle toll collection at the toll booths without stopping the vehicles. Vehicular networks have been shown to particularly useful for traffic management. For instance, Vehicle to infrastructure solution for road tolling is widely deployed.

B. Vision Enhancement

In vision enhancement, drivers are given a clear view of vehicles and obstacles in heavy fog conditions and can learn about the existence of vehicles hidden by obstacles, buildings, and by other vehicles.

C. Weather Conditions

Either vehicle sensors (wipers movement, grip control, outside thermometer, etc.); if not available/reliable, weather information can be updated/requested by an application via DSRC. In post-crash notification, a vehicle involved in an accident would broadcast warning messages about its position to trailing vehicles so that it can take decision with time in hand as well as pass information to the highway patrol for support. Parking Availability Notification (PAN) helps to find the availability of space in parking lot in a certain geographical area as per the weather conditions. For the convenience of the vehicle, highway and urban area maps are available which avoid the traffic jam and accident conditions and also provide shortest path in critical situation which saves the time

D. Driver Assistance

Vehicular networks can also be used to support driving military exercises, by providing drivers with information that they might have missed or might not yet be able to see. By [20] having vehicles exhibiting abnormal driving patterns, such as a dramatic change of direction, send messages to inform cars in their locality, drivers can be warned earlier of potential hazards, and therefore get more time to react and avoid accidents. Other applications of vehicular networks to driver assistance include supporting decision making.

E. Automatic Parking

Automatic Parking is an application through which a vehicle can park itself without the need for driver intervention. In order to be able to perform an automatic parking, a vehicle needs accurate distance estimators and/or a localization system with sub-meter precision.

F. Safety

Safety applications include immediate collision warning, forward obstacle detection and avoidance, emergency message dissemination, highway/rail collision avoidance, left/right turn assistant, lane changing warning, stop sign movement assistant and road-condition warning, intersection decision support, cooperative driving (e.g. collision warning, lane merging, etc. [21,22]).

G. Searching Roadside Locations and vehicle's Direction

For unknown passenger help to find the shopping center, hotels, gas stations, etc., in the nearby area along the road. GPS, sensors and database from the nearest roadside base station are capable of calculating information

H. Entertainment

A number of applications aim to entertain passengers who spend a very long period in transit. FleetNet [27] that provides Internet access, as well as communication between passengers in cars in the same vicinity, allowing them to play games. A pure V2V based solutions cannot address these application domains and there is a definite need for V2I infrastructure and VANETs have this V2I support as well.

VANETs would support life-critical safety applications, Safety warning applications, electronic toll collection, internet access, automatic parking, roadside service finder, etc. Table 2. shows the comparisons between the above application on the bases of priority, latency, and network traffic and message range.

We believe that main applications of VANETs are divided into two categories. One is safety applications and another one is non-safety application. In safety applications communications are usually of broadcast type where as in non-safety applications communication is on demand only request response bases (e.g. gaming mobile commerce, multimedia, streaming).

IV. CHALLENGES OF VANET COMMUNICATION

A. Security

esides the introduction and management of trust also the security of message content is a big issue for vehicle to vehicle communication. The content of a received message has to be verified within a short time to be able to use the information as soon as possible.

Table 2. Comparison of VANETs applications

Applications	Priority	Allowable Latency(ma)	Network Traffic	Message Range(m)
Life-Critical Safety	Class1	100	Event	300
Safety Warning	Class 2	100	Periodic	50-300
Electronic Toll Collection	Class 3	50	Event	15
Internet Access	Class 4	500	Event	300
Automatic parking	Class 4	500	Event	300
Roadside Service Finder	Class 4	500	Event	300

Fundamentally, in [23] VANET security should guarantee for the few main issues:-

B. Authentication

The authentication service is concerned with assuring that the communication is authentic in its entities. Vehicle should react to events only with disseminating messages generated by legal senders. Therefore we need to authenticate the senders of these messages.

C. Integrity

The integrity service deals with the stability of a stream of messages. It assures that messages are received as sent, without modification, insertion, reordering, or replays.

D. Confidentiality

This service provides the confidentiality to the communication content. It guarantees the privacy of drivers against unauthorized observers.

E. Accessibility

A kind of attacks can result in the loss or diminution in the accessibility. Even a robust communication channel can still suffer some attacks (such as deny of service) which can bring down the network. Therefore, availability should be also supported by alternative means.

An important feature of VANET security is the digital signature as a building block. Infrastructure communications or communications inter-vehicle through, authentication (using signatures) is a fundamental security requirement.

F. Scalability

The term scalability means that the number of users and/or the traffic volume can be increased with reasonably small performance degradation or even network outage and without changing the system components and protocols.

G. Reliability

Due to the brief communication time, it is difficult to assure the reliable message reception and acknowledgement between communication vehicles on opposite directions. In vehicular ad hoc networks a majority of the messages that are transmitted will be periodic broadcast messages that announce the state of a vehicle to its neighbors. So in case of broadcast messages it needs more reliability. In [24], authors proposed to use a group of vehicles carrying the messages to improve the reliability.

H. Confidentiality

Confidentiality issue is totally related with the security. Vehicles are very costly devices, so the user those who are accepting need to protect their personal data. So there are number of methods to protect user private data. One way to protect data, collect information for a long time from number of source nodes and evaluate that data [25].

I. Media Access Control

To create wide scale vehicular ad hoc networks, changes need to be made to the media access control (MAC) layer [26]. The aim of MAC layer is to access to shared medium, which is the wireless channel. If no method is used to coordinate the transmission of data, then a large number of collisions would occur and the data sent would be lost.

V. CONCLUSION

VANET is not a new research field in network communication. MANET and VANET both share some common features of network. In this paper, we have explained few QoS parameters such as data latency, efficient bandwidth utilization and packet delivery ratio of VANETs, which affects the performance of network communication. However, the performance of VANETs depends heavily on the mobility model, routing protocol, vehicular density, driving environment and many other factors. There are still quite a few parameters that have not been carefully investigated yet like network fragmentation, delay-constrained routing, efficient resource utilization, and delay-tolerant network. Focus of our future work would be on the above said parameters. Nevertheless, VANET shows its unique characteristics which impose both applications and challenges to the research communities.

VI. REFERENCES

- 1) M. Wellens, B. Westphal P. Mahonen, Performance evaluation of IEEE 802.11-based WLANs in vehicular scenarios, IEEE Vehicular Technology Conference, VTC, 2007, pp 1167-1171.
- 2) Spring, (2007) Status of Project IEEE 802.11p, http://grouper.ieee.org/groups/802/11/Reports/tg_update.htm [Cited on: Sept. 2007], (2007).
- 3) K. D. Wong, K. Tepe, W. Chen, M. Gerla, Inter vehicular communication, IEEE Wireless Communications, vol. 13, issue no. 5, October 2006, pp-6-7.
- 4) M. Boban, G. Misek, O. Tonguz, "What is the best achievable QoS for unicast routing in VANETs?", in The 3rd IEEE Workshop on Automotive Networking and Applications, IEEE, New Orleans, USA, December 2008, pp. 1-10.
- 5) "IEEE Trial-Use Standard for Wireless Access in Vehicular Environments (WAVE)-Multi Channel Operation", IEEE Std 1609.4-2006, pp. c1-74.
- 6) Z. Niu, W. Yao, Q. Ni, and Y. Song, "DRReq: A QoS routing algorithm for multimedia communications in vehicular ad hoc networks", in Proc. of the International Conference on Wireless Communications and Mobile Computing, ACM, New York, USA, 2007, pp. 393-398.
- 7) Antonios Skordylis, Niki Trogoni "Delay-bounded routing in vehicular ad-hoc networks", in the 9th ACM international symposium on Mobile ad hoc networking and computing, ACM, New York, USA 2006, pp: 108 - 119.
- 8) Tung-Shih Su, Chih-Hung Lin, Wen-Shyong Hsieh, "A novel QoS-aware Routing for ad hoc networks", in Proc. of Joint Conference on Information Sciences, JCIS 2006, Las Vegas, USA, October 2006.
- 9) Chenn-Jung Huang, Yi-Ta Chuang, You-Jia Chen, Dian-Xiu Yang, I-Fan Chen, QoS-aware roadside base station assisted routing in vehicular networks, The Engineering Applications of Artificial Intelligence, NY, USA, Dec. 2009, pp. 1292-1301.
- 10) Qi Xue and Aura Ganz, "Ad hoc QoS on-demand routing (AQOR) in mobile ad hoc networks", in Journal of Parallel and Distributed Computing, Vol-63, Issue 2, Feb 2003, pp. 154-165.
- 11) National Highway traffic Safety administration, U.S. Department of Transportation, Vehicle Safety Communication Project: Task 3 Final Report, DOT HS 809-859, Department of Transportation, Mar 2005.
- 12) H. Dajing, J. Shengming and R. Jianqiang, A Link Availability Prediction Model for Wireless Ad Hoc Networks, International Workshop on Wireless Networks and Mobile Computing, Taipei, Taiwan, April 2000, pp 7-11.
- 13) Zhaomin Mo, Hao Zhu, Kia Makki, Niki Pissinou, "MURU: A multi-hop routing protocol for urban vehicular ad-hoc networks", in Proc of Mobile and Ubiquitous Systems: Networking & Services, 2006 3rd IEEE Annual International Conference, Jul 2006, pp. 1-8.
- 14) V. Naumov and T. R. Gross, "Connectivity-aware routing (CAR) in vehicular ad-hoc networks", in Proc. of 26th IEEE International Conference on Computer Communications, IEEE, May 2007, pp. 1919-1927.
- 15) G. Liu, B.-S. Lee, B.-C. Seet, C.H. Foh, K.J. Wong, and K.-K. Lee, "A routing strategy for metropolis vehicular communications", in Proc. of the International Conference on Information Networking (ICOIN), 2004, pp. 134-143.

- 15) S.Y. Wang, C.C. Lin, Y.W. Hwang, K.C. Tao, and C.L. Chou, —A Practical Routing Protocol for Vehicle-Formed Mobile Ad-hoc Networks on the roads”, in Proc. of the 8th IEEE International Conference on Intelligent Transportation Systems, 2005, pp. 161–165.
- 16) Grzegorz Wolny, —Modified DMAC Clustering Algorithm for VANETs,” Third International Conference on Systems and Networks Communications, IEEE Computer Society, October 2008, pp.268-273.
- 17) T. Nadeem, P. Shankar, and L. Iftode, —A comparative study of data dissemination models for vanets,” in Mobile and Ubiquitous systems: Networking & Services, 3rd Annual International Conference on Mobile and Ubiquitous Systems, 2006, pp. 1–10.
- 18) Schoch, Elmar, and Frank Kargl, and Michael Weber, —Communication Patterns in VANETs”, IEEE Communications Magazine November 2008, pp. 119-125.
M. Kim, E.-K. Lee, and Y. Lee, —Simulation of intersection collision avoidance system in wireless sensor networks”, in IEEE Geosciences and Remote Sensing Symposium (IGARSS), Denver, CO, Aug. 2006.
- 19) T. ElBatt, S. Goel, G. Holland, H. Krishnan, J. Parikh, —Cooperative Collision Warning Using Dedicated Short Range Wireless Communications”, 3rd ACM International Workshop on VANETs, Los Angeles, California, USA, 2006.
- 20) F. Dötzer, F. Kohlmayer, T. Kosch, M. Strassberger —Secure Communication for Intersection Assistance”, in Proc of the 2nd International Workshop on Intelligent Transportation, Hamburg, Germany, March 15-16, 2005.
- 21) Qian, Yi, Nader Moayeri. —DESIGN SECURE AND APPLICATION-ORIENTED VANET”, National Institute of Standards and Technology. Apr 2009: <http://w3.antd.nist.gov/pubs/Yi-Paper7.pdf>.
- 22) H. Wu, R. Fujimoto, R. Guensler, M. Hunter, "MDDV: A Mobility-Centric Data Dissemination Algorithm for Vehicular Networks," in 1st ACM workshop on vehicular ad hoc networks, Oct. 2004, pp. 47 – 56.
- 23) F. Doetzer, —Privacy issues in vehicular ad hoc networks”, Privacy Enhancing Technologies, Lecture Notes in Computer Science, Springer 2005, pp. 197–209.
- 24) Zeyun N. , Wenbing Y. , Qiang N., Yonghua S., —Study on QoS Support in 802.11e-based Multi-hop Vehicular Wireless Ad Hoc Networks”, in Proc. IEEE International Conference on Networking, Sensing and Control, London, April 2007, pp-705-71
- 25) H. Hartenstein, B. Bochow, A. Ebner, M. Lott, M. Radimirsch, D. Vollmer, —Position-aware ad hoc wireless networks for inter-vehicle communications: the Fleetnet project,” in Proceedings of the 2nd ACM international symposium on Mobile ad hoc networking & computing, Oct. 2001, pp. 259 - 262.

Parametric-Activity Based System Development Method Selection

GJCST Computing Classification
D 2.8, D 2.2

Ajay Jangra¹ Deepa Sharma² Parbhat³ Priyanka⁴

Abstract -In software development life cycle many models have been developed to evaluate and improve capabilities. This paper proposed two enhance tables which provide necessary guidelines to the developer/ organisation on decision making regarding selecting System Development Methodology (SDM) approach by “Comparing traditional and object oriented SDM”. This work is novel in the sense it identify five new parametric activities from SDLC and evaluating characteristic behavior of corresponding to behavior of traditional and object oriented methodology. Furthermore on considering (complexity, testing effort, cost etc.) five parameters are assigned with some weightage distinguish for both system development methodologies. The required result may depend on organization’s decision that how well they create software according to how they define and execute their processes.

Keywords-System development methodology, SDLC, Parametric activity. Object –oriented system design, user, developer.

I. INTRODUCTION

Two important phases of system development are: External and Internal. External development deals with the implementation, planning, preparation of manuals, & installing. Internal Development deals with Software development & performance and testing. Selection process consists of several steps i.e. Requirement analysis, System specifications, Request for proposal, Estimation & validation. The main criteria for software selection depends upon benchmarking which is a evaluation technique where the software purchaser compare the software with other to find the best of Speed & cost by pursue the Reliability, Functionality, Capacity, Flexibility, Usability, Security, Performance, Serviceability, Minimal costs which are the quality factors for SDLC.[1,2,3,9]

II. LITRERATURE SURVEY

A methodology is a route for solving the problems of the current system or for structure a new one. There are many methodologies for the design and development of systems which include: Systems Development Life Cycle (SDLC), Object-Oriented Analysis and Design and many others (Dennis, Wixom, Teagarden, 2002)[4]. The SDLC is more commonly known as Structured Systems Analysis & Design. Structured methodologies allow the analyst to break down complicated systems into smaller, clearly defined and more manageable parts. The structured systems

Development Methodology life cycle moves toward a step by step procedure that goes from one phase to another. The first object-oriented languages came into existence during the 1960's and 1970's with Simula and Smalltalk. However, it was not in anticipation of several years later that the Object-Oriented Analysis and Design (OOAD) methodology came into being (Larman, 2004)[5]. First in 1982 Object-Oriented Design emerged as independent topic (G. Booch, 1982), and later in 1988 Object-Oriented Analysis was introduced by S. Shlaer and S. Mellor (1988) and S. Bailin (1988)[6]. Many different object-oriented analysis and design methods evolved since then such as J. Rumbaugh (1991), P. Coad and E. Yourdon (1991)[7] and many others. The OOAD methodology uses an object-oriented perception rather than a functional perception as in the SSAD methodology. An object is a person, place or thing initially drawn from the problem domain which has three aspects to it: what it knows (its identity and certain attributes), who it knows (relationships to other objects) and what it does (its methods it is responsible for performing on its data) (Norman, 1996)[8].

III. SYSTEM DEVELOPMENT METHODOLOGIES

In paper Tabular guidelines for system development methodology [16] two new C-tables (characteristic & cost (efforts)) were proposed which helps developer/client to select a suitable system development methodology. There are different ways to develop an appropriate system. System development life cycle (SDLC) provides an overall framework for managing the process of system development. Traditional approach and object-oriented approach use the SDLC as a project management framework. There are two main approaches to SDLC: *Predictive* and *Adaptive*. (i) *Predictive approach* assumes project can be planned out in advance (ii) *Adaptive approach* is more flexible, assumes project cannot be planned out in advance. SDLC describes as problem solving methodology which describes software in different stages such as: Organization recognizes problem (project planning), Project team investigates, understands problem and solution requirements (analysis), Solution is specified in detail (design), System that solves problem is built and installed (implementation), System used, maintained, and enhanced to continue to provide intended benefits (support). [1, 2, 14, 15,]

About^{1,2}- CSE deptt. U.I.E.T. Kurukshetra University, India.
(phone +91-9466027922, e-mail er_jangra@yahoo.co.in,)
(phone +91 9729596190, e-mail reach4deepa@gmail.com)

About³ CSE deptt. J.M.I.T. Radaur. Haryana India.
(phone +91-9896713172, e-mail parbhatverma@rediffmail.com)

About⁴ ECE deptt. K.I.T.M. Kurukshetra, India.
(phone +91- 9466751345, priyanka.jangra@gmail.com)

IV. TABULAR ANALYSIS OF PARAMETRIC ACTIVITIES OF TRADITIONAL AND OBJECT ORIENTED APPROACH

The objective is to develop an effective system which suggests whether to go for a traditional approach or object-oriented approach to develop software according to requirements. Proposed C-table in the next segment briefly analyze two approaches with their activities and stages respectively.

A. Proposed C-tables

Here, we proposed five parametric activities (SDLC) and discuss/ Enlist corresponding characteristics, behavior and functions of traditional and object oriented approach. The proposed C¹ table analyzes functional behavior of traditional and objects oriented approach which aggregates

developer vision about the characteristics and behavior of software.

B. Proposed C¹ table for SDLC

The structured approach is well established. There is a lot of CASE tools exist to support development. Most development projects set their own standards that are adopted for analysis and design. The distinct stages make it easier to schedule, distribute work among a number of people. It is easier to express a system in terms of its functions than its data. The structured methods are based on functional decomposition expressed using DFDs. Class diagrams are more similar to ERDs, which are more difficult to model. Entity Relationship Diagrams (ERDs) contain most of the information of the Class diagrams.

Parametric Activities	Traditional	Object oriented
Planning	Define problem and scope. Produce detailed schedule. Confirm project feasibility. resource management	Define problem and scope Produce detailed schedule Confirm project feasibility Staff the project
Analysis	Gather information to learn problem domain Define requirements Build prototypes Generate alternatives Review recommendations	Defines types of objects users deal with Shows use cases are required to complete tasks
Design	Integrate the network Design the application architecture, user interfaces system interfaces and integrate the database and system controls	Defines object types needed Shows objects interaction. Refines the object for implementation with specific language of environment.
Implementation	Construct Verify, test and Convert data. Train users and document and Install the system.	Writing statements in programming language to define what each type of object does.
Support	Maintain, Enhance system Repairs and updates small upgrades expand system capabilities. Support users	CASE tools are designed to help analysts complete system development tasks

Table C¹: Parametric Activity characteristics of traditional and Object oriented methods

The main difference is that classes also define functionality. The development process is both top-down and bottom-up. The problem is partitioned in terms of objects and classes, which is a top down activity. Re-use is considered at all points, during analysis, design and implementation

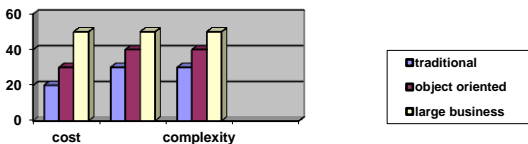


Fig 1

Existing designs, frameworks, patterns, components, classlibraries are considered for re-use. This is a bottom up activity. [10,11,12,13]

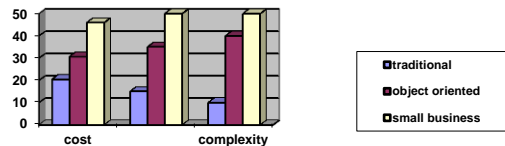


Fig.2

From reference [16] the cost, complexity, testing effort in fig [1,2] required for large and small business software development. The design and implementation will differ only in the level of detail. With corresponding to the above fig. rough estimated weightage are assigned to proposed parametric activities of SDLC and we draw table 2 and fig.3 which shows the cost/effort of both methodologies w.r.t these parametric activities. The effort required with the Object-oriented approach compared to the traditional approach has a difference. Design is much more complex than with traditional development, because of re-use, but

coding requires less effort so does testing. Implementation of traditional is more complex than object oriented approach.

V. PROPOSED C^2 TABLE

Life cycle stages	Traditional approach	Object oriented approach
Requirement	20	20
Design	10	25
Coding	25	20
Implementation	20	15
Support/Testing	25	20

Table C^2

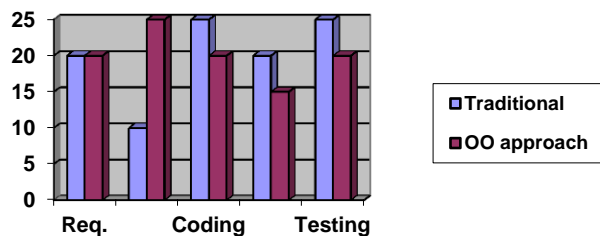


Fig. 3 cost (efforts) v/s Parametric activities for traditional and Object oriented methods

VI. CONCLUSION

Traditional and OOSD are completely different in many terms. With the help of proposed C tables (C^1, C^2) we found OOSD is complex at design time and structured approach is simple. User/developer agreed on selection of software development methodology on the basis of tables. The proposed work helps in planning, staffing, organizing to developer so that he may easily estimates the development level base requirements(resource, efforts). this work helps to forecast the required development efforts and resources in advance which helps developers to manage the software development process efficiently.

VII. REFERENCE

- 1) Daisuke Horie, Toshio Kasahara, Yuichi Goto, and Jingde Cheng, A New Model of Software Life Cycle Processes for Consistent Design, Development, Management, and Maintenance of Secure Information Systems, Eighth IEEE/ACIS International Conference on Computer and Information Sciences, 2009
- 2) Mohammad A. Rob, "Issues of Structured VS. Object Oriented Methodology of System Analysis and Design, Issues in Information Systems, 2004
- 3) Kenneth Pefkaros, "Using Object Oriented analysis and design over traditional structured analysis and design", International Journal of Business Research, March, 2008
- 4) Dennis, A., Wixom, B., and Tegarden, D., Systems Analysis and Design An Object-Oriented Approach with UML, John Wiley & sons, New York, New York, 2002.
- 5) Larman, C., Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design and Iterative Development, 3rd ed., Prentice Hall, Upper Saddle River, New Jersey, 2004
- 6) Bailin, S., Remarks on Object-Oriented Requirements Specification, Computer Technology Associates, Laurel, MD, 1988
- 7) Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F., and Lorensen, W., Object-Oriented Modeling and Design, Prentice-Hall, Englewood Cliffs, New Jersey, 1991.
- 8) Norman, R., Object-Oriented Systems Analysis and Design, Prentice Hall, Upper Saddle River, New Jersey, 1996.
- 9) Craig Larman, Victor R. Basili, "Iterative and Incremental Development: A Brief History," Computer, vol. 36, no. 6, pp. 47-56, June 2003
- 10) Walt Scacchi, Institute for Software Research, University of California, Irvine "Process Models in Software Engineering", Encyclopedia of Software Engineering, 2nd Edition, John Wiley and Sons, Inc, New York, December 2001
- 11) Scacchi, W., Understanding Software Process Redesign using Modeling, Analysis and Simulation. Software Process --Improvement and Practice 5(2/3):183-195, 2000.
- 12) Booch, G., "Object-Oriented Design", Ada Letters Volume 1 (3), 1982, 64-76.
- 13) Bailin, S., Remarks on Object-Oriented Requirements Specification, Computer Technology Associates, Laurel, MD, 1988
- 14) Coad, P., and Yourdon, E., Object-Oriented Analysis, 2nd ed., Yourdon Press, Englewood Cliffs, New Jersey, 1991
- 15) Mauro Pezz`e, Michal Young, Testing Object Oriented Software, 26th International Conference on Software Engineering (ICSE'04)
- 16) A.Jangra, Deepa Sharma, J.Singh, Priyanka, "Tabular guidelines for system development methodology", international Journal of Ultra Scientist of Physical Sciences, ISSN : 0970-9150(2010) Accepted for publication in vol.22 no. 2 aug.2010.

Analyzing Horizontal Integration among Soa, Cloud Computing and Grid Computing

GJCST Computing Classification
C 1.4, C 2.1

Mr. Krunal D. Trivedi¹, Mr. Kamaljit I. Lakhtaria², Mrs. Palak R. Shah³, Dr. N.J. Patel⁴

Abstract- Cloud computing is among emerging technologies for high speed, complex computing. Cloud computing means anything that involves delivering hosted services over the Internet. These services are broadly divided into three categories: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS). Virtual resources and computing assets are accessed through the cloud, including not only externally hosted services but also those provided globally by companies. This provides the basis for the next generation of enterprise data centre which, like the Internet, will provide extreme scalability and fast access to networked users.

In this article, we analyze different standards for Messaging, Security, standards for Application Development and the role of Open Virtualization Format (OVF) in virtualized environment for Cloud Computing. We have suggested security aspect with a proposed architecture for relationship and intersection between Service Oriented Architecture, Grid Computing and Cloud Computing.

Keywords- SOA, Cloud computing, IaaS, PaaS, SaaS.

I. INTRODUCTION

A. To Cloud Computing And Architecture

A cloud service has three discrete characteristics that discriminate it from traditional hosting. It is sold on demand, typically by the minute or the hour; it is elastic -- a user can have as much or as little of a service as they want at any given time; and the service is fully managed by the provider like payment and any search services.

The diagram illustrates (Figure 1) the component architecture. Virtualized Infrastructure allows business services to move directly across virtualized infrastructure resources in a very efficient manner.

About-^{1,2}MCA Department, Acharya Motibhai Patel Institute of Computer Studies, Ganpat University, Mehsana Gujarat, INDIA, Email: trivedi.kunal@gmail.com, Email: palak.kunal@gmail.com

About-³MCA Department, Atmiya Institute of Technology & Science, Rajkot, Gujarat, INDIA

Email: kamaljit.ilakhtaria@gmail.com

About-⁴professor and head U V patel college of engineering E-mail narendraj.patel@ganpatuniversity.ac.in

Security standards, Protocols are defined in Security Management. This Layer also defined the various processes, procedures, and practices necessary for implementing a security program.

Virtualized applications are decouple from the underlying hardware, operating system, storage, and network to enable flexibility in deployment. Virtualized Application servers that can decorate of grid execution coupled with Service Oriented Architectures. And this intersection of Cloud, SOA and Grid provide the greatest scalability to meet the business requirements.

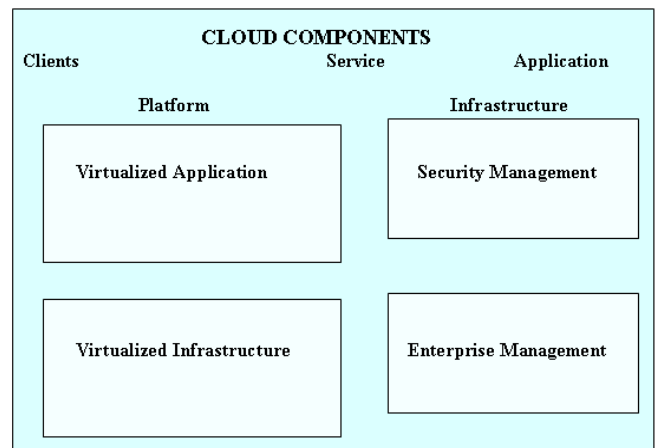


Figure 1: Cloud Components Architecture

Enterprise management provides top-down, end-to-end management of the virtualized infrastructure and applications for business solutions. This layer also handles the full lifecycle of virtualized resources and provides additional common infrastructure elements for service level management, policy management, license management, and disaster recovery. Here is the detail explanation of each Cloud Components.

Table 1: Examples of individual Cloud Components

Cloud Components	Examples
Clients	Mobile like iPhone, Windows Mobile. Thin client like CherryPal, Zonbu. Thick client like Microsoft Internet Explorer, Mozilla Firefox, Google Chrome etc.
Services	Identity (Open Authentication, Open ID) Payments (Amazon Flexible Payments Service, Google Checkout) Mapping (Google Maps, Yahoo! Maps, MapQuest) Search (Alexa, Google Custom Search, Yahoo! BOSS) Video Games (OnLive) Live chat (LivePerson)
Application	Peer-to-peer / volunteer computing (Bittorrent, Skype) Web application (Twitter)
Platform	Amazon.com, eBay, Google, iTunes, and YouTube.
Infrastructure	Full virtualization (GoGrid, Skytap, iland) Compute (Amazon EC2, Rackspace Cloud Servers, Savvis) Platform (Force.com)

B. Example Of Cloud Computing *Www.Facebook.Com*

Face Book statistics are mentioned here

- i. Facebook may have 175 million active users, 11m of which are Canadians. (As on January 2010)
- ii. Facebook is the world's largest photo site, at over 850 million images uploaded per month.
- iii. Facebook uses partitioned/denormalized MySQL databases, but uses extensive caching (using "mem-cache-D") to minimize the number of database requests:
 - 10,000 PHP webservers, 1,000 memcached servers, with 15TB RAM
 - approximately 2000 MySQL servers with 25 terabytes of RAM
 - 10 million requests/sec from the web, reduced to 500K/sec to MySQL through the use of memcached.
- iv. The Facebook front-end tier dumps logs and database extractions into Hadoop/Hive. Hive is Facebook's data warehouse technology that is built over Hadoop: it support an SQL-like query language with support for joins, grouping, and aggregation. From Hive, data summaries are exported into Business Intelligenece applications and/or into an Oracle RAC for analysis, including usage trends, ad-hoc queries, ad optimization, and spam detection. For Facebook, their Hive data warehouse stores 10TB of new data per day (2TB compressed) and totals 500TB compressed, over 600+ nodes with 50+ engineers supporting the warehouse.

II. CLOUD COMPUTING STANDARDS

Cloud Computing have many preferable standards for Messaging, Security and for Application developer.

Cloud computing growth facilitated with powerful and light weight open source standards

A. Open Virtualization Format

The Open Virtualization Format Specification was prepared by the System Virtualization, Partitioning, and Clustering Working Group of the DMTF. OVF 1.0 is the First industry standard to enable interoperability within virtualized environments.

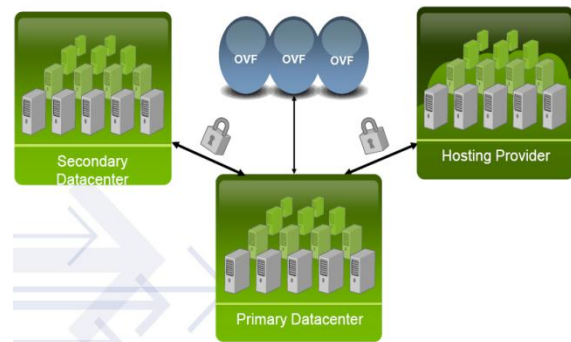


Figure 2: How Open Virtualized Format deals with provider to Data Centre

The Open Virtualization Format (OVF) Specification describes an open, secure, portable, efficient and extensible format for the packaging and distribution of software to be run in virtual machines. The key properties of the format are as follows:

- i. Optimized for distribution: OVF supports content verification and integrity checking based on industry-standard public key infrastructure.
- ii. Supports both single VM and multiple-VM configurations: OVF supports both standard single VM packages and packages containing complex, multi-tier services consisting of multiple interdependent VMs.

- iii. Vendor and platform independent: OVF does not rely on the use of a specific host platform, virtualization platform, or guest operating system.
- iv. Extensible: OVF is immediately useful — and extensible. It is designed to be extended as the industry moves forward with virtual appliance technology. It also supports and permits the encoding of vendor-specific metadata to support specific vertical markets.
- v. Localizable: OVF supports user-visible descriptions in multiple locales, and it supports localization of the interactive processes during installation of an appliance. This capability allows a single packaged appliance to serve multiple market opportunities.
- vi. Open standard: OVF has arisen from the collaboration of key vendors in the industry, and it is developed in an accepted industry forum as a future standard for portable virtual machines.
- vii. Security: The OVF specifies procedures and technologies to permit integrity checking of the virtual machines (VM) to ensure that they have not been modified since the package was produced. This enhances security of the format and will help to alleviate security concerns of users who adopt virtual appliances produced by third parties.

B. LAMP, JSON, AJAX For Application Developer

LAMP- Linux, Apache, Postgre SQL, and PHP (or Perl or Python) is open source web platform that can be used to run dynamic web sites and servers. JSON- Java Script Object Notation is a lightweight computer data interchange format and Asynchronous Java Script Xml are the application standards that can provide that are commonly used across the Internet in browsers, for transferring data, sending messages, and securing data.

C. Web Services, Atom And Atom Publishing Protocol For Messaging

- i. Web Services: For consistent and loosely-coupled architectures. Cloud are constructed with web services as the system interface such as Amazon Web Services, Google Apps, and Salesforce.com CRM.
- ii. Atom: The Atom Syndication Format is an XML language used for web feeds. And it allows software programs to check for updates on a website published. To provide a web feed, a site owner may use specialized software such as a content management system that publishes a list or "feed" of recent articles or content in a standardized and machine-readable format. Web feeds are used by the blog community to share recent entries' headlines, full text, and even attached multimedia files. Atom also provides a standard way to export an entire blog, or parts of it, for backup or for importing into other blog systems.

- iii. Atom Publishing Protocol: The Atom Publishing Protocol (AtomPub or APP) is a simple HTTP-based protocol for creating and updating web resources. APP mechanisms required for the exchange of rich and meaningful content via a process known as Representational State Transfer in cloud application. Atom publishing protocol operates on collections of Web resources.

D. Open Authentication And Open ID For Security

Security standards define the processes, procedures, and practices necessary for implementing a security program. These standards have several processes and procedure those are applying to cloud related IT activities and ensure a secure environment. And the environment is maintained by privacy and security of confidential information. Security standards are based on a set of key principles intended to protect this type of trusted environment. Messaging standards, especially for security in the cloud, must also include nearly all the same considerations as any other IT security endeavours. The following protocols are used in the cloud environment.

- i. Open Authentication: Open Authentication is an open protocol, initiated to allow secure API authorization in a simple, standardized method for various types of web applications. Open Authentication provides users access to their data while protecting account credentials. Open Authentication also allows users to grant access to their information, which is shared by the service provider and consumers without sharing all of their identity.
- ii. Open ID: Open ID is an open, decentralized standard for user authentication and access control that allows users to log onto many services using the same digital identity. It is a single-sign-on (SSO) method of access control. As such, it replaces the common log-in process (i.e., a log-in name and a password) by allowing users to log in once and gain access to resources across participating systems.

III. INTERSECTION BETWEEN CLOUD, SERVICE ORIENTED ARCHITECTURE AND GRID COMPUTING

Cloud Computing expands SOA by adding scalability and Grid Computing. Cloud computing is a means of computing 'services' that are provided over the internet and can scale dynamically and is based on virtualized resources. Cloud Service Providers includes IBM , Amazon , Google , Microsoft , Yahoo etc. These applications may need many computers. And here the computing resources are shared by multiple clients/consumers. Internally, they run a grid, but it is hidden from you.

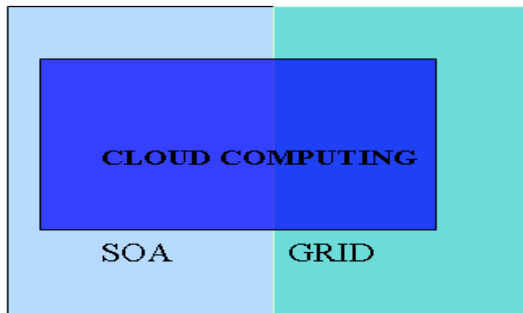


Figure 3: Intersection between SOA, Cloud and Grid.

SOA provides the evolutionary step to cloud computing by creating the necessary interfaces from the IT infrastructure to the cloud outside the enterprise. Cloud computing essentially becomes an extension of SOA.

SOA is derived from architecture and a methodology. Since cloud computing is typically driven from the view of business resources that are needed,

Another evolutionary step that cloud computing has taken from the SOA model is to architect and design services into the cloud so that it can expand and be accessed as needed. Expanding services in an SOA is typically a difficult and expensive process.

However, cloud computing is accelerating the adoption of SOA by providing aspects of SOA on-demand. SOA can learn a lot from the clouds, and the clouds can learn a lot from SOA."

Differentiating between SOA and cloud computing can be confusing because they overlap in some areas but are fundamentally different. SOA delivers web services from applications to other programs, whereas the cloud is about delivering software services to end users and running code. Thus the cloud-versus-SOA debate is like comparing apples and oranges. The true success of SOA application depends widely on its deployment in the cloud and taking advantage of its elasticity.

IV. CHALLENGES AND SECURITY ISSUES FOR CLOUD ENVIRONMENT

- i. Cloud applications undergo constant feature additions, and users must keep up to date with application improvements to be sure they are protected. The speed at which applications will change in the cloud will affect both the SDLC and security.
- ii. Data integrity means ensuring that data is identically maintained during any operation (such as transfer, storage, or retrieval). Put simply, data integrity is assurance that the data is consistent and correct. Ensuring the integrity of the data really means that it changes only in response to authorized transactions. This sounds good, but you must remember that a common standard to ensure data integrity does not yet exist.
- iii. Any information stored locally on a computer can be stored in a cloud, including email, word

processing documents, spreadsheets, videos, health records, photographs, tax or other financial information, business plans, PowerPoint presentations, accounting information, advertising campaigns, sales numbers, appointment calendars, address books, and more. The entire contents of a user's storage device may be stored with a single cloud provider or with many cloud providers. Whenever an individual, a business, a government agency, or other entity shares information in the cloud, privacy or confidentiality questions may arise.

- iv. Even if data is securely stored in a cloud, many factors can temporarily disrupt access to the data, such as network outages, denial of service attacks against the service provider, and a major failure of the service provider infrastructure. It may be a challenge to host and maintain intranet and access restricted sites (government, defense, institutional.)
- v. If information is encrypted while passing through the cloud, who controls the encryption/decryption keys? Is it the customer or the cloud vendor?
- vi. If your data is not available or is completely wiped out due to a disaster, can they replicate that data for you? How soon? How much of it?
- vii. The virtual machines must be self-defending, effectively moving the perimeter to the virtual machine itself. Enterprise perimeter security (i.e., firewalls, demilitarized zones, network segmentation, intrusion detection and prevention systems [IDS/IPS], monitoring tools, and the associated security policies) only controls the data that resides and transits behind the perimeter.

V. CONCLUSION

Expanding services in an SOA is typically a difficult and expensive process. So Cloud architecture attached to grid computing ensures that the SOA applications take advantage of the elasticity of cloud computing environment and grid computing able to process a service within a finite amount of time. Cloud computing has been criticized for limiting the freedom of users and making them dependent on the cloud computing provider, and some critics have alleged that it is only possible to use applications/services that the provider willing to offer. Virtualization related issues will be satisfactorily resolved and network intrusion detection and prevention system for cloud is appearing as this new technology matures.

VI. REFERENCES

- 1) Andy Bechtolsheim Chairman & Co-founder, Arista Networks- Cloud Computing, November 12th, 2008.
- 2) Daniel J. Abadi, Data Management in the Cloud: Limitations and Opportunities.
- 3) David Linthicum, SOA cloud computing relationship leaves some folks in a fog, Mar 09, 2009.
- 4) Ejaz Jamil, What really is SOA. A comparison with Cloud Computing, Web 2.0, SaaS, WOA, Web Services, PaaS and others.
- 5) Fred cummins, SOA and Cloud: Business Necessity for Small to Medium Sized Businesses, September 2009.
- 6) Joe McKendrick, Connecting the Dots Between SOA and Cloud, June 29, 2009.
- 7) John W. Rittinghouse, James F. Ransome: Cloud Computing Implementation, Management and Security.
- 8) Marcia Savage, Security challenges with cloud computing services, Information Security magazine 21 Sep 2009 | SearchSecurity.com .
- 9) Nikita Ivanov, 20 Real-Life Challenges of Cloud Computing.
- 10) Oracle White Paper in Enterprise Architecture – Architectural Strategies for Cloud Computing.
- 11) Paul Krill, The cloud-SOA connection, February 2009.
- 12) SOA Consortium, Finding the Intersections of SOA and Cloud Computing.
- 13) Sun Cloud Computing.
- 14) The Aspen Institute, Identity in the Age of Cloud Computing: The next-generation Internet's impact on business, governance and social interaction, 2009.
- 15) [http://www.dmtf.org/standards/published documents/DSP0243_1.1.0.pdf](http://www.dmtf.org/standards/published/documents/DSP0243_1.1.0.pdf)

Multi Domain Metadata Standard Harvesting

G. Sivaraman¹, K.Thangadurai²

{ GJCST Computing Classification
H 3.5,H 3.1,F m }

Abstract- The rapid growth of Internet resources, digital collections and libraries are constructed with the help of metadata schemas. Each metadata schema has been designed based on the requirements of the particular user community, intended users, type of materials, subject domain, the depth of description, etc. Problems arise when building large digital libraries or digital information resource with metadata records prepared according to related schemas. Most of the users do not know or understand the underlying structure of the digital collection; but in reality, they are experiencing difficulties in retrieval. The challenge will be overcome through metadata harvesting. This paper is reviewing this harvesting with example.

Keywords- metadata, metadata standard, harvesting, crosswalk, interoperability, harmonization.

I. INTRODUCTION

Information retrieval from heterogeneous resources is quite difficult. Because of the information holding follow different material administration and different metadata implementation techniques. There is two or more type of metadata standards are used in same subject domain or in same type of resource. In building a large digital library or digital collection, an issue often encountered is that the resource may have used different schemas and description methods to create their metadata records. Users want to retrieve information through one search what digital objects freely available from a variety of collections rather than searching each collection individually. User community can be developed to attain harvesting it will be possible to facilitate the exchange and sharing of data prepared according to different metadata schemas and to enable cross-collection searching. This article analyzes some of the methods currently used to achieve harvesting in a broader context, that is, among different metadata schemas and applications.

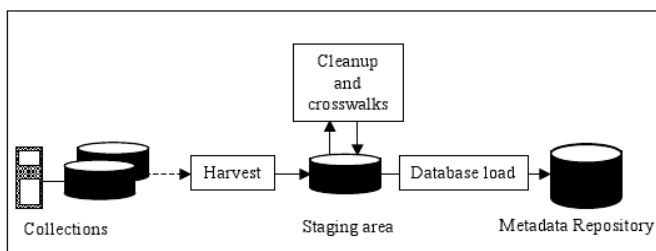


Fig.1. Collection of record store into metadata repository

About- ¹Department of Computer Science, M.G.R.College, Hosur-635109, Tamilnadu, India

About² Department of Computer Science, Government Arts College(Men), Krishnagiri- 635 001, Tamilnadu, India E-mail: sivaphd_guru@yahoo.co.in

II. HARVESTING

Harvesting refers to the gathering together of metadata from a number of distributed repositories into a combined data store. In other words, harvesting is a technique for extracting metadata from individual repositories and collecting it in a central catalog.

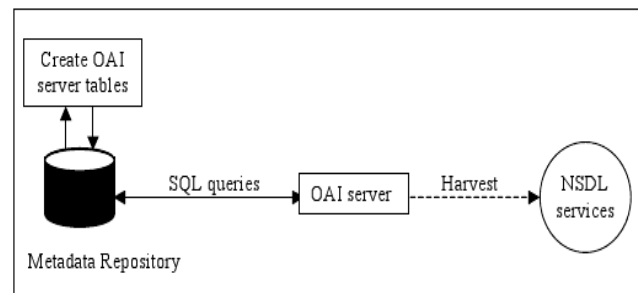


fig.2. Process of Harvesting

Resources across the network can be searched more flawlessly using defined metadata standards and shared transfer protocols between these standards. Different metadata standards are available like the Dublin Core, LOM etc. For the accurate retrieval of information using metadata, the different metadata standards should be able to operate between themselves. Hence the concept of metadata harvesting arose.

Harvesting achieve in three ways, Crosswalk, Interoperability and Harmonization.

III. CROSSWALK

A crosswalk is a specification for mapping one metadata standard to another. Crosswalks provide the ability to make the contents of elements defined in one metadata standard available to communities using related metadata standards.

A crosswalk is defined as a mapping of the elements, semantics, and syntax from one metadata scheme to those of another. The predominant method used is direct mapping or establishing equivalency between and among elements in different schemas. Equivalent fields or elements are mapped in order to allow conversion from one to the other. Most of the crosswalk effort to date has been in the form of mapping between two metadata schemas; mapping among multiple schemas has also been attempted.

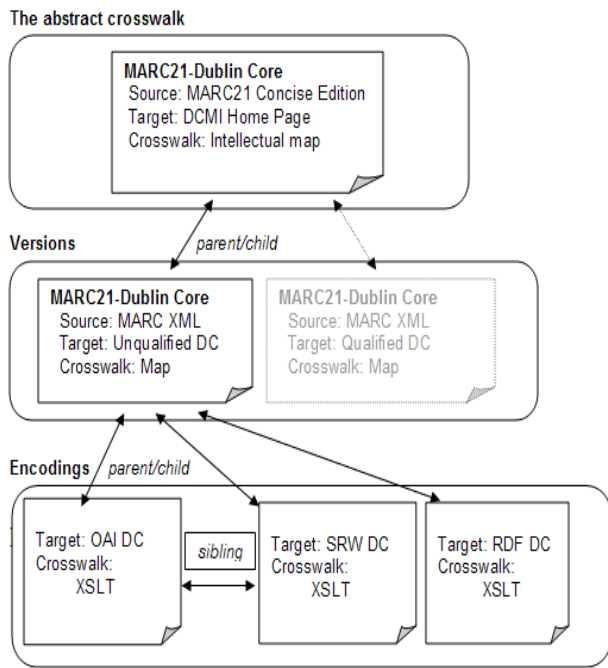


Fig. 3. Levels of Metadata Abstraction

There have been a substantial number of crosswalks. Some examples are:

- MARC21 to Dublin Core
- MARC to UNIMARC
- VRA to Dublin Core
- ONIX for books to MARCXML
- FGDC to MARC
- EAD to ISAD(G)
- ETD-MS to MARCXML
- Dublin Core/MARC/GILS
- ADL/FGDC/MARC/GILS
- MARC/LOM/DC
- Etc., etc., etc.

The crosswalk approach appears to be more workable when mapping from complex to simpler schema. An example is the crosswalk between the Dublin Core and MARC. Because of different degree of depth and complexity, crosswalk works relatively well when mapping MARC fields to Dublin Core elements but not vice versa, because MARC is a much more complex schema. One of the problems identified is the different degrees of equivalency: one-to-one, one-to-many, many-to-one, and one-to-none. Also, while crosswalk works well when the number of schemes involved is small, mapping among multiple schemas is not only extremely tedious and labor intensive but requires enormous intellectual efforts. For example, a one-way crosswalk requires one mapping process (A-->B), and a two-way crosswalk requires two mapping processes (A-->B and B-->A). When the process becomes more and more cumbersome the more schemas are involved. For example, a crosswalk involving three schemas would require six (or three pairs of) mapping processes ,a

four-schema crosswalk would require twelve (or six pairs of) mapping processes and a five-schema crosswalk would require twenty mapping processes.

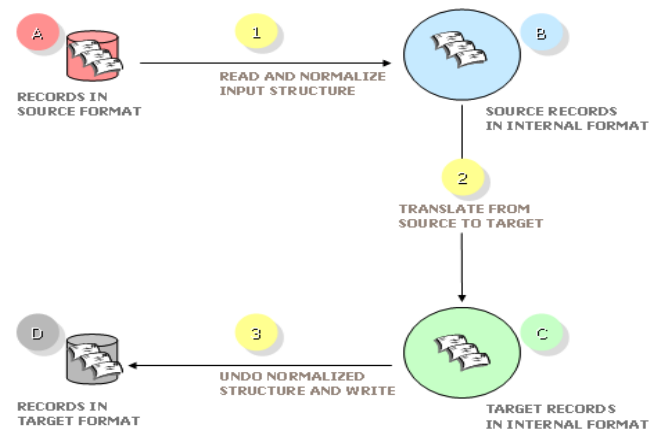
IV. INTEROPERABILITY

Interoperability means that the compatibility of two or more systems such that they can exchange information and data without any special manipulation.

In recent years, numerous projects have been undertaken in the information community to achieve interoperability among different metadata schemas. Some of these efforts are outlined below.

- Uniform standard
- Application profiling/adaptation/modification
- Derivation
- Switching schema
- Lingua franca

Fig. 4. Process of Interoperability



A. Uniform Standard

In this approach, all participants of a consortium, repository, etc., use the same schema, such as MARC/AACR or the Dublin Core. However, although it is a conceptually simple solution, it is not always feasible or practical, particularly in heterogeneous environments serving different user communities where components or participating collections contain different types of resources already described by a variety of specialized schemas. This method is only viable at the beginning or early stages of building a digital library or repository, before different schemas have been adopted by different participants of the collection or repository. Examples of uniform standardization include the MARC/AACR standards used in union catalogs of library collections and the Electronic Thesis and Dissertations Metadata Standard (ELD-MS) based on the Dublin Core used by members of the Networked Digital Library of Thesis and Dissertations (NDLTD).

B. Application Profiling/Adaptation/Modification

In the heterogeneous information environment, different communities manage information that has different characteristics and requirements. There often is no one metadata schema that meets all needs, that is, ~~one-size-~~does-not-fit-all. ~~To~~ accommodate individual needs, in this approach, an existing schema is used as the basis for description in a particular digital library or repository, while individual needs are met through specific guidelines or through adaptation or modification by:

- i. Creating an application profile (a set of policies) for application by a particular interest group or user community.
- ii. Adapting an existing schema with modification to cater to local or specific needs, that is, a DTD of an existing schema.

C. Derivation

In a collection of digital databases where different components have different needs and different requirements regarding depths, an existing complex schema such as the MARC format may be used as the ~~source~~ or ~~model~~ from which new and simpler individual schemas may be derived. This approach would ensure a similar basic structure and common elements, while allowing different components to vary in depth and details. For example, both the MODS (Metadata Object Description Schema) and MARC Lite are derived from the MARC21 standard, and the TEI Lite is derived from the full Text Encoding Initiative (TEI).

D. Switching Schema

In this model, an existing schema is used as the switching mechanism among multiple schemas. Instead of mapping between every pair in the group, each of the individual metadata schemas is mapped to the switching schema. This model reduces drastically according to the number of mapping processes required. The switching schema usually contains elements on a fairly broad level. Examples of using switching schemas include the Picture Australia project and the Open Archive Initiative (OAI). Both use the Dublin Core as the switching schema.

E. Lingua Franca

If no existing schema is found to be suitable for use as a switching schema, an alternative is the use of a lingua franca. A lingua franca acts as a superstructure, but is not a ~~schema~~ in itself. In this method, multiple existing metadata schemas are treated as satellites of a superstructure (lingua franca) which consists of elements common or most widely used by individual metadata schemas. This model facilitates cross-domain searching but is not necessarily helpful in data conversion or data exchange. However, the lingua franca model allows the retention of the richness and granularity of individual schemas.

The lingua franca superstructure is built from a set of core attributes that are common to many or most of the existing schemas used by participants in a digital library or

repository. An example is the ROADS template, which uses a set of broad, generic attributes.

V. HARMONIZATION

Harmonization is refers to the ability of different systems to exchange information about resources. Metadata created in one system and then transferred to a second system will be processed by that second system in ways which are consistent with the intentions of the metadata creators (human or software).

Different forms of Harmonization:

A. Extensibility

The ability to create structural additions to a metadata standard needs application-specific or community-specific. Given the diversity of resources and information, extensibility is a critical feature of metadata standards and formats.

B. Modularity

The ability to combine metadata fragments adhering to different standards. Modularity metadata extensions from different sources should be usable in combination without causing ambiguities or incompatibilities.

C. Refinements

The ability to create semantic extensions, i.e., more fine-grained descriptions that are compatible with more coarse-grained metadata, and to translate a fine-grained description into a more coarse-grained description.

D. Multilingualism

It has ability to express, process and display metadata in a number of different linguistic and cultural circumstances. One important aspect of this is the ability to distinguish between what needs to be human-readable and what needs to be machine-processable.

Harmonization then refers to the ability to use several different metadata standards in combination in a single software system. The rest of the deliverable will analyse the different groups of standards and try to find obstacles to harmonization.

VI. CONCLUSION

In the open, networked environment enable multiple user communities using a multitude of standards for description of digital resources, the need for harvesting among metadata schemas is over-riding. Currently, mapping metadata schemas still require enormous effort even with all the assistance computer technology can provide. If the information community is to provide optimal access to all the information available across the board of digital libraries and depositories, information professionals must give high priority to the task of creating-and maintaining-the highest feasible level of exchange methods among schemas and new information services.

VII. REFERENCES

- 1) ALCTS/CCS/Committee on Cataloging: Description and Access Task Force on Metadata. (last updated 2004). Summary Report. <http://www.libraries.psu.edu/tas/jca/ccda/tf-meta3.html>
- 2) Guenther, Rebecca, and Sally McCallum (2002). New metadata standards for digital resources: MODS and METS. ASIST Bulletin, 29(2)
- 3) Heery, Rachel M, Andy Powell, and Michael William Day. (Mar. 1998). Metadata: CrossROADS and interoperability [computer file]. Ariadne (Online) no. 14
- 4) Gillman, D., Appel, M., and Highsmith, S (1997). Building A Statistical Metadata Repository, Second IEEE Metadata Conference 1997: <http://computer.org/conferen/proceed/meta97/papers/dgillman/dgillman.html>
- 5) LaPlant, W. P. Jr., Lestina, G. J. Jr., Gillman, D. W., and Appel, M. V. (1996), "Proposal for a Statistical Metadata Standard", Census Annual Research Conference, Arlington, VA., March 18-21, 1996.
- 6) Consortium for the Computer Interchange of Museum Information (CIMI): http://www.cimi.org/downloads/CIMI_profile/profile.htm.
- 7) Baker, T., and Dekkers, M., (2002), CORES Standards Interoperability Forum Resolution on Metadata Element Identifiers. <http://www.cores-eu.net/interoperability/cores-resolution>

Maximum Likelihood Criterion Based On Polygon Inequality

{ *GJCST Computing Classification*
G.O, H 4.2 }

Hsin-Kun Lai

Abstract--This letter proposes a maximum likelihood criterion for soft-decision decoding algorithm of binary linear block codes over AWGN channel based on natural polygon inequality. In that criterion, Hamming weights are computed instead of Euclidean distances. As a result, decoding complexity is reduced in both generating test patterns and computing Euclidean distances.

Keywords- polygon inequality, soft-decision decoding, Hamming weight, Euclidean distance.

I. INTRODUCTION

Basically, error control coding with forward error-correction can be categorized into hard decision decoding (HDD) and soft decision decoding (SDD). HDD algorithms have the advantage of low decoding complexity but the disadvantage of low error-correction performance. On the contrary, SDD algorithms have the advantage of better error-correction performance by two to three dB coding gain with respect to that of HDD algorithms [1], but the disadvantage of high decoding complexity. For better error-correction performance, many studies proposed some SDD algorithms based on such as the threshold of E_b/N_0 [2], stop criterion [3], reliabilities [4], voting [5] or for a particular code [6]-[7] to reduce decoding complexity with slight or without degradation in error-correcting performance. To reduce decoding complexity, this letter proposes a soft-decision maximum likelihood criterion for SDD algorithm by computing Hamming weights (HWs) rather than Euclidean distances (EDs). If the criterion is satisfied, the code word is the maximum likelihood (ML) one. Hence, generating test patterns (TPs) is unnecessary. Certainly, computing EDs is not required. Thus, decoding complexity is reduced. The higher the SNR is, the more chance the code word satisfies the criterion. In the next section, The Chase algorithm [8] is introduced first and the decoding algorithm with the proposed criterion is presented. In addition, the criterion is defined and proved as well. Following that, simulation results show that the decoding complexity with the help of proposed criterion is reduced. In turn, the strength and the weakness of decoding algorithm with the proposed criterion are summarized in the conclusion.

*H.-K. Lai is with the department of electronic engineering, Chin Min Institute of technology, No.110, Syuefu Rd., Toufen Township, Miaoli County 351, Taiwan, R.O.C.
(telephone: +886-37-605535
email: larry@ms.chinmin.edu.tw)*

II. DECODING ALGORITHM WITH THE PROPOSED CRITERION

Assume that $C(n,k,d_{min})$ denotes a code with code length n , message length k , minimum Hamming distances d_{min} between code words, and number of error-correcting

capability $t = \lfloor \frac{d_{min}-1}{2} \rfloor$. And further, algorithm with the proposed criterion is termed as natural polygon inequality (NPI) based SDD algorithm. Now, consider a code word $C=[c_1 c_2 \dots c_n]$ is transmitted using BPSK modulation, where c_i denoted the i th bit of C . And suppose that the received signal at the output of demodulator is given as $Y=[y_1 y_2 \dots y_n]$. Let $A=[\alpha_1 \alpha_2 \dots \alpha_n]$, which is absolute value $|Y_i|$, be the reliability word of Y . In addition, R is the word received from the output of hard decision (HD) with Y as the input; that is $R=HD(Y)$, where HD given as (1).

$$HD(y_i) = \begin{cases} 0, & y_i \leq 0 \\ 1, & y_i > 0 \end{cases} \quad (1)$$

With availability of R and A , the decoding steps of Chase algorithm are as follows:

Step 1. Do HDD to R

$$C_R = HDD(R) \quad (2)$$

Step 2. Do C_R exclusive-OR operation with R as in (3) and then error pattern (EP_R) is found.

$$EP_R = C_R \oplus R \quad (3)$$

If there is no error, the decoding process finishes with C_R . If not, continue the following steps.

Step 3. Sort the reliabilities $A=[\alpha_1 \alpha_2 \dots \alpha_n]$ in ascending order and which gives a new reliability vector $A_S=[\alpha_{S1} \alpha_{S2} \dots \alpha_{Sn}]$, where $\alpha_{Si} < \alpha_{Sj}$ for $i < j$. Also, a position vector $P_S=[p_{S1} p_{S2} \dots p_{Sn}]$ keeps the original position of $[\alpha_{S1} \alpha_{S2} \dots \alpha_{Sn}]$.

Step 4. By use of P_S , Chase algorithm generates a number of TP according to the positions of bits with least reliabilities in R .

Step 5. Do R exclusive-or operations with TPs as given in (4),

$$Z_S = R \oplus TPs \quad (4)$$

Step 6. Do HDD to words Z_S as given in (5).

$$CCs=HDD(Zs) \tag{5}$$

Step 7. Find out the code word in the code word candidates (CCs) for which has the minimum ED from R.

The decoding process of NPI algorithm is the same as that of Chase except with one more step after step3 by testing (6). Utilizing the position word P_S and programming skill of indirect addressing, C_R is a ML code word if (6) satisfied.

$$\sum_{i=1}^{d_{\min}-W\{EP_{R_S}\}} (EP_{R_S}) = W\{EP_{R_S}\} \tag{6}$$

where

EP_{R_S} : sorted EP_R in ascending order according to P_S.

$W\{EP_{R_S}\}$:Hamming weights of EP_{R_S} .

Equation (6) implies that the error bits are included in the least reliable bits. If (6) holds for the associated C_R, decoding is completed. Hence, generating TPs and computing EDs are unnecessary from step4 to step7 for the NPI algorithm compared to that of Chase algorithm. If EP_{R_S} does not satisfy (6), NPI algorithm follows from step4 to the last step of Chase algorithm to check if C_R is a ML code word.

Before proving that (6) determines the associated C_R a ML code word, a theorem of natural polygon inequality is introduced based on the theorem of triangle inequality.

Theorem- Natural polygon inequality (NPI)

Suppose that a polygon is formed with M edges, where the length of edges $E_1 \leq E_2 \leq \dots \leq E_M$ and $E_M < E_1 + E_2 + \dots + E_{M-1}$. Then, the summation of edges' length has following natural inequality (7) and which is termed as the NPI.

$$\sum_{i=1}^{\lfloor \frac{M}{2} \rfloor} E_i \leq \sum_{i=\lfloor \frac{M}{2} \rfloor + 1}^M E_i \tag{7}$$

Now, we prove that (6) determines the associated C_R a ML code word.

Proof- Since $A_S = [\alpha_{S1} \alpha_{S2} \alpha_{S3} \dots \alpha_{Sn}]$ is the sorted reliability and if (6) is satisfied, the maximum ED_{C_R} of code word C_R from the received word R is

$$ED_{C_R} \leq \sum_{i=1}^{d_{\min}-W\{EP_{R_S}\}} \alpha_{S_i}$$

Moreover, any other code word C_X with minimum ED_{C_x} from the received word R is

$$ED_{C_x} \geq \sum_{i=d_{\min}-W\{EP_{R_S}\}+1}^{d_{\min}} \alpha_{S_i}$$

Referring to (7), we observed that

$$ED_{C_R} \leq ED_{C_x}$$

Hence, C_R is a ML code word. The following example helps to clarify the proof above.

Example: Assume that Golay (23,12,7) is used and C_R satisfies (6). Then referring to (6), the maximum value of ED_{C_R} is

$$ED_{C_R} \leq \alpha_{S_4} + \alpha_{S_5}, \quad \text{in case of } W\{EP_{R_S}\} = 2.$$

And any other code word C_X with the minimum value of ED_{C_x} is

$$ED_{C_x} \geq \alpha_{S_1} + \alpha_{S_2} + \alpha_{S_3} + \alpha_{S_6} + \alpha_{S_7}$$

when $W\{EP_{R_S}\} = 2$

Since $A_S = [\alpha_{S1} \alpha_{S2} \alpha_{S3} \dots \alpha_{Sn}]$ is the sorted reliability in ascending order, the reliabilities inequality $\alpha_{S_4} + \alpha_{S_5} \leq \alpha_{S_6} + \alpha_{S_7} + \alpha_{S_3} + \alpha_{S_2} + \alpha_{S_1}$ naturally holds.

Intuitively, ED_{C_R} is less than ED_{C_x} . C_R is a ML code word.

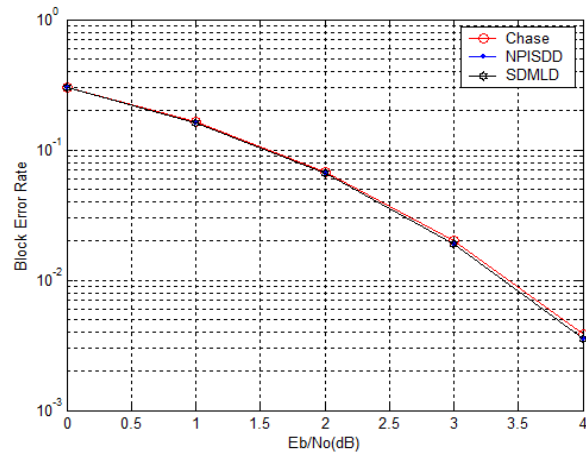


Fig.1. Error-correcting performances

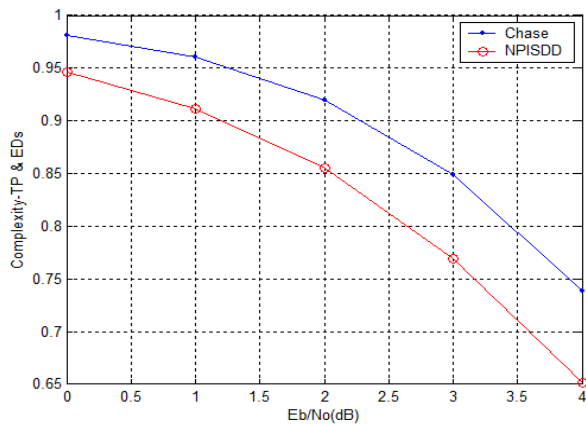


Fig.2. Comparison of decoding complexity in terms of TPs and EDs

III. SIMULATION RESULTS

As shown in Fig. 1, the proposed NPI algorithm and the Chase algorithm has almost the same error-correcting performance as the soft-decision ML decoding. In addition, the decoding complexity in generating TPs and computing EDs is reduced by around from 4% to 9% for the proposed NPI algorithm from 0 dB to 4 dB of Eb/No compared to that of Chase algorithm. In other words, around 4% to 9% in the decoding process is computing HWs in stead of EDs.

IV. CONCLUSION

We summarize the advantage and disadvantage of NPI algorithm by making a comparison with those of Chase algorithm. Obviously, the NPI algorithm has the advantage of less decoding complexity in generating TPs and computing EDs by around 9% at 4 dB of Eb/No as shown in Fig 2. However, this is a trade-off of one more step of testing with (6). Nevertheless, (6) is a soft-decision maximum likelihood criterion without computing EDs and which is the innovative point in this letter.

V. REFERENCES

- 1) Wozencraft, J.M. & Jacobs, I.M. (1965). Principles of Communication Engineering. Illinois: John Wiley & Sons.
- 2) Tanaka, H. & Kakigahara, K. (1983). Simplified correlation decoding by selecting possible code words using erasure information. IEEE Trans. Info. Theory, 29, 743-748.
- 3) Kaneko, T., Nishijima, T., Inazumi, H., & Hirasawa, S. (1994). An efficient maximum likelihood decoding algorithm for linear block codes with algebraic decoder. IEEE Trans. Info. Theory, 40, 320-327.
- 4) Mahran, A., & Benaissa, M.(2003). Adaptive Chase algorithm for block turbo codes. Electron. Lett.,39, 617-619.
- 5) [5] Kuo, C.F., Lai, H.K., & Lu, E.H. (2007). Half-Chase algorithm of non-soft decision decoding based on majority, Proc. ICACT' 07, 619-621.
- 6) [6] Lu, E.H., Wu, H.P., Cheng, Y.C., & Lu, P.C. (1995). Fast algorithms for decoding the (23,12) binary Golay code with four-error-correcting capability. International Journal of Systems Science, 26, 937-945.
- 7) [7] Hackett, C.M. (1981). An efficient algorithm for soft –decision decoding of the (24,12) extended Golay code, IEEE Trans. Communi., com-29, 909-911.
- 8) [8] Chase, D. (1972). A class of algorithms for decoding block codes with channel measurement information, IEEE Trans. Info. Theory, it-18, 170-182.

Global Journals Guidelines Handbook 2010

www.GlobalJournals.org

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

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

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

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



-

Auxiliary Memberships

ANNUAL MEMBER

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

PAPER PUBLICATION

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



Process of submission of Research Paper

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

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

The Author can submit the paper either online or offline. The authors should prefer online submission.

Online Submission: There are three ways to submit your paper:

(A) (I) 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.

© Copyright by Global Journals | Guidelines Handbook

Scope

The Global Journals welcome the submission of original paper, review paper, survey article relevant to the all the streams of Philosophy and knowledge. The Global Journals is parental platform for Global Journal of Computer Science and Technology, Researches in Engineering, Medical Research, Science Frontier Research, Human Social Science, Management, and Business organization. The choice of specific field can be done otherwise as following in Abstracting and Indexing Page on this Website. As the all Global Journals are being abstracted and indexed (in process) by most of the reputed organizations. Topics of only narrow interest will not be accepted unless they have wider potential or consequences.

2. ETHICAL GUIDELINES

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

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

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

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

- 1) Substantial contributions to conception and acquisition of data, analysis and interpretation of the findings.
- 2) Drafting the paper and revising it critically regarding important academic content.
- 3) Final approval of the version of the paper to be published.

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

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

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

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

Please mention proper reference and appropriate acknowledgements wherever expected.

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

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

3. SUBMISSION OF MANUSCRIPTS

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



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

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

4. MANUSCRIPT'S CATEGORY

Based on potential and nature, the manuscript can be categorized under the following heads: Original research paper: Such papers are reports of high-level significant original research work.

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

Research articles: These are handled with small investigation and applications

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

5. STRUCTURE AND FORMAT OF MANUSCRIPT

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

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

- (a) Title should be relevant and commensurate with the theme of the paper.
- (b) A brief Summary, "Abstract" (less than 150 words) containing the major results and conclusions.
- (c) Up to ten keywords, that precisely identifies the paper's subject, purpose, and focus.
- (d) An Introduction, giving necessary background excluding subheadings; objectives must be clearly declared.
- (e) Resources and techniques with sufficient complete experimental details (wherever possible by reference) to permit repetition; sources of information must be given and numerical methods must be specified by reference, unless non-standard.
- (f) Results should be presented concisely, by well-designed tables and/or figures; the same data may not be used in both; suitable statistical data should be given. All data must be obtained with attention to numerical detail in the planning stage. As reproduced design has been recognized to be important to experiments for a considerable time, the Editor has decided that any paper that appears not to have adequate numerical treatments of the data will be returned un-refereed;
- (g) Discussion should cover the implications and consequences, not just recapitulating the results; conclusions should be summarizing.
- (h) Brief Acknowledgements.
- (i) References in the proper form.



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

The Editorial Board reserves the right to make literary corrections and to make suggestions to improve 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, ought to include:

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

Abstract, used in Original Papers and Reviews:

Optimizing Abstract for Search Engines

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

Key Words

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

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



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

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

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

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

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

Acknowledgements: Please make these as concise as possible.

References

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

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

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

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

Tables, Figures and Figure Legends

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

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

Preparation of Electronic Figures for Publication

Even though low quality images are sufficient for review purposes, print publication requires high quality images to prevent the final product being blurred or fuzzy. Submit (or e-mail) EPS (line art) or TIFF (halftone/photographs) files only. MS PowerPoint and Word Graphics are unsuitable for printed pictures. Do not use pixel-oriented software. Scans (TIFF only) should have a resolution of at least 350



dpi (halftone) or 700 to 1100 dpi (line drawings) in relation to the imitation size. Please give the data for figures in black and white or submit a Color Work Agreement Form. EPS files must be saved with fonts embedded (and with a TIFF preview, if possible).

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

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

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

6. AFTER ACCEPTANCE

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

6.1 Proof Corrections

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

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

(Free of charge) from the following website:

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

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

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

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

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

6.3 Author Services

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

6.4 Author Material Archive Policy

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



6.5 Offprint and Extra Copies

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

INFORMAL TIPS FOR WRITING A COMPUTER SCIENCE RESEARCH PAPER TO INCREASE READABILITY AND CITATION

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

Techniques for writing a good quality Computer Science Research Paper:

1. Choosing the topic- In most cases, the topic is searched by the interest of author but it can be also suggested by the guides. You can have several topics and then you can judge that in which topic or subject you are finding yourself most comfortable. This can be done by asking several questions to yourself, like Will I be able to carry our search in this area? Will I find all necessary recourses to accomplish the search? Will I be able to find all information in this field area? If the answer of these types of questions will be "Yes" then you can choose that topic. In most of the cases, you may have to conduct the surveys and have to visit several places because this field is related to Computer Science and Information Technology. Also, you may have to do a lot of work to find all rise and falls regarding the various data of that subject. Sometimes, detailed information plays a vital role, instead of short information.

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

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

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

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

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

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

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

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



- 10. Bookmarks are useful:** When you read any book or magazine, you generally use bookmarks, right! It is a good habit, which helps to not to lose your continuity. You should always use bookmarks while searching on Internet also, which will make your search easier.
- 11. Revise what you wrote:** When you write anything, always read it, summarize it and then finalize it.
- 12. Make all efforts:** Make all efforts to mention what you are going to write in your paper. That means always have a good start. Try to mention everything in introduction, that what is the need of a particular research paper. Polish your work by good skill of writing and always give an evaluator, what he wants.
- 13. Have backups:** When you are going to do any important thing like making research paper, you should always have backup copies of it either in your computer or in paper. This will help you to not to lose any of your important.
- 14. Produce good diagrams of your own:** Always try to include good charts or diagrams in your paper to improve quality. Using several and unnecessary diagrams will degrade the quality of your paper by creating "hotchpotch." So always, try to make and include those diagrams, which are made by your own to improve readability and understandability of your paper.
- 15. Use of direct quotes:** When you do research relevant to literature, history or current affairs then use of quotes become essential but if study is relevant to science then use of quotes is not preferable.
- 16. Use proper verb tense:** Use proper verb tenses in your paper. Use past tense, to present those events that happened. Use present tense to indicate events that are going on. Use future tense to indicate future happening events. Use of improper and wrong tenses will confuse the evaluator. Avoid the sentences that are incomplete.
- 17. Never use online paper:** If you are getting any paper on Internet, then never use it as your research paper because it might be possible that evaluator has already seen it or maybe it is outdated version.
- 18. Pick a good study spot:** To do your research studies always try to pick a spot, which is quiet. Every spot is not for studies. Spot that suits you choose it and proceed further.
- 19. Know what you know:** Always try to know, what you know by making objectives. Else, you will be confused and cannot achieve your target.
- 20. Use good quality grammar:** Always use a good quality grammar and use words that will throw positive impact on evaluator. Use of good quality grammar does not mean to use tough words, that for each word the evaluator has to go through dictionary. Do not start sentence with a conjunction. Do not fragment sentences. Eliminate one-word sentences. Ignore passive voice. Do not ever use a big word when a diminutive one would suffice. Verbs have to be in agreement with their subjects. Prepositions are not expressions to finish sentences with. It is incorrect to ever divide an infinitive. Avoid clichés like the disease. Also, always shun irritating alliteration. Use language that is simple and straight forward. put together a neat summary.
- 21. Arrangement of information:** Each section of the main body should start with an opening sentence and there should be a changeover at the end of the section. Give only valid and powerful arguments to your topic. You may also maintain your arguments with records.
- 22. Never start in last minute:** Always start at right time and give enough time to research work. Leaving everything to the last minute will degrade your paper and spoil your work.
- 23. Multitasking in research is not good:** Doing several things at the same time proves bad habit in case of research activity. Research is an area, where everything has a particular time slot. Divide your research work in parts and do particular part in particular time slot.
- 24. Never copy others' work:** Never copy others' work and give it your name because if evaluator has seen it anywhere you will be in trouble.



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

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

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

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

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

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

31. Adding unnecessary information: Do not add unnecessary information, like, I have used MS Excel to draw graph. Do not add irrelevant and inappropriate material. These all will create superfluous. Foreign terminology and phrases are not apropos. One should NEVER take a broad view. Analogy in script is like feathers on a snake. Not at all use a large word when a very small one would be sufficient. Use words properly, regardless of how others use them. Remove quotations. Puns are for kids, not grunt readers. Amplification is a billion times of inferior quality than sarcasm.

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

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

34. After conclusion: Once you have concluded your research, the next most important step is to present your findings. Presentation is extremely important as it is the definite medium 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

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

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

- The **major constraint** is that you must independently make all content, tables, graphs, and facts that are offered in the paper. You must write each part of the paper wholly on your own. The Peer-reviewers need to identify your own perceptive of the concepts in your own terms. NEVER extract straight from any foundation, and never rephrase someone else's analysis.
- Do not give permission to anyone else to "PROOFREAD" your manuscript.
- **Methods to avoid Plagiarism is applied by us on every paper, if found guilty, you will be blacklisted by all of our collaborated research groups, your institution will be informed for this and strict legal actions will be taken immediately.)**
- To guard yourself and others from possible illegal use please do not permit anyone right to use to your paper and files.



CRITERION FOR GRADING A RESEARCH PAPER (COMPILATION)
BY GLOBAL JOURNALS

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

Topics	Grades		
	A-B	C-D	E-F
<i>Abstract</i>	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
<i>Introduction</i>	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
<i>Methods and Procedures</i>	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
<i>Result</i>	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
<i>Discussion</i>	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
<i>References</i>	Complete and correct format, well organized	Beside the point, Incomplete	Wrong format and structuring



Index

A

accelerates · 3
Access · 8, 12, 30, 40, 48, 49, 50, 51, 52, 53, 54, 72, 75, 76, 89
accounting · 30, 48, 50, 51, 52, 53, 84
acquiring · 24
activity · 12, 33, 55, 67, 78, 79, V, XI
adapted · 25, 27, 28, 30
adjacent · 3, 4, 5, 33, 62, 63
advantage · 11, 14, 17, 24, 46, 84, 90, 92, III
agency · 84
algorithm · 2, 3, 4, 5, 6, 7, 13, 14, 19, 20, 21, 22, 23, 51, 57, 73, 76, 90, 91, 92
AOP · 28, 29, 30, 31, 33, 35
application · 5, 6, 7, 8, 10, 11, 19, 26, 29, 30, 31, 32, 48, 53, 61, 63, 66, 67, 68, 73, 74, 75, 79, 82, 83, 84, 88
applications · VI
Architecture · 2, 12, 20, 66, 68, 71, 81, 85
Aspect · 2, 28, 29, 30, 31, 33, 34, 35
attributes · 25, 29, 30, 31, 48, 49, 50, 51, 52, 53, 57, 78, 88
authenticate · 13, 48, 50, 51, 53, 75
authentication · 10, 11, 12, 13, 14, 15, 16, 17, 18, 48, 49, 50, 52, 53, 61, 72, 75, 83
Authentication · 8, 10, 12, 13, 48, 50, 52, 53, 54, 75, 82
Authenticator · 49, 50, 51, 52, 53, 54
authorization · 11, 48, 50, 83
automation · 7, 14

B

biometrics · 10, 13, 14, 15, 17, 18
Biometrics · 2, 13, 17, 18
block · 9, 11, 16, 61, 62, 63, 64, 75, 90, 92

C

choose · IX, X, XV
Classifier · 55
Cloud · 2, 81, 82, 83, 84, 85
Common · VII
commonalities · 66, 67
complement · 55, 56, 59, XVI

complexity · 2, 6, 25, 27, 30, 31, 57, 62, 63, 64, 66, 68, 70, 71, 78, 79, 87, 90, 91, 92
composing · 6
Composition · 55, 59
computing · 2, 10, 30, 72, 76, 77, 81, 82, 83, 84, 85, 90, 91, 92
concatenation · 50, 51
concrete · 24, 29, 33, 48, 66
Concurrency · 8, 9
concurrent · 9, 52
configuration · 9, 12, 37, 48, 50, 52, 53, 66, 67, 68, 69, 70, 71
Configuration · 2, 12, 39, 62, 66, 67, 68, 69, 71
Conflicts · 2, 66
CoordinateObserver · 33
Copy · 2, 61, 62, 63, 64, 65
corresponding · 17, 26, 29, 52, 62, 63, 64, 78, 79, III, IX
could · XVI
coupling · 29, 30, 31, 32, 33
crosscutting · 28, 29, 30, 31, 32, 33
crosswalk · 86, 87
cryptographic · 13, 14, 15, 17, 51, 52, 54
Cryptography · 13, 14, 18, 54

D

Darzynkiewicza · 59
database · 3, 4, 6, 8, 9, 10, 11, 12, 13, 14, 15, 37, 39, 48, 50, 56, 75, 79, 82
davidgornshtein · 12
DBMS · 8, 10, 11, 12
decision · V, XVI, XVII, XVIII
decoding · 51, 90, 91, 92
Delay · 19, 20, 23, 73, 76
demonstrate · 10, XVI
design · 2, 13, 15, 17, 24, 28, 29, 31, 32, 33, 34, 37, 39, 52, 63, 67, 68, 78, 79, 80, 84, V, VI, XII, XVII
detachment · 74
detection · 16, 52, 61, 62, 63, 64, 68, 71, 75, 82, 84
determination · 2, 17
developer · 29, 31, 67, 78, 79, 80, 82
development · 11, 13, 24, 25, 26, 28, 29, 31, 33, 34, 57, 58, 61, 64, 67, 78, 79, 80, XIV
Digital · 2, 61, 62, 64, 65, 87
dimensions · 57, 62
Dipeptide · 55, 57, 59



dipeptides · 57
discrete · 2, 3, 4, 6, 63, 81
distribution · 38, 40, 57, 82, V

E

enabling · 11, 36, 38
encryption · 11, 14, 51, 52, 72, 84
engineers · 1, 33, 63, 68, 82
Enhancing · 17, 77
Enterprises · 27, 71
Evaluation · 6, 17, 21, 34, 39, 40, 46, 47, 57
exonuclease · 55
exoribonucleases · 55, 56, 58
explanation · 81, XVIII
explanations · 14, XVII
extension · 11, 24, 25, 26, 29, 48, 57, 84
Extraction · 2, 10, 13, 15

F

Feature · 8, 13, 14, 15, 18, 62, 66, 67, 69, 71
fingerprint · 13, 14, 17, 18
forensics · 61, 65
forgeries · 61, 63, 64, 65
forgery · 61, 62, 63, 64, 65
functionality · 8, 9, 11, 12, 40, 48, 53, 79
fuzziness · 24, 25, 26
fuzzy · 13, 14, 15, 16, 17, 24, 25, 26, 27, VIII
Fuzzy · 2, 13, 17, 18, 24, 25, 26, 27

G

gathering · I
guarantees · 19, 75

H

handled · 50, 54, 67, VI
harmonization · 86, 88
harvesting · 86, 88
hoaxes · 61
hydrolytic · 56

I

laaS · 81
illumination · 64
Image · 18, 61, 62, 64, 65

Implementation · 2, 18, 26, 34, 48, 52, 79, 80, 85
Implications · 34
inbuilt · 57, 58
information · 8, 9, 10, 11, 12, 13, 14, 19, 24, 25, 29, 36, 37, 38, 39, 40, 48, 49, 50, 51, 52, 53, 57, 58, 60, 64, 68, 69, 72, 73, 74, 75, 76, 79, 83, 84, 86, 87, 88, 92, III, V, VI, VII, VIII, IX, X, XIV, XV, XVI, XVII, XVIII
Information · 6, X
infrastructure · 68, 72, 74, 75, 81, 82, 84
initiated · 52, 53, 83
instinctively · 2
interactive · 7, 37, 40, 46, 83
interconnected · 10
interoperability · 52, 82, 86, 87, 89
Interoperability · 86, 87, 89
intersection · 2, 3, 4, 6, 31, 73, 75, 77, 81

J

jeopardize · 19
jitter · 19, 20, 21, 23
Jitter · 19, 20, 21, 23

L

Layer · 68, 69, 81
LDAP · 8, 10, 11, 12
likelihood · 31, 90, 92
localization · 60, 75, 83

M

Machine · 18, 55, 57, 60, 65
Management · 6, 7, 2, 8, 19, 20, 23, 24, 27, 39, 40, 48, 67, 71, 80, 81, 85, V
mapping · 11, 25, 33, 63, 86, 87, 88
matching · 14, 17, 61, 62, 63
matrix · 3, 4, 5, 6, 62, 63
mentioned · 13, 15, 16, 17, 19, 24, 29, 30, 82, III, IV, V
metadata · 83, 86, 87, 88, 89
methodology · 14, 15, 24, 25, 28, 56, 78, 80, 84, XV
metrics · 28, 29, 30, 31, 33
modularize · 28
Move · 2, 32, 61, 62, 63, 64, 65
Mulling · 73
multimedia · 19, 23, 73, 75, 76, 83
Multimedia · 23, 65

N

Network · 2, 11, 18, 20, 23, 48, 53, 54, 72, 75
Networking · 23, 76, 77

O

Object · 8, 11, 27, 34, 35, 78, 79, 80, 83, 88
obviously · 4, 19, XVI
occurrence · 61
organizations · V
Oriented · 2, 27, 28, 29, 34, 35, 78, 80, 81
Orthogonal · 7, 72

P

PaaS · 81, 85
packet · 11, 19, 20, 21, 48, 49, 50, 51, 52, 53, 72, 73, 74, 76
Parametric · 2, 78, 79, 80
participating · 83, 87
persistent · VII
Polygons · 2
potential · 26, 28, 29, 31, 37, 55, 59, 64, 73, 74, V, VI
Precision · 58, 64
procedure · V, VI, XV
Process · 2, III
Programming · 2, 28, 29, 34, 35, 39
Protocol · 2, 19, 23, 48, 51, 52, 53, 54, 77, 83
pseudo · 4, 5, 52

Q

quantization · 63
questionnaire · 39

R

Radius · 48
RADIUS · 2, 10, 48, 49, 50, 51, 52, 53, 54
randomly · 13, 15
reconstruction · 16
record · XIII
Repository · 2, 56, 66, 68, 69, 89
Requirements · 67, 68, 70, 80
Retina · 2, 13
Retinal · 13, 15
retransmission · 19, 52
ribonuclease · 55, 57
ribonucleases · 55, 56, 57, 58, 59

Ribonucleases · 2, 55, 56, 58, 59
robust · 6, 15, 63, 64, 65, 73, 74, 75
robustness · 26, 54, 63

S

Scheme · 13, 17, 18
SDLC · 78, 79, 84
Search · VII
security · 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 48, 51, 52, 53,
54, 72, 75, 76, 81, 83, 84
segments · 2, 3, 4, 5, 6, 7, 51, 61
Server · 2, 8, 9, 10, 11, 12, 28, 48, 49, 53, 59
SIGCSE · 7
significant research · VI
simulator · 19, 36
SOA · 81, 83, 84, 85
Soft · 2, 13, 18, 25, 27
Splicing · 61
standard · 11, 14, 25, 29, 33, 37, 48, 54, 57, 59, 72, 82, 83, 84,
86, 87, 88, IV, VI, XIII, XV
summary · 8, 39, 49, 63, XI, XIV, XVIII
Support · 5, 8, 10, 12, 54, 55, 57, 60, 77, 79, 80
Symposium · 7, 17, 18, 34, 77, I
synchronized · 53, 68, XV
system · 2, 3, 4, 8, 10, 11, 12, 13, 14, 15, 17, 19, 20, 24, 25, 26,
28, 29, 30, 31, 33, 34, 36, 37, 38, 39, 40, 41, 46, 47, 48, 52,
53, 55, 62, 64, 67, 69, 71, 73, 75, 77, 78, 79, 80, 81, 83, 84,
88, IX, XV

T

TCP · 2, 19, 21, 23, 48
technique · XV
Therefore · IX
transformation · 10, 14, 15, 16, 17
translation · 15, 16
Transmission · 2, 19, 23, 73

U

Ubiquitous · 76, 77
UDP · 2, 19, 21, 23, 48, 50, 52
university · 36, 37, 38
unsurprisingly · 19
user · 2, 3, 4, 9, 10, 11, 12, 13, 14, 15, 16, 17, 24, 25, 26, 27, 37,
39, 40, 46, 48, 49, 50, 51, 52, 53, 57, 59, 76, 78, 79, 81, 83,
84, 86, 87, 88
Utilizing · 91



V

Vault · 2, 13, 17, 18

Vector · 55, 57, 60

vehicular · 72, 74, 76, 77

Venkateshmurthy · 65

virtualized · 81, 82, 83

vulnerability · 12

W

warehouse · 26, 82

Wittenstein · 41





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