Global Journal of Computer Science and Technology

discovering thoughts and inventing future

Volume 10 Issue 9 Version 1.0 Online ISSN: 0975-4172 Print ISSN: 0975-4350

18 Technology Reforming Ideas

highlights

- Tam With Odour Interface
- Self-Organization Theory
- Service Productivity Optimization
- DWDM Network Design

September 2010

ENG

Global Journal of Computer Science and Technology

Global Journal of Computer Science and Technology

Volume 10 Issue 9 (Ver. 1.0)

Global Academy of Research and Development

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

Global Journals Inc.

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

Sponsors: Global Association of Research
Open Scientific Standards

Publisher's Headquarters office

Global Journals Inc., Headquarters Corporate Office, Cambridge Office Center, II Canal Park, Floor No. 5th, ***Cambridge (Massachusetts)***, Pin: MA 02141 United States
USA Toll Free: +001-888-839-7392
USA Toll Free Fax: +001-888-839-7392

Offset Typesetting

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

Packaging & Continental Dispatching

Global Journals, India

Find a correspondence nodal officer near you

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

eContacts

Press Inquiries: *press@globaljournals.org*
Investor Inquiries: *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

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

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

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

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

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

Dr. Söhnke M. Bartram
Department of Accounting and FinanceLancaster University Management 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 SciencesPurdue 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

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

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

Contents of the Volume

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

Validating Tam With Odour Interface In Atm Machines

GJCST Computing Classification
D.4.6, K.6.5, J.3

[1]Rashed Abdullah ,[2]Santos Henrique

***Abstract*- Many studies were and still are carried out to discover the user acceptance technology acceptance that mostly depends on ease of use and usefulness. Biometrics in general has the characteristics needed for acceptance such as easiness and usefulness. In this paper we introduce a new biometric interface for ATM machines, using human odour biometric for authentication. We validate TAM in the Portuguese culture in term of new interface acceptance (using odour in ATM machine). Results show that perceived easiness has significant effect with intention to use. However, perceived usefulness has not significant effect with intention to use. In addition results show that security issues should be embedded within TAM.**

Keywords-Biometrics, biometrics technology,
Authentication, odour, smell sense, Banking, Recognition, interface.

I. Introduction

Human–computer interaction tools are often used cross-culturally before being tested for suitability and validity. As new tools emerge, they must be cross-culturally validated to ensure that they work with all audiences (Oshlyansky et al., 2008), not just those in the country in which they were developed.

Day by day, the number of online banking customers is continuously growing. This demands more research, focusing on security issues relevant to the clients' side of online banking systems. Security is vital in banking and financial sectors, where users use their identity to transfer property and security is controlled by owning a particular artifact or tool (Coventry et al., 2003) and (Rashed & Santos, 2010b).

IT tools made the daily life much easier, but not secure (Sukhai, 2008). IT tools evolving everyday and the information overloading makes it difficult for individuals to recall their user names and passwords within all those technological artifacts. Moreover, users select easier passwords that they can easily remember (Coventry et al., 2003), which is considered a security tradeoff. Biometric techniques can be used as a solution for these problems: users can confirm their personal identities without being asked for their tokens or PINs and without a requirement to remember anything (Coventry et al., 2003). There are three types of authentication (Bala, 2008): 1 something that you know, like a PIN; 2) something you know, like a PIN; 2) something you have, like a passport, A driver‘s license, an ID card, or an ATM card; 3) something that you are (Biometrics), like fingerprints, signature, ear shape, odour, keystroke, voice, finger geometry, iris, retina, DNA, hand geometry (Prashanth et al., 2009) or odour (Rashed & Santos, 2010a).

Dogs use their noses to recognize things via odour (Wikipedia, 2009). When they search they can use their memory, since smell sense is linked to memory and emotion (Brewster et al., 2006).

Human beings smell is not used with such accuracy. The reasons may be summarized in the lacking of research and the IT tools to enable these devices to work (Brewster et al., 2006). Moreover, this field is under development (Korotkaya, 2009) and so it is much less well understood with comparison with image based or voice recognition (Brewster et al., 2006).

The first mechanical cash issuer was removed after six months due to the lack of customer acceptance (MIT, 2003). For that reason, Technology Acceptance Model (TAM), figure 1, (or any suitable model) should be used to measure the acceptance of such machines. We validate TAM in term of odour acceptance as user interface in the ATM machines OTM).

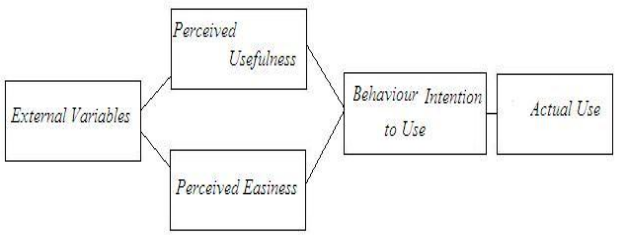


Fig. 1: Technology Acceptance Model (Rashed & Santos, 2010b)

This paper is organised as following: in section 2 we demonstrate the literature review. We present the problem statement and our approach in sections 3 and 4, respectively. In section 5 some conclusions are presented.

II. Literature Review

Some studies discussed the security within the banking sector and the technological solutions used. (Mao & Palvia, 2006) extended a US based research model and applied it to Chinese culture, contacting by e-mail 30 organisations. They compared their findings with the existing studies and found that their findings supported previous US studies. Moreover, they confirmed the suitability and applicability of TAM, TRA and IDT to study IT acceptance in Chinese culture.

[1]*Dr. Abdullah A. Rashed is with the Research Centre Algoritmi, University of Minho, Portugal, (telephone: 00351917410255 here email: rashed@dsi.uminho.pt)*
[2]*Prof. Henrique Santos is with the Research Centre Algoritmi, University of Minho, Portugal (telephone: +351 253 510 319 email: hsantos@dsi.uminho.pt)*

(Kllopping& McKiinney, 2004) studied consumer e-commerce as a technology adoption process and evaluated the suitability of Task-Technology Fit (TTF) and TAM. They discussed the use of the TAM to predict online shopping activity, both the intention to shop online and actual purchases. They suggested two alterations to the use of TAM—perceived ease of use was not linked to perceived usefulness, and perceived usefulness is directly linked to actual use. Moreover, they found that the (TTF) model is a valuable addition to TAM for online shopping tasks.

(Tibenderana & Ogao, 2008) studied the acceptance and use of digital library services by the library end-users in Ugandan universities. They suggested modifying The Unified Theory of Acceptance and Use of Technology (UTAUT) by replacing —effort expectancy" and —voluntariness" with "relevancy", —awareness" and —benefits" factors. They developed the Service Oriented UTAUT (SOUTAUT) model whose dependent constructs predict users' acceptance and use of e-library services. Their study found that relevancy moderated by awareness plays a major factor in acceptance and use of e-library services.

(De Magalhães et al., 2006) focused on user authenticating as a complex problem. They discussed the ethical issues for many times authentication and thought that they must be done without the collaboration of the authenticated user. They introduced keystroke dynamics biometrical technology as solution and difficult of intrusive issues, when used in collaborative mode.

(AlZomai et al., 2008) discussed the authentication problems of security in online banking of using SMS for transactions. They stated that online banking security should be enhanced focusing on usability more than security technical and mechanisms. They suggested the SMS authorization scheme. Their experiment aimed to simulate the online bank using website to do the transactions. They attacked their approach to make sure that it would work properly. Their attack succeeded in 21%. They justified that as user should have more experience.

(Singh et al., 2008) presented a shared family accounts framework. Their experiment consisted of sixteen families that included eight who used individual profiles at home, and eight who shared a single profile. They concluded that it is necessary to bridge the gap between authentication systems and social practice leads to a weakening of overall security. In addition, they stated that could lead to a lessening of the security of digital banking systems by having a security design that takes into account the importance of social and cultural practice.

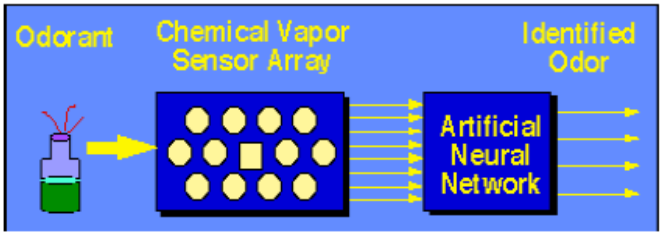
(Hao et al., 2009) discussed the two problems of online banking: first problem online security lacked to the attention and research that should focus on the security issues related to the client side. The second problem related to the huge number of security product that would increase the difficult level to test the category and standards of the security. They presented a scheme to design a compliance testing system for the security of online banking. The presented scheme aimed to obtain suggestions from testers that might help to design security testing and identify potential vulnerabilities in current online banking systems.

(Heckle et al., 2007) studied the biometrics acceptance. 24 participants were asked to participate in a role play to evaluate the use of fingerprint in online book purchasing. Their findings showed that respondents were comfortable after perceiving the benefit of using that biometrics. They suggested explaining obvious benefits would lead to more perception of usability and higher rate in acceptance.

(Coventry et al., 2003) discussed consumer-driven usability and user acceptance of biometrics tools. They focused on using iris as ATM interface. The applied qualitative, laboratory and field based studies. Their results showed that 90% were satisfied with iris verification method and they would select it over signatures or PINs.

(Saadé et al., 2007) studied the viability of TAM in multimedia. They conducted a comparative study consisting of 362 students. Results showed that TAM was a solid theoretical model where its validity could extend to the multimedia and e-learning context. They concluded that the multimedia learning system users is an important step towards a better understanding of the user behavior on the system and a multimedia acceptance model.(Rashed & Santos, 2010a) studied the acceptance of odour in authentication systems among youth users in the Arab countries. They found that presenting the tool in acceptable form would accelerate the acceptance and adoption of this tool.Technology continues to evolve and improve, so more work is required to address the usability issues which will be a key to successful implementation of biometrics within a general public application such as banking. Three patents related to odour machines are registered (Fukui et al., 1989), (Kao et al., 1999) and (Friedli & Gaussmann, 2005). Moreover (Korotkaya, 2009) introduced the electronic noses (ENoses) model as an odour recognition as shown in figure 1. Now the question is would users accept to use odour in ATM machines?

Figure 2: E-Nose (Korotkaya, 2009)



III. Discussion

Aiming to answer our research question, a questionnaire was distributed via website to a local university community. We selected e-mail as the target technology (postgraduate students in the department).

The constructs used existing scales from previous studies. Questionnaire contained 11 statements and fully anchored 5-point Likert scales were used with end points being for one —extremely disagree" and five —extremely agree". The instrument was developed in English. We received 48 responses. The main findings are:

i. Age of the respondents was within the interval [20-30] that represents youth people as shown in figure 3.

15-19	20-30	31-39	More than 40
8	22	17	1

Figure 3: ages of the respondents

ii. Education level: most of the respondents obtained high education as shown in Figure 4.

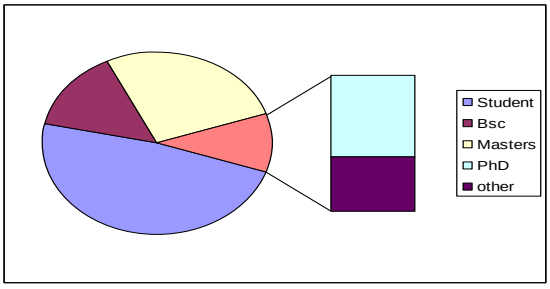


Fig. 4: Education level of the education

iii. The specialization of the sample is shown in figure 5 (the context is a technological university department).

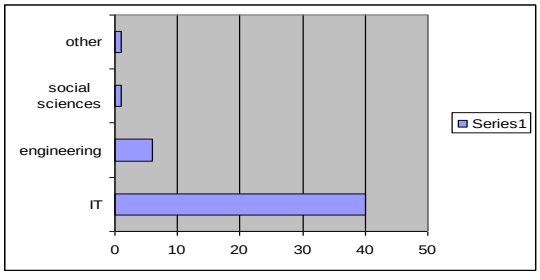


Fig. 5: specialization of the sample

iv. 19 participants (40%) believe it would be easy to use odour as an authentication technique in ATM machine, whereas only 8 (25%) think it would not be easy (see Figure 6).

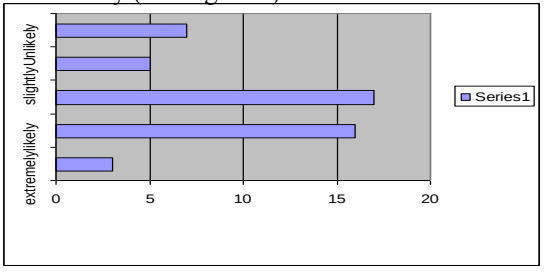


Fig. 6: easiness of odour in ATM machine

v. In contrast of the expected hypothesis, 15 respondents (31%) found odour for authentication in ATM machine would not improve their performance and 14 (29%) thought that it would improve their performance.

vi. 21 respondents (44%) intended to use odour as ATM authentication system and 15 respondents (31%) did not decide.

vii. The majority, 22 respondents (46%), point that using odour in ATM machines would be good idea.

viii. The majority of the respondents had not used odour in ATM machines.

ix. In comparison with other authentication mediums like iris, finger print, retina, and ear shaped, the respondents ranked odour as shown in figure 7.

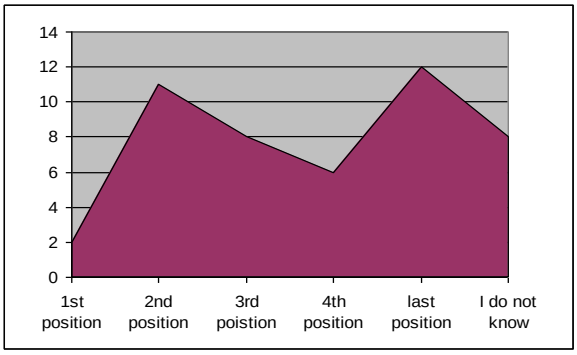


Figure 7: ranking odour with comparison with other biometrics

x. The main comments submitted by the sample were the following:

1) Is it secure? Using odour as an authentication method in Bank ATM? Odour is easier to be copied and this makes the system absolutely insecure.
2) Odour seems to be easy to be hijacked; I don't think that it is safe.
3) Odour is an innovative idea, but attention to its security. Otherwise it can be used for other less sensitive fields.

The standard deviations were calculated and mostly equal for all study parameters that showed it was consistent and significant ….

IV. Proposed Solution

Odour automated teller machine (OTM), as discussed here, is a computerised telecommunications device that provides the customers of a financial institution access to the financial transactions in public places without the need for a human clerk or bank teller. Customers are identified by their odour.We assume that OTM would have its own shape as it would allow only one customer to use it at the same time and be alone with some special cabinet. The overall OTM system can be described by the following:

i. A place for only one person.
ii. Thermal system that help to extract the odour from users and for detecting living bodies.
iii. Odour sensors to detect the smell and the associated module for improved authentication.
iv. Another solution is that users who need to authenticate themselves should carry their e-citizen cards.
v. This card should obtain (or provide?) both:
   1) Microprocessor that can compare the extracted odour and the biometrics stored information; and
   2) Stored biometrics: encrypted digitized format stored in the card.

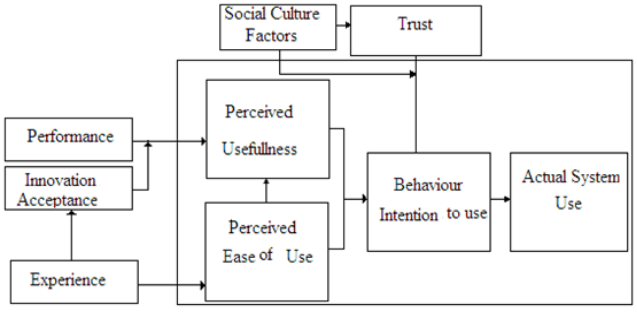


Figure 8: modified TAM

For the functional and non-functional requirements; Authors believe that they will be the same as traditional ATM in (Mylopoulos et al., 2007).

V. Conclusion and Future Work

Most of the respondents were youth and they found it good idea and easy to use odour as an interface in authentication system and useful. Moreover they reported they would use it frequently if it would be available. The majority of the respondents thought that it would not improve the performance. The biometric technologies create the challenge of avoiding attacks before they take place (De Magalhães et al., 2006). Results show that perceived the usefulness does not have a significant effect on intention to use. However, there is significant effect between easiness of use and intension to use. Moreover, due to the type of the technology and risk that might affect; security is still first element for the users (figure 8), so we suggest adding user concerns to the model. The reason is that user would prefer difficult tool if it is secure.It is concluded that the problem is how we could present the odour to usersIt is a challenge to apply this approach due to:

i. Acceptance by the customers due to its simplicity: customers have to do nothing whereas people used to apply very complicated authentication depending on the importance of that issue.
ii. Many people might be worries (comments of the respondents) that this approach might be hijacked easily (smell of odour on clothes) and needs to be strengthened with another approach that enhances it is performance.

VI. References

1) AlZomai M., AlFayyadh B., Audun Jøsang A., cCullagh A.(2008), An Experimental Investigation of the Usability of Transaction Authorization in Online Bank Security Systems, Proceedings of the sixth Australasian conference on Information security - Volume 81, Wollongong, NSW, Australia , pp:65-73.
2) Bala D. (2008), Biometrics and Information Security, Proceedings of the 5th annual conference on Information security curriculum development, Kennesaw, Georgia, ISBN:978-1-60558-333-4, pp: 64-66
3) Brewster S., McGookin D., Miller C.(2006), Olfoto: Designing a Smell-based Interaction, Proceedings of the SIGCHI conference on Human Factors in computing systems, Montréal, Québec , ISBN:1-59593-372-7, pp: 653 - 662
4) Coventry L., De Angeli A., Johnson G.(2003),Usability and Biometric Verification at the ATM Interface, Proceedings of the SIGCHI conference on Human factors in computing systems, Ft. Lauderdale, Florida, USA, ISBN:1-58113-630-7, pp: 153 - 160.
5) De Magalhães S., Kenneth R. Santos H., Sérgio P., Kenneth P., Santos, Santos H. (2006), Keystroke Dynamics : Stepping Forward in Authentication, "GESTS International Transactions on Computer Science and Engineering", .ISSN 1738-6438, pp:29-30.
6) Friedli P., Gaussmann A. (2005), System for Security Checking or Transport of Persons by an Elevator, Application number: 10/829,489, Publication number: US 2005/0138385 A1, Filing date: Apr 22, 2004, Inventors: , U.S. Classification 713, International Classification H04K001/00.
7) Fukui K., Shigemori T., Ehara K.(1989), Smell Sensing Element and Smell Sensing Device, Patent number: 5047214, Filing date: Mar 8, 1989, Issue date: Sep 10, 1991, Assignees: New Cosmos Electric Co., Ltd., Primary Examiner: Howard Hampel, U.S. Classification 422/98; 422/96; 422/88; 73/310.6; 338/34, International Classification G01N 2712.
8) Hao C., Jean-Pierre C., Castillo O., Douglas C., Feng D., Lee J. (2009), Security Testing and Compliance for Online Banking in Real-World, IMECS 2009: International MultiConference of Engineers and Computer Scientists , VOLS I AND II Book Series: Lecture Notes in Engineering and Computer Science, pp: 1039-1043.

We would like to thank Dr. Leonel Santos from information system department (DSI) from University of Minho, for his help. We thank also www.estibyan.com web site for their cooperation.

9) Heckle, R. , Patrick A., Ozok A. (2007), Perception and Acceptance Of Fingerprint Biometric Technology, ACM International Conference Proceeding Series; Vol. 229, Proceedings of the 3rd symposium on Usable privacy and security, Pittsburgh, Pennsylvania, PP: 153 - 154, ISBN:978-1-59593-801-5
10) Kao I., Milman I., Schneider D., Willard R. (1999), Authentication Framework for Multiple Authentication Processes , Patent number: 6651168, Filing date: Jan 29, 1999, Issue date: Nov 18, 2003, , Assignees: International Business Machines, Corp. Primary Examiner: Gilberto Barrn, Secondary Examiner: Doug Meislahn, Attorneys: Morgan & Finnegan, LLP, Joseph C. Redmond, Jr., Application number: 9/240,492, U.S. Classification 713/185; 713/182; 713/155, International Classification H04L 930
11) Kllopping I. McKiinney E. (2004), Extending the Technology Acceptance Model and the Task-Technology Fit Model to Consumer E-Commerce, Information Technology, Learning, and Performance Journal, Vol. 22, No. 1, Spring 2004 35
12) Korotkaya Z. (2009), Biometric Person Authentication: Odor, available via http://www.it.lut.fi/kurssit/03-04/010970000/seminars/Korotkaya.pdf.
13) Mao E., Palvia P.(2006), Testing an Extended Model of IT Acceptance in the Chinese Cultural Context , ACM SIGMIS Database, Volume 37 , Issue 2-3 (Spring-Summer 2006), SESSION: Research contributions, pp : 20 - 32, ISSN:0095-0033.
14) MIT School of Engineering (2003), Inventor of the Week: Range Estimation Trainer, available via http://web.mit.edu/invent/iow/simjian.html.
15) Mylopoulos J., Wang Y., McIlraith S., Yu Y (2007) An automated approach to monitoring and diagnosing requirements, International conference on Automated software engineering, Pages: 293-302, Atlanta, Georgia, USA Atlanta, Georgia,
16) Oshlyansky L., Cairns P., Thimbleby H. (2008) , Validating the Unified Theory of Acceptance and Use of Technology (UTAUT) tool cross-culturally, BCS-HCI '07: Proceedings of the 21st British HCI Group Annual Conference on HCI 2008: People and Computers XXI: HCI,, Volume 2,pp: 83-86.
17) Prashanth C. , Ganavi S., Mahalakshmi T. ,Raja K.,Venugopal K., Patnaik L. (2009), Iris Feature Extraction Using Directional Filter Bank, for Personal Identification, Proceedings of the 2nd Bangalore Annual Compute Conference on 2nd Bangalore Annual Compute, Article No. 6, ISBN:978-1-60558-476-8.
18) Rashed A., Santos H. (2010a), Odour User Interface for Authentication: Possibility and Acceptance: Case Study, The International MultiConference of Engineers and Computer Scientists 2010 (IMECS2010), (The 2010 IAENG International Conference on Bioinformatics), Hong Kong.
19) Rashed A., Santos H., (2010b), Multimodal Biometrics and Multilayered IDM for Secure Authentication, accepted, ICGS3 6th International Conference for on Global Security, Safety and Sustainability, 1-3 September 2010, Braga, Portugal.
20) Rashed A., Santos H., (2010c), OTM Machine Acceptance: in the Arab Culture, accepted, ICGS3 6th International Conference for on Global Security, Safety and Sustainability, 1-3 September 2010, Braga, Portugal.
21) Saadé R., Nebebe F., Tan W. (2007), Viability of the "Technology Acceptance Model" in Multimedia Learning Environments: A Comparative Study, Interdisciplinary Journal of Knowledge and Learning Objects Volume 3, 2007.
22) Singh, S., Cabraal A., Demosthenous C., Astbrink G., Furlong M.(2008), Password Sharing: Implications for Security Design Based on Social Practice, Computer Supported Cooperative Work, Proceedings of the ACM 2008 conference on Computer supported cooperative work, San Diego, CA, USA, Pages 669-678, ISBN:978-1-60558-007-4
23) Sukhai N. (1998), Access Control & Biometrics, Access control & biometrics, Proceedings of the 1st annual conference on Information security curriculum development, Kennesaw, Georgia, ISBN:1-59593-048-5, pp: 124 - 127
24) Tibenderana P., Ogao P. (2008) , Acceptance And Use of Electronic Library Services in Ugandan Universities, International Conference on Digital Libraries, Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries, Pittsburgh PA, PA, USA , SESSION: Platforms and users in digital libraries
25) Wikipedia webpage, (2009), Dog, Retrieved from http://en.wikipedia.org/wiki/Dog.

A Novel RFID Data Mining System: Integration Of Effective Sequential Pattern Mining And Fuzzy Rules Generation Techniques

[1]Barjesh Kochar and [2]Rajender Singh Chhillar

GJCST Computing Classification
I.2.3 ,H.2.8

***Abstract*- Data warehousing and Data mining find enormous applications; RFID technology is one among them. A RFID data warehousing system with novel data cleaning, transformation and loading technique has been proposed in the previous work. The system has been dedicatedly implemented in one of the significant RFID applications tracking of goods in warehouses. The warehoused RFID data is in specific format and so an effective mining system is required to mine the needed information from the database. The existing mining algorithms are inefficient in extracting the information from the warehoused RFID data. In this paper, a novel data mining system is proposed, which effectively extracts the information regarding the nature of movement of the RFID tags. The proposed mining system generates an intermediate dataset (Idataset) from the warehoused dataset. From the I-dataset,sequential patterns are mined with different pattern length combinations. From the mined sequential patterns, fuzzy rules are generated, which depicts the nature of movement of the RFID tags. The implementation results show that the proposed mining system performs well by extracting the significant RFID tags and its combinations and the nature of movement of the tags.**

Keywords- Data mining system, sequential pattern mining, RFID data, I-dataset, fuzzy rules, RFID readers, RFID tags.

I. Introduction

Enormous data collection has been possible throughout the previous decades as a result of the recent development of information, and availability of low-priced storage. The ultimate intention of this huge data collection is for the utilization of this information to realize competitive benefits, by analyzing the data i.e., determining formerly unidentified patterns in data that can direct the process of decision making [1] [17]. Conventional data analysis methods are generally based on direct handling of the data by a person and are not extendable to large data sets.Fundamental tools are available in database technology for efficient storage and searching of large data sets. But,helping humans to analyze and understand large masses of data continues to be a challenging and unresolved issue. New methods and intelligent tools offered by the upcoming data mining field assure to meet these challenges. The term data mining, which is also called Knowledge Discovery in Databases (KDD), is defined as —The non-trivial extraction of implicit, previously unknown, and potentially useful information from data" [2] [15].

1Barjesh Kochar and 2Rajender Singh Chhillar

Data mining [3] as a multidisciplinary united effort from databases, machine learning, and statistics, is winning in turning masses of data into small valuable pieces. The ultimate objective of a data mining task in a real-world application might be e.g. to permit a company either to enhance its marketing, sales, and customer support operations or to identify a fraudulent customer through better understanding of its customers. Data mining methods have been effectively applied in various fields including marketing [22], manufacturing, process control, fraud detection [21], bioinformatics, information retrieval,adaptive hypermedia, electronic commerce and network management [4]. Broadly, the data mining tasks can be categorized into two types: Descriptive mining and Predictive mining [11]. The descriptive mining denotes the method in which the fundamental characteristics or common properties of the data in the database are portrayed. The method of predictive mining figures out patterns from the data so that predictions can be made. The predictive mining methods include tasks like Classification, Regression and Deviation detection.

Mining of information from a huge database finds many latest and emerging applications. Radio Frequency Identification (RFID) is one among the fields that incorporate the sequential pattern mining in RFID database. Radio Frequency Identification (RFID) is a high-speed, realtime, precise information gathering and processing technology, which distinctively identifies object by radiofrequency signal [14]. RFID technology can help a broad variety of organizations and individuals, for instance, hospitals and patients, retailers and customers, and manufacturers and distributors all through the supply chain to accomplish considerable productivity gains and efficiencies [23]. It is also extensively applied in logistics, traffic management, medical treatment, national defense,mining, and endlessly infiltrated to new fields [12]. The heart of an RFID system is the —RFID tag" which encloses an integrated circuit and a transponder for data storage and transmission [25]. The tag can be tied to an object and be read out with reader hardware either in the form of a handreader or a RFID gate [24]. The tags are radically different from printed barcodes in their capability to hold data, at which range the tags can be read, and the nonexistence of line-of-sight constraints [26].

The RFID technology and its applications identified as one of the most hopeful technologies of the century are in a slow maturation period. Warehousing the RFID data and mining information from the warehoused RFID data are the major

issues that persist in the technology. In the previous work, we have proposed a novel data cleaning, transformation and loading technique to warehouse the RFID data efficiently. The warehoused data is in a specific format and so mining the required information needs an effective mining system.But, the existing mining algorithms are ineffective in their way of mining the information from the warehoused RFID data. To overcome this issue, in this paper, we have proposed an effective data mining system to mine the warehoused RFID data. In the proposed system, firstly, an intermediate dataset is generated from the warehoused RFID dataset. Secondly, a sequential pattern mining technique is proposed to mine the reader pattern from the intermediate dataset. Finally, fuzzy rules are generated based on the mined sequential pattern to make the decisions from the path followed by the RFID tags. The rest of the paper is organized as follows: Section 2 gives a brief introduction about the Sequential Pattern Mining and Section 3 makes A short review about the recent literary works. Section 4 details the proposed RFID data mining system with required mathematical formulations and tabulations. Section 5 discusses about the implementation results and Section 6 concludes the paper.

II. SEQUENTIAL PATTERN MINING

Mining Frequent Itemsets from transaction databases is A basic task for several types of knowledge discovery such as association rules, sequential patterns, and classification [5].Sequential pattern mining, which finds out frequent patterns in a sequence database, is a significant issue among the various data mining issues [7]. Discovering sequential patterns from a huge database of sequences is a significant issue in the field of knowledge discovery and data mining [20]. The sequential pattern mining problem was first presented by Agrawal and Srikant [19]. Mining Sequential Patterns in huge databases has become a significant data mining task with wide range of applications [9] [18].The aim of sequential pattern mining is to discover all frequent sequential patterns with a user-specified minimum support [10].The sequential pattern is a prearranged list of frequent itemsets in a customer‘s sequence and discloses the customer‘s purchasing manners in transactional databases, from which the policy-maker may make a knowledgeable business decision [16]. It extracts patterns that occur more frequently than a user-specified minimum support while preserving their item occurrence order [9]. It is helpful in various applications for enhancing the quality of analysis, including the analysis of web surfing, transactional customers‘ behaviors, network alarm patterns, business analysis, web mining, security and bio-sequences analysis [16]. For instance, supermarkets frequently gather customer purchase records in sequence databases in which A sequential pattern would signify a customer‘s purchasing behavior. In such a database, every purchase would be denoted as a set of items purchased, and a customer sequence would be a sequence of such itemsets [8]. The issue of sequential patterns discovery was motivated by retailing industry problems. Nevertheless, the results apply to several scientific and business domains, like stocks and markets basket analysis, natural disasters (e.g. earthquakes), DNA sequence analyses, gene structure analyses, web log click stream analyses, and so on [13]. In this task, time is the most significant factor, particularly when the results are required in a limited period of time [6].

III. RELATED WORKS

Jigyasa Bisaria et al. [27] have put forward a rough set orientation to the problem of constraint driven mining of sequential pattern. They have divided the search space of sequential patterns utilizing indiscernibility relation from the theory of rough sets and they have put forward an algorithm that permitted pre-visualization of patterns and imposition of diverse types of constraints in the mining task. The algorithm C-Rough Set Partitioning was more than ten times faster than the naive algorithm SPRINT that was founded on different types of regular expression constraints. Shigeaki Sakurai et al. [28] have detailed that all frequent sequential patterns included in sequential data are efficiently found out by the sequential mining methods. Those methods assessed the frequency by utilizing the support, which is the previous criterion that satisfied the Apriori property.Nevertheless, since the patterns were general and the analysts cannot acquire new knowledge from the patterns, the discovered patterns do not always correspond to the interests of the analysts. To find out sequential patterns that are more appealing for the analysts, they have put forward A criterion, namely, the sequential interestingness. They have illustrated how the criterion is related to the support and that the criterion fulfilled the Apriori property. Moreover,founded on the proposed criterion, they have put forward an efficient sequential mining method. Finally, by applying the method to two types of sequential data, they have demonstrated the effectiveness of the proposed method. Themis P. Exarchos et al. [29] have offered a methodology with two process for sequence classification that utilizes sequential pattern mining and optimization. In the first stage, a sequence classification model, founded on a set of sequential patterns was defined and two sets of weights were set up, one for the patterns and the other for classes. In the second stage, the weight values were assessed by utilizing an optimization technique to accomplish optimal classification accuracy. By changing the number of sequences, the number of patterns and the number of classes, extensive appraisal was done on the methodology, and it has compared with similar sequence classification approaches. Chen [30] has found out a data structure, Up Down Directed Acyclic Graph (UDDAG), for efficient sequential pattern mining. Along both ends of detected patterns, UDDAG has permitted bidirectional pattern growth. Therefore, in $\log 2(k+1)$ levels of recursion, a length-k pattern can be identified, which has resulted in fewer levels of recursion and faster pattern growth. When minSup is large such that the average pattern length is close to 1, UDDAG and PrefixSpan have similar performance because the problem degrades into frequent item counting problem. Nevertheless, UDDAG scales up much better. In

scalability tests, UDDAG often surpassed PrefixSpan by almost one order of magnitude. UDDAG was also significantly faster than Spade and LapinSpam. Barring extreme cases, UDDAG has utilized memory comparable to that of PrefixSpan and less than that of Spade and LapinSpam. Moreover, extension of UDDAG towards applications that involve searching in large spaces has been made possible due to its special feature.Jirachai Buddhakulsomsiri and Armen Zakarian [31] have offered a sequential pattern mining algorithm which has enabled the extraction of hidden knowledge from a huge automotive warranty database by product and quality engineers. Elementary set concept and database manipulating methods were utilized by the algorithm to look for patterns or relationships among instances of warranty claims over time. IF-THEN sequential rules represent these patterns, where the IF part has included one or more instances of warranty problems at one time, and the THEN part has included warranty problem(s) that happen at a later time. When the sequential patterns were generated, the algorithm utilized rule strength parameters to filter out insignificant patterns so that only significant rules were supported. The Significant patterns have provided knowledge of one or more product failures that leads to future product fault(s). Warranty data mining application from the automotive industry has demonstrated the effectiveness of the algorithm. For the automotive case, A discussion on the sequential patterns created by the algorithm and their interpretation were also given.Ya-Han Hu et al. [32] have illustrated multi-time-interval sequential patterns, a variation of time-interval sequential patterns, which can be disclosed the time-intervals between all pairs of items in a pattern to offer more time-related knowledge. As a result, two efficient algorithms, called the MI-Apriori and MI-PrefixSpan algorithms were developed by them to solve this problem. Experimental results have illustrated MI-PrefixSpan algorithm was faster than the MIApriori algorithm, however in long sequence data, MIApriori algorithm has better scalability.Ya-Han Hu et al. [33] have concentrated on mining sequential patterns in the business-to-business (B2B) environment. Utilizing traditional methods in the B2B environment may end up in numerous uninteresting and meaningless patterns and extensive computational time due to very long sequences of customers, and every customer often purchasing majority of the items. They have set up three conditions (constraints) to solve those problems, namely, compactness, repetition, and recency and consider them together with frequency in choosing sequential patterns. To find out frequent sequential patterns which have fulfilled the conditions, an efficient algorithm has been built up. Computational efficiency and effectiveness of their proposed method in the B2B environment in extracting useful sequential patterns has been confirmed by observed results.

IV. THE RFID DATA MINING SYSTEM

The RFID data has been warehoused effectively by means of a novel data cleaning, transformation and loading technique, which was dedicatedly proposed for RFID data. The warehoused data has been compressed and loaded in A specific format and so extraction of required knowledge from them is difficult. The proposed novel RFID data mining system effectively mines the required knowledge from the warehoused RFID data. The proposed system is comprised of three stages, namely, Generation of I-dataset, sequential pattern mining and generation of fuzzy rules. In the first stage, a dataset is generated from the warehoused RFID data. In the second stage, sequential patterns are mined using the proposed mining technique and in the third stage, fuzzy rules are generated from the mined sequential patterns. Prior to detail the proposed mining system, the format of the warehoused RFID data is briefed in the following sub-section for easy understanding of the proposed mining system.Let R_i ; $0 \le i \le N_R - 1$be the number of RFID readers present and $_j T$; $0 \le j \le N_T - 1$ be the RFID tags which are in movement, where, N_R is the number of readers and N_T is the number of Tags present in the warehouse. The tags may enter into any reader at any time while in process. Each RFID tag has its own Electronic Product Code (EPC), which represents the class to which $T\,j$ belongs. The Tag representation with EPC code can be given asT_{kj} ;$k \in [EPC_1, EPCNc]$, where,Nc is the number of product classes available.Generally, the tags are given I.D. only by their EPC code.But, here the tags are given numerical IDs along with the EPC code. As an example, assuming the Tag representation is T_{10} ,T_{11} ,T_{21} ,T_{32} ,T_{42} ,T_{52} , …., -1 $NcNT\,T$, then,T_{10} represents the tag that belongs to the code class EPC_1 and code ID _0‘ and, T_{11} represents the tag that belongs to the code class $EPC1$ and code ID _1‘ and so on. The EPC of a tag defines the product to which the tag is attached. For example, EPC_1 may be the code for the product _pen‘;EPC_2 may be the code for the product _calculator‘, and so on. The RFID readers may be present in any of the chambers of a warehouse and they monitor the movement of the RFID tags (i.e. every product present in the warehouse).Hence, when the tag moves from one location to the other, the Reader present in that location makes an entry in the database by its tag ID and the time of presence of the tag.Thus obtained RFID data has been cleaned, transformed and loaded effectively in the data warehouse by our novel data cleaning, transformation and loading technique, which was done in the previous work. An exemplary view of the warehoused data is given in the Table

Table I: An exemplary view of warehoused RFID data with three readers

S.No	Reader R_1		Reader R_2		Reader R_3	
	Tags IN	Tags OUT	Tags IN	Tags OUT	Tags IN	Tags OUT
1	[3,20,30]	[5,10]	[6,12]	[8,14]	[31,21]	[15,40]
2	[11,1,3,8]	[20]	[13,7,20]	[6,12,17]	[6,12,17]	[31, 19]
3	[27,25]	[30, 3, 11]	[3,11]	[13]	[30,1β3]	[21]
4	[24,9]	[1,27]	[27,30]	[3,7]	[1]	[6,12,30]
5	[15]	[8,24,9]	[9,24]	[27, 11]	[27, 8]	[27]

The RFID data for only three readers are given in the table. In the Table I, the tags that entered into the corresponding reader‘s range and the tags that left out from the corresponding reader‘s range for some five interrogations are given. The interrogation period, the frequency of occurrence and the availability index are given in the Table

Table 2: The warehoused RFID data with interrogation period, frequency of occurrence and availability index for three readers

S. No	Reader R_1			Reader R_2			Reader R_3		
	Transformed Interrogation Time	Frequency of Occurrence	Availability Index	Transformed Interrogation Time	Frequency of Occurrence	Availability Index	Transformed Interrogation Time	Frequency of Occurrence	Availability Index
1	1.08	7	[3,5,6]	0.09	6	[41,34]	0.06	7	[14,15]
2	2.1	5	[8,10]	0.01	5	[21,12,13]	1.2	6	[23,1]
3	1.3	3	[16,3,9]	0.08	3	[1,4,5]	1.0	4	[1,7,10]
4	1.9	2	[15,17]	1.4	2	[4,2,7,8]	1.9	2	[35,16]
5	4.5	1	[24,20,1,4]	4.1	2	[23,15]	3.8	1	[9]

The proposed data mining system extracts the information of the flow of path for every tag. It generates rules based on the mined data so that the nature of each product (in terms of its RFID tag) can be understood well. To accomplish this, the transformed data present in the Table I is sufficient for the proposed system.

A. GENERATION of I-dataset

In the first process of the proposed data mining system, the warehoused data is processed and an I-dataset is generated.The I-dataset is generated by querying the warehoused dataset. The proposed system intends to mine the information in the dataset given in the Table I. Firstly, the dataset given in the Table I is mathematically represented for further querying process. This can be accomplished by representing the set of Tag IDs that are entering into the Reader‘s range and the set of Tag IDs that are leaving out from the Reader‘s Range from Table I. The set representation is given as $\{T^{IN}\}_{il}$ and $\{T^{OUT}\}_{il}$: $0 \le \le -1$ / N_1 where, / N is the number of interrogations performed by the reader, such that, $\{T^{IN}\}_{il} \subseteq \{R^{IN}\}_i$ and $\{T^{OUT}\}_{il} \subseteq \{R^{OUT}\}_i$. Hence, two sets, $\{R^{IN}\}_j$ and $\{R^{OUT}\}_j$, are generated, where, each set is comprised of N_1 subset. But, the cardinality of each subset present in $\{R^{IN}\}_i$ (here, $i \in (0, N_R - 1)$) need not be equal to each other. Also, the cardinality of $\{R^{OUT}\}_i$ need not be equal to each other. Each subset is comprised of the Tag IDs that entered and left out from the Reader. As $\{R^{OUT}\}_i$ $\{R^{IN}\}_i$ stated earlier, the set is queried to generate the Idataset Querying can be performed either over the set or over the set Here, querying is performed over the set, $\{R^{OUT}\}_i$. The I-dataset is comprised of all the RFID tags and the readers that cover the tags in their way. This can be obtained as follows

$$\{P(l)\}_j = \begin{cases} R_i ; \ if \ T_j \in \{T^{OUT}\}_{il} & \forall i \\ -1; \ otherwise \end{cases} \quad (1)$$

$$P_j = P_j - \alpha \quad (2)$$

In Eq. (1), the condition is checked by varying for all i and the Path set {P} is determined. For every Tag, a set {P} is determined (i.e. *Pj for T j*). Generally, some tags may be present in a reader‘s range for a long time (i.e. the tags may not appear in Tout for many interrogation cycles). In such circumstances, the set {P} holds an element of value -1 in the corresponding interrogation cycle. The filtered path set is obtained in the Eq. (2), where, $\alpha = \{-1\}$, in which the readers that covered the tags are present and then sorted in ascending order based on the Tag IDs.

Thus, by querying, (mathematically represented in Eq. (1) and Eq. (2)) the readers under which each tag has covered are collected and so the set {P} is determined and the Idataset is generated from the set {P}. For instance, the {P} is assumed *to be as:* $P_0 = \{R_1, R_3, R_5, R_{10}, R_{13}\}$,

P_1 = {R_2 , R_4 , R_1, R_6} , P_2 = {R_{11}, R_3, R_5 , R_1, R_3} , P_3 = {R_7 , R_9 ,R_5 ,R_7 ,R_{12} , R_4} and P_4 = {R1,R9 ,R4} . The exemplary I-dataset determined from the aforesaid set {P} is tabulated below.

Table 3: The exemplary I-dataset for four RFID tags

S.No	Tag IDs	Readers
1	T_0	R_1,R_3,R_5,R_{10},R_{13}
2	T_1	R_2,R_4,R_1,R_6
3	T_2	R_{11},R_3,R_5,R_1,R_3
4	T_3	R_7,R_9,R_5,R_7,R_{12},R_4
5	T_4	R_1,R_9,R_4

The field ‗Tag IDs' present in the I-dataset, which is given in the Table III, is comprised of all the Tag IDs and the field ‗Readers' for a record named *T*0 is comprised of all the Reader IDs that covered the Tag *T*0 in its course. The Idataset is subjected to sequential pattern mining so that the required path information is extracted from the I-dataset.

B. Mining Sequential RFID Data Patterns

In the I-dataset and the set P, a generalized Tag ID has been used. But, practically, the Tag IDs are represented with EPC code. So, from this instant, the Tag IDs are grouped as per the product with which the tags are attached. To accomplish this, it is assumed that the tags which has first NP number of IDs belongs to *the EPC* code *EPC*1, the tags which has second NP number of IDs belongs to the EPC code EPC2 , and so on. Hence, a new tag ID representationTab ; $EPC_1 \le a \le EPC_{N_-}$ And $0 \le b \le N_{P-1}$ is given for T_j; $0 \le j \le N_{T-1}$, where, b = j%NP and a= 1+ (j / NP) flow extracted from the set {P} is given as

$$T_{ab} = \left(R_h^{(0)} \rightarrow R_h^{(1)} \rightarrow R_h^{(2)} \rightarrow \cdots \rightarrow R_h^{(|T_{ab}|-1)}\right) \quad (3)$$

where, hÎ[0, NR −1] and | Tab | is the size of the pattern length.. Then, from Tab , a single length (L = 1) sequential pattern is mined. The pseudo code for mining the single length sequential pattern is given in Fig. 1.

```
Initialize S, sup
for a = 1 to Nc − 1
    set supa0 to 1
    set supa1 to 0
    set x to 1
    Assign Ta0 to Sa0
    for b = 1 to Np − 1
        if Tab ∈ S
            for y = 1 to |S|-1
                if Say is Tab
                    increment supay
                end if
            end for
        else
            Assign Tab to Sax
            increment supax
            increment x
            set supax to 0
        end if
    end for
end for
```

Fig 1: Pseudo code for mining single length sequential patterns from the Idataset

By mining as per the given pseudo code, the sequential pattern with L = 1(i.e. for every individual EPC code) is obtained from the pattern set {S}. Each pattern has its own support, i.e. frequency of occurrence of the pattern, which is obtained in the support set sup . From the obtained S_l (as $L = 1$), the sequential patterns with support greater than supmin ($\text{sup} \ge \text{supmin}$)are selected for further mining operations and the remaining patterns that doesn't satisfy supmin are neglected. Then, the sequential patterns with $L > 1$are mined. The pseudo code for mining the sequential patterns with $L > 1$ is given in the Fig. 2. The mined sequential patterns using the algorithm, which is illustrated in the pseudo code, are checked for the minimum support.The sequential patterns those have support greater than supmin is selected for further process.

```
for L = 2 to Np
    for every L - length combination
        S^(L)_{x1x2...xL} ← S^(1)_{x1} ∩ S^(1)_{x2} ∩ ··· S^(1)_{xL} : {xi; 1 ≤ i ≤ L} ∈ (0, Nc + 1)
        for y = 0 to |S_{x1x2...xL}| - 1
            sup^(L)_{x1x2...xL}(y) ← min{sup^(1)_{x1}(z), sup^(1)_{x2}(z) ··· sup^(1)_{xL}(z)}:
                S^(1)_{xi}(z) = S^(L)_{x1x2...xL} ; z ∈ [0, |S_{xi}| - 1]
        end for
    end for
end for
```

Fig 2: Pseudo code for mining sequential patterns with L > 1 from the Idataset

Hence, sequential patterns that are commonly present in the path flow are mined with different combination of their support. The obtained maximum length sequential patterns have strong support as it survives all the lower length combinations.

C. Generation of Fuzzy Rules Set

The sequential patterns mined by the mining algorithm with different combinations are subjected to generate fuzzy rules. The sequential patterns, $S^{(L)}_{x_g} : x_g \in (0, N_c + 1),\ 1 \le g \le L,\ 1 \le L \le N_c$, which are obtained from the mining algorithm are mathematically detailed as

$$S^{(1)}_{x_1} = \left(R_h^{(0)} \rightarrow R_h^{(1)} \rightarrow \cdots \rightarrow R_h^{\left(\left|S^{(1)}_{x_1}(z)\right|-1\right)}\right);\ 0 \le z \le \left|S^{(1)}_{x_1}\right| - 1 \quad (4)$$

$$S^{(2)}_{x_1x_2} = \left(R_h^{(0)} \rightarrow R_h^{(1)} \rightarrow \cdots \rightarrow R_h^{\left(\left|S^{(2)}_{x_1x_2}(z)\right|-1\right)}\right) \quad (5)$$

$$S^{(3)}_{x_1x_2x_3} = \left(R_h^{(0)} \rightarrow R_h^{(1)} \rightarrow \cdots \rightarrow R_h^{\left(\left|S^{(3)}_{x_1x_2x_3}(z)\right|-1\right)}\right) \quad (6)$$

$$S^{(N_c)}_{x_1x_2\cdots x_{N_c}} = \left(R_h^{(0)} \rightarrow R_h^{(1)} \rightarrow \cdots \rightarrow R_h^{\left(\left|S^{(N_c)}_{x_1x_2\cdots x_{N_c}}(z)\right|-1\right)}\right) \quad (7)$$

Hence, Nc kind (length) of pattern sets are obtained from the mining algorithm. Each pattern set consists of $\left|S^{(L)}_{x_g}\right|$

number of sequential patterns. For every kind of pattern set, a set of support values are also obtained, which is comprised of the support values for every sequential pattern present in the pattern set. To generate fuzzy rules, R_{TH}, termed as rule threshold, is set. Probably, the R_{TH} is generated within the interval (0,1). Then, N_e elements are taken from the pattern in sliding fashion and the rules are generated. Generally, the fuzzy rules are of the form, *if* $IN_1 = i_1, IN_2 = i_2, \cdots, IN_n = i_n$ then $OUT = i_{out}$. The same format is followed here and the fuzzy rules are generated. The step-by-step procedure of the generation of fuzzy rules is given below.

Step 1-Initialize $L = N_c$

Step 2-Get the L^{th} pattern set and initialize $z = 0$

Step 3-Get the z^{th} sequential pattern from the L^{th} pattern set and initialize $p = 0$

Step 4- Select a sub-pattern as follows.

$$IN_q = R_h^{(p+q)},\ 0 \le q \le N_e - 1 \qquad (8)$$

Step 5-Find all the sub-patterns as IN from $S^{(L)}_{x_1 x_2 \cdots x_L}$ similar to that of the selected pattern and select the following sub-pattern as OUT from the corresponding sequential pattern. For example, a sequential pattern of length L = 3 with a combination of the pattern set that corresponds to EPC code 1, 2 and 4 are given as

$$S^{(3)}_{124} = \begin{pmatrix} R_1 \rightarrow & R_2 \rightarrow & R_3 \rightarrow & R_4 \\ R_3 \rightarrow & R_1 \rightarrow & R_2 \rightarrow & R_4 \\ R_1 \rightarrow & R_2 \rightarrow & R_5 \rightarrow & R_6 \end{pmatrix} \qquad (9)$$

The support values for every sequential pattern, which is present in the above pattern set is given as

$$\sup^{(L)}_{x_1 x_2 \cdots x_L} = \begin{pmatrix} 5 \\ 6 \\ 7 \end{pmatrix} \qquad (10)$$

From the above pattern set, the selected IN and OUT for fuzzy rules are given as

$$IN = R_1 \rightarrow R_2 \qquad (11)$$

$$OUT = \begin{pmatrix} R_3 \rightarrow R_4 \\ R_4 \\ R_5 \rightarrow R_6 \end{pmatrix} \qquad (12)$$

Step 6-Sort the selected patterns based on the support values and generate fuzzy rules by considering the IN subpattern for _if' statement, OUT sub-pattern for _then' statement and the corresponding support value is given asfuzzy score. The fuzzy rule for the aforesaid example can be generated as

if IN = $R_1 \rightarrow R_2$ then OUT = $R_5 \rightarrow R_6$ with fuzzy score = 7

if IN = $R_1 \rightarrow R_2$ then OUT = R_4 with fuzzy score = 6

if IN = $R_1 \rightarrow R_2$ then OUT = R_3 ® R_4 with fuzzy score = 5

Step 7-Increment p and go to Step 4 until p < Ne

Step 8-Increment z and go to Step 3 until $z < \left| S^{(L)}_{x_1 x_2 \cdots x_L} \right|$

Step 9-Decrement L and go to Step 2 until L > 0

Step 10-Get all the fuzzy rules separately for every L^{th} pattern set. From the fuzzy rules, the nature of every tag (based on EPC code) as well as the significant combination of different tags can be understood. The significance of the combination of the tags is decided by the combination which has support greater than supmin .

V. Results and Discussion

The data mining system, which is proposed in this paper, is implemented in the working platform of JAVA (version JDK 1.6). The proposed system has been evaluated in the RFID application, tracking of goods in warehouses. Here, it is assumed that the warehouse holds some six stationery goods such as Pencil, Pen, Notebook, Diary, Paper clips, Memo holder and so Nc = 6 . The RFID tags are affixed with each product and so each product has its EPC code. The warehouse is considered to have 200 products/per class (NT = 200) and eight RFID readers (i.e. NR = 8) are available. From the warehoused data, the I-dataset is generated in the format, as given in the Table III. Then, the sequential patterns are generated with a minimum support supmin = 1. Some of the sequential patterns, which are generated from the I-dataset, are given in the Table IV. With the aid of the sequential patterns, the fuzzy rules are generated and the rules corresponding to the pattern are given in the Table V. The number of sequential patterns with different lengths and the number of fuzzy rules with different pattern combinations are depicted in the Fig. 3 and Fig. 4.

Table 5: Generated Sequential Patterns and its support with different possible lengths

Length	Pattern Id	Pattern	Support
	S_1	$R_3 \rightarrow R_8 \rightarrow R_4$	2
	S_1	$R_3 \rightarrow R_1 \rightarrow R_8 \rightarrow R_6 \rightarrow R_5$	1
	S_2	$R_3 \rightarrow R_2 \rightarrow R_1$	2
1	S_2	$R_4 \rightarrow R_7 \rightarrow R_1$	1
	S_3	$R_3 \rightarrow R_2 \rightarrow R_1 \rightarrow R_4$	2
	S_3	$R_4 \rightarrow R_3 \rightarrow R_2$	1
	S_4	$R_1 \rightarrow R_3 \rightarrow R_6$	2
	S_4	$R_3 \rightarrow R_2 \rightarrow R_3 \rightarrow R_4 \rightarrow R_7$	1
	S_5	$R_4 \rightarrow R_1$	2
	S_5	$R_7 \rightarrow R_1 \rightarrow R_2 \rightarrow R_5$	1
	S_6	$R_4 \rightarrow R_8 \rightarrow R_2$	2
	S_6	$R_3 \rightarrow R_8 \rightarrow R_1 \rightarrow R_6$	1
2	S_{12}	$R_3 \rightarrow R_8 \rightarrow R_4$	1
	S_{12}	$R_3 \rightarrow R_1 \rightarrow R_6$	1
	S_{13}	$R_4 \rightarrow R_3 \rightarrow R_4$	1
	S_{13}	$R_3 \rightarrow R_8 \rightarrow R_2 \rightarrow R_7$	1
	S_{14}	$R_7 \rightarrow R_4 \rightarrow R_8 \rightarrow R_5$	1
	S_{14}	$R_3 \rightarrow R_8 \rightarrow R_2 \rightarrow R_7$	1
	S_{15}	$R_3 \rightarrow R_1 \rightarrow R_6$	1
	S_{15}	$R_7 \rightarrow R_8 \rightarrow R_4 \rightarrow R_1$	1
	S_{16}	$R_7 \rightarrow R_3 \rightarrow R_2 \rightarrow R_6$	1

	S_{16}	$R_6 \rightarrow R_3 \rightarrow R_4$	1
	S_{23}	$R_4 \rightarrow R_8 \rightarrow R_7 \rightarrow R_5$	1
	S_{23}	$R_6 \rightarrow R_3 \rightarrow R_5 \rightarrow R_1$	1
	S_{24}	$R_3 \rightarrow R_3 \rightarrow R_1 \rightarrow R_4 \rightarrow R_6$	1
	S_{24}	$R_3 \rightarrow R_3 \rightarrow R_1$	1
	S_{25}	$R_6 \rightarrow R_1 \rightarrow R_3 \rightarrow R_2$	1
	S_{25}	$R_3 \rightarrow R_1 \rightarrow R_6 \rightarrow R_8$	1
	S_{26}	$R_4 \rightarrow R_1 \rightarrow R_3 \rightarrow R_6$	1
	S_{26}	$R_6 \rightarrow R_4 \rightarrow R_8 \rightarrow R_5$	1
	S_{34}	$R_7 \rightarrow R_3 \rightarrow R_3$	1
	S_{34}	$R_6 \rightarrow R_1 \rightarrow R_3 \rightarrow R_5$	1
	S_{35}	$R_7 \rightarrow R_3 \rightarrow R_3$	1
	S_{35}	$R_7 \rightarrow R_3 \rightarrow R_1$	1
	S_{36}	$R_6 \rightarrow R_3 \rightarrow R_4 \rightarrow R_8$	1
	S_{36}	$R_3 \rightarrow R_8 \rightarrow R_1 \rightarrow R_6$	1
	S_{45}	$R_4 \rightarrow R_6 \rightarrow R_1 \rightarrow R_3 \rightarrow R_7$	1
	S_{45}	$R_4 \rightarrow R_1 \rightarrow R_3 \rightarrow R_7$	1
	S_{46}	$R_6 \rightarrow R_4 \rightarrow R_2 \rightarrow R_7$	1
	S_{46}	$R_8 \rightarrow R_2 \rightarrow R_3 \rightarrow R_7 \rightarrow R_6$	1
	S_{56}	$R_8 \rightarrow R_3$	1
	S_{56}	$R_4 \rightarrow R_4 \rightarrow R_7 \rightarrow R_6 \rightarrow R_3$	1
	S_{123}	$R_3 \rightarrow R_1 \rightarrow R_6$	1
	S_{134}	$R_3 \rightarrow R_8 \rightarrow R_3 \rightarrow R_7$	1
	S_{136}	$R_6 \rightarrow R_3 \rightarrow R_4$	1
3	S_{233}	$R_4 \rightarrow R_8 \rightarrow R_3$	1
	S_{343}	$R_7 \rightarrow R_3 \rightarrow R_3$	1
	S_{356}	$R_7 \rightarrow R_3 \rightarrow R_1$	1
	S_{436}	$R_3 \rightarrow R_6 \rightarrow R_2 \rightarrow R_7$	1
	S_{456}	$R_8 \rightarrow R_3$	1

Table 5: Fuzzy rules with fuzzy score generated from the extracted sequential patterns

Length	Pattern combination	Fuzzy Rule	Fuzzy score
	S_{123}	if IN=R_3 ,then OUT= $R_1 \rightarrow R_6$	1
	S_{134}	if IN=$R_3 \rightarrow R_8$,then OUT= $R_3 \rightarrow R_7$	1
	S_{134}	if IN=$R_8 \rightarrow R_3$,then OUT= R_7	1
	S_{136}	if IN=R_6 ,then OUT= $R_3 \rightarrow R_4$	1
3	S_{233}	if IN=R_4 ,then OUT= $R_8 \rightarrow R_3$	1
	S_{343}	if IN=R_7 ,then OUT= $R_3 \rightarrow R_3$	1
	S_{343}	if IN=R_7 ,then OUT= $R_3 \rightarrow R_3$	1
	S_{436}	if IN=$R_3 \rightarrow R_6$,then OUT= $R_2 \rightarrow R_7$	1
	S_{436}	if IN=$R_4 \rightarrow R_2$,then OUT= R_7	1
2	S_{12}	if IN=R_3 ,then OUT= $R_6 \rightarrow R_4$	1
	S_{12}	if IN=R_3 ,then OUT= $R_1 \rightarrow R_6$	1
	S_{13}	if IN=R_6 ,then OUT= $R_3 \rightarrow R_4$	1
	S_{13}	if IN=$R_3 \rightarrow R_8$,then OUT= $R_3 \rightarrow R_7$	1
	S_{14}	if IN=$R_7 \rightarrow R_6$,then OUT= $R_8 \rightarrow R_3$	1
	S_{14}	if IN=$R_4 \rightarrow R_8$,then OUT= R_3	1
	S_{15}	if IN=R_3 ,then OUT= $R_1 \rightarrow R_6$	1
	S_{15}	if IN=$R_7 \rightarrow R_8$,then OUT= $R_4 \rightarrow R_1$	1
	S_{16}	if IN=$R_7 \rightarrow R_3$,then OUT= $R_3 \rightarrow R_6$	1
	S_{16}	if IN=$R_3 \rightarrow R_3$,then OUT= R_6	1
	S_2	if IN=R_3 ,then OUT= $R_3 \rightarrow R_6$	2
	S_{23}	if IN=$R_4 \rightarrow R_8$,then OUT= $R_7 \rightarrow R_3$	1
	S_{23}	if IN=$R_8 \rightarrow R_7$,then OUT= R_3	1
	S_{24}	if IN=$R_3 \rightarrow R_3$,then OUT= $R_1 \rightarrow R_4 \rightarrow R_6$	1
	S_{24}	if IN=$R_3 \rightarrow R_1$,then OUT= $R_4 \rightarrow R_6$	1
	S_{25}	if IN=$R_6 \rightarrow R_1$,then OUT= $R_3 \rightarrow R_2$	1
	S_{25}	if IN=$R_3 \rightarrow R_1 \rightarrow R_6 \rightarrow R_8$	1
	S_{26}	if IN=$R_4 \rightarrow R_1 \rightarrow R_3 \rightarrow R_6$	1
	S_{26}	if IN=$R_6 \rightarrow R_4 \rightarrow R_8 \rightarrow R_5$	1
	S_{34}	if IN=$R_7 \rightarrow R_3 \rightarrow R_3$	1
	S_{34}	if IN=$R_6 \rightarrow R_1 \rightarrow R_3 \rightarrow R_5$	1
	S_{35}	if IN=$R_7 \rightarrow R_3 \rightarrow R_3$	1
	S_{35}	if IN=$R_7 \rightarrow R_3 \rightarrow R_1$	1
	S_{36}	if IN=$R_6 \rightarrow R_3 \rightarrow R_4 \rightarrow R_8$	1
	S_{36}	if IN=$R_3 \rightarrow R_8 \rightarrow R_1 \rightarrow R_6$	1
	S_{45}	if IN=$R_4 \rightarrow R_6 \rightarrow R_1 \rightarrow R_3 \rightarrow R_7$	1
	S_{45}	if IN=$R_4 \rightarrow R_1 \rightarrow R_3 \rightarrow R_7$	1
	S_{46}	if IN=$R_6 \rightarrow R_4 \rightarrow R_2 \rightarrow R_7$	1
	S_{46}	if IN=$R_8 \rightarrow R_2 \rightarrow R_3 \rightarrow R_7 \rightarrow R_6$	1
	S_{56}	if IN=$R_8 \rightarrow R_3$	1
	S_{56}	if IN=$R_4 \rightarrow R_4 \rightarrow R_7 \rightarrow R_6 \rightarrow R_3$	1
	S_1	if IN=R_3 ,then OUT= $R_8 \rightarrow R_4$	2
	S_1	if IN=$R_7 \rightarrow R_1$,then OUT= $R_4 \rightarrow R_3 \rightarrow R_3$	2
	S_7	if IN=R_3 ,then OUT= $R_3 \rightarrow R_1$	2
	S_7	if IN=$R_3 \rightarrow R_2$,then OUT= $R_1 \rightarrow R_4$	2
	S_3	if IN=$R_3 \rightarrow R_1$,then OUT= R_4	2
	S_4	if IN=R_3 ,then OUT= $R_3 \rightarrow R_6$	2
	S_4	if IN=$R_3 \rightarrow R_2$,then OUT= $R_3 \rightarrow R_4 \rightarrow R_7$	1
	S_5	if IN=R_4 ,then OUT= R_1	2
	S_5	if IN=$R_4 \rightarrow R_3$,then OUT= $R_4 \rightarrow R_6$	2
	S_6	if IN=R_4 ,then OUT= $R_8 \rightarrow R_2$	2
	S_6	if IN=$R_4 \rightarrow R_3$,then OUT= $R_4 \rightarrow R_1$	2

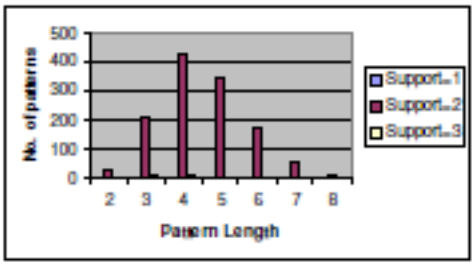


Fig 3: Number of sequential patterns with different pattern length and with the support value

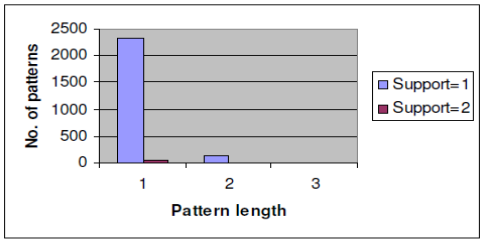


Fig 4: Number of fuzzy rules with different pattern combinations along with the support value

Rules Analysis-The rules define the nature of movement of every tag in terms of EPC and different tag combinations.For instance, in the given Table V, the tags with EPC code, 1, 2 and 5 commonly get into the range of R_1 and R_6 ,when they entered into the range of R_5 with a fuzzy score of 1. Hence, the nature of significant combination of the tags is well understood from the rules that are generated from the proposed mining system.

VI. Conclusion

In this paper, we have proposed a data mining system for mining the information that are relevant to the nature of movement of tags, which are affixed in the warehouse goods. The implementation results have shown that the proposed mining system mines the knowledge from the warehoused data by generating I-dataset, mining sequential patterns and then by generating fuzzy rules from the sequential patterns. The outcome of the system, fuzzy rules, has detailed the nature of the tag movement with a fuzzy score. Given a part of the tag (indirectly it refers to A product) movement, the fuzzy rules hold the remaining path of the tag (product). In this manner, different length combinations of the tags have been considered and the movement has been understood. The movements are not considered for all the tags and its combinations, but only for some significant tags and combinations. Hence, with the aid of the proposed data mining system, the tracking of goods in large warehouses can be performed effectively and also the extracted information may be helpful for the management of warehouse.

VII. References

1) Bin Li and Dennis Shasha, "Free Parallel Data Mining", ACM SIGMOD Record, Vol.27, No.2, pp.541-543, June 1998.

2) Anand, Bell and Hughes, "EDM: A general framework for data mining based on evidence theory", Data and Knowledge Engineering, Vol.18, No.3, pp.189-223, 1996.
3) Agrawal, Imielinsk and Swami, ―Database Mining: A Performance Perspective", IEEE Transaction Knowledge and Data Engineering, vol. 5, no. 6, pp. 914-925, 1993.
4) Chen and Liu, "Data mining from 1994 to 2004: an application-oriented review", International Journal of Business Intelligence and Data Mining, Vol.1, No.1, pp.4-11, 2005.
5) Shankar and Purusothaman, " Utility Sentient Frequent Itemset Mining and Association Rule Mining: A Literature Survey and Comparative Study", International Journal of Soft Computing Applications, Vol.10, No.4, pp.81-95, 2009.
6) Mourad Ykhlef and Hebah ElGibreen, "Mining Sequential Patterns Using Hybrid Evolutionary Algorithm", WASET, Vol.60, No.149, pp.863-870, 2009.
7) Jia-Dong Ren and Xiao-Lei Zhou, "A New Incremental Updating Algorithm for Mining Sequential Patterns",Journal of Computer Science, Vol.2, No.4, pp.318-321, 2006
8) Hye-Chung Kum, Joong Hyuk Chang and Wei Wang, "Benchmarking the effectiveness of sequential pattern mining methods", Data & Knowledge Engineering, Vol.60, pp. 30–50, 2007.
9) Jiawei Han, Hong Cheng, Dong Xin and Xifeng Yan, "Frequent pattern mining: current status and future directions", Data Mining and Knowledge Discovery, Vol.15, No.1, pp.55-86, January 2007.
10) Dunren Che1 and Wei Zheng, "Smart support functions for sequential pattern mining", Journal of Computational Methods in Science and Engineering, Vol.6, No.2, pp.255- 263, 2006.
11) Yashpal Singh and Alok Singh Chauhan, "Neural Networks In Data Mining", Journal of Theoretical and Applied Information Technology, Vol.5, No.6, pp.36- 42,2009.
12) Guangming Wang and Gonglian Jin, "Research and Design of RFID Data Processing Model Based on Complex Event Processing", in Proceedings of International Conference on Computer Science and Software Engineering, Vol.5, pp. 1396-1399, 2008.
13) Dhany Saputra, Dayang R.A.Rambli and Oi Mean Foong, "Mining Sequential Patterns Using I-PrefixSpan", International Journal of Computer Science and Engineering, Vol. 2, No.2, pp.49-554, 2008.
14) C.M. Roberts, "Radio frequency identification (RFID)", Computers & Security, Vol.25,pp. 18 – 26, 2006.
15) Hatim A. Aboalsamh, "A novel Boolean algebraic framework for association and pattern mining", WSEAS Transactions on Computers, Vol.7, No.8, pp.1352-1361,August 2008.
16) Kuen-Fang Jea, Ke-Chung Lin and I-En Liao, "Mining hybrid sequential patterns by hierarchical mining technique", International Journal of Innovative Computing,Information and Control, Vol.5, No.8, August 2009.
17) Sathiyamoorthi and Murali Bhaskaran, "Data Mining for Intelligent Enterprise Resource Planning System", International Journal of Recent Trends in Engineering, Vol. 2, No. 3,pp.1-5, November 2009.
18) Jian Pei,Jiawei Han, Behzad Mortazavi-Asl, Jianyong Wang, Helen Pinto, Qiming Chen, Umeshwar Dayal and Mei-Chun Hsum, "Mining Sequential Patterns by Pattern- Growth: The PrefixSpan Approach", IEEE Transactions on Knowledge and Data Engineering, Vol. 16, No. 10, pp.1-17, October 2004.
19) Agrawal and Srikant, ―Mining Sequential Patterns," in proceedings of International Conference on Data Eng., pp.3-14, Mar. 1995.
20) Garofalakis, Rastogi and Shim, "Mining sequential patterns with regular expression constraints", IEEE Transactions on Knowledge and Data Engineering, Vol.14, No.3, pp.530-552, June 2002.
21) Jayanthi Ranjan and Vishal Bhatnagar, "A Review of Data Mining Tools In Customer Relationship Management", Journal of Knowledge Management Practice, Vol. 9, No. 1, March 2008.
22) Michael J. Shaw, Chandrasekar Subramaniam, Gek Woo Tan and Michael E. Welge, "Knowledge management and data mining for marketing", Decision support systems, Vol.31, No.1, pp.127-137, May 2001.
23) Asghar Sabbaghi and Ganesh Vaidyanathan, ―Effectiveness and Efficiency of RFID technology in Supply Chain Management: Strategic values and Challenges", Journal of Theoretical and Applied Electronic Commerce Research, Vol. 3, No. 2, p.p. 71-81, 2008, ISSN 0718–1876.
24) Lahiri S., RFID source book, IBM press, Upper Saddle River, NJ, 2006.
25) Henning Baars, Hans-Georg Kemper, Heiner Lasi and Marc Siegel, ―Combining RFID Technology and Business Intelligence for Supply Chain Optimization – Scenarios for Retail Logistics", Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008), Waikoloa, Big Island, Hawaii, ISBN: 0-7695-3075-3.
26) Asif, Z., Mandviwalla, M., "Integrating the supply chain with RFID: a technical and business analysis", Communications of the Association for Information Systems, Vol. 15, No. 24, pp.393-427, 2005.
27) Jigyasa Bisaria, Namita Srivastava and Kamal Raj Pardasani, "A Rough Set Model for Sequential Pattern Mining with Constraints", International Journal of Computer and Network Security, Vol.1, no.2, pp.16-22, November 2009.

28) Shigeaki Sakurai, Youichi Kitahara and Ryohei Orihara, "A Sequential Pattern Mining Method based on Sequential Interestingness", International Journal of Computational Intelligence, Vol. 4, No.4, pp.252-260, 2008.
29) Themis P. Exarchos, Markos G. Tsipouras, Costas Papaloukas and Dimitrios I. Fotiadis, "A two-stage methodology for sequence classification based on sequential pattern mining and optimization", Data & Knowledge Engineering, Vol.66, pp.467–487, 2008.
30) Chen, "An Up Down Directed Acyclic Graph Approach for Sequential Pattern Mining", IEEE Transactions on Knowledge and Data Engineering, Vol.21, No.12, pp.1-1, 2009.
31) Jirachai Buddhakulsomsiri and Armen Zakarian, "Sequential pattern mining algorithm for automotive warranty data", Computers and Industrial Engineering, Vol.57, No.1, pp.137-147, August 2009
32) Ya-Han Hu, Tony Cheng-Kui Huang, Hui-Ru Yang and Yen-Liang Chen, "On mining multi-time-interval sequential patterns", Data & Knowledge Engineering, Vol.68, No.10, pp.1112-1127, October 2009
33) Ya-Han Hu, Yen-Liang Chen and Kwei Tang, "Mining sequential patterns in the B2B environment", Journal of Information Science, Vol. 35, No. 6, 677-694 (2009).

Face Recognition for Single and Different Facial Expressions

GJCST Computing Classification
I.5.4 , I.4.6

Rashedur M Rahman[1] Anirban Das[2] M Russel[3] Md. Shazzad Maruf[4]

***Abstract*- This paper presents and analyzes the performance of Principle Component Analysis (PCA) based technique for face recognition. We consider recognition of human faces with two facial expressions: single and differential. The images that are captured previously constitute the training set. From these images eigenspaces/eigenfaces are calculated. The image that is going to be recognized through our system is mapped to the same eigenspaces. Next two classification techniques, namely distance based and neural network based classifier are used to classify the images as recognized or non-recognized. In this research we categorize the face images into six different test cases and discuss the performance of each test case with various performance metrics. Our experimental results demonstrate that the neural network technique outperforms the distance based classifier in most of the test cases.**

Keywords- principal component analysis, neural network, face recognition, image processing.

I. Introduction

Face is an essential element of focus of our daily life. We convey our identity and emotions through our face and different expressions of faces respectively. Though human faces are complex in shape, face recognition is not difficult for a human brain whereas for a computer this job is not easy. The complexity of recognition is prominent and several algorithms are reported in literature [1,5,7,8] that could achieve the recognition with high degree of accuracy. Face recognition system is widely used in different areas that include a) criminal record and identification, b) Robot vision, c) security system, d) human computer interaction, e) image and field processing. Face recognition system is divided into two categories, i) appearance based and ii) component based. For appearance based, we consider the holistic feature or the whole face image as our feature for recognition. On the other hand, in component based face recognition, we consider geometrical relationship of different components of face such as eye, nose, lip etc as the features of a recognition system. Principal Component Analysis (PCA) [7,12] is a fast and efficient technique that is widely used for appearance based face recognition. This technique is also used for dimensionality reduction in different areas that include image processing, signal processing and data mining.

About [1]Rahman, R.M. is with the Department of Electrical Engineering and Computer Science, North South University, Dhaka, Bangladesh (telephone: 8802-8852000-1507 email: rashedur@northsouth.edu) About-[2]Das, A., 3Russel, M, 4Maruf, S. Department of Electrical Engineering and Computer Science, North South University, Dhaka, Bangladesh.

The eigenfaces approach is chosen for this study considering its capability of recognizing real time images where orientation, illumination and distortion are continuously changing. This work focuses on how the images with real time attributes affect the recognition feature of eigenfaces technique. Our primary objective for this research is to minimize the complexity in calculation for bigger matrices. For example, if we have 120 pictures with the size of (180 × 200), we will have a very big number while calculating the one dimensional vector from 2D matrix (by calculating 180 × 200 × 120) which is a very big number. By using the eigenvectors, we could minimize the use of all the images and reduce them for example 40 pictures which will also bring down our total calculation to (180 × 200 × 40). Though, we are using lesser amount of data, we will still get the same level of accuracy. Besides, we could even make the size even smaller by changing the order of matrix multiplication which in turn reduces the principal components, and the end we could work only on (40×120) matrix with the same level of accuracy Rest of the paper is organized as follows. Section 2 describes the methodologies used in this research in detail. Section 3 describes about the system and execution flows of different components in the system. Section 4 presents and analyzes the result. Finally Section 5 concludes and gives direction of future research.

II. Methodologies

Our face recognition system consists of several steps. Each of the steps is described in detail in below:

A. Initialization And Finding Principal Components

At first we take images. These images are nothing but the matrix which has pixel intensity at different rows and columns. This image could be viewed as a vector also. If an image has height, h and width, w, then we could formulate this image as w vectors, where each vector has h dimensions. The rows of the images are placed one after another like the Figure1 below:

Fig 1: Formation of the face's vector from the face's images

The vector which is w×h represents our image and this image has a certain space so this is called image space. If we have N images, we have image space dimension as N× w×h. In this image space all images are represented by w by h pixels. These images under same image space look like each other. They all have two eyes, a nose, a mouth etc located at the same image space.

Now we will build the face space from the image space. The main task of building a face space is to describe the face images. The basis vector of this space is called principal component. The dimension of the face space will be M× w×h. In the face space all pixel is not relevant and each pixel depends on the neighbors. So the dimension of face space is less than the dimension of the image space. We could find the principle components of the faces by finding the eigenvectors of the covariance matrix of the set of face images. This eigenvectors are basically a set of features which characterize to the maximum variations between face images. Each of this images that comes from the image space contribute more of less to the eigenfaces. So we can display eigenvector as a sort of ghostly faces which we call eigenfaces. Actually eigenfaces do not exist in real world. We could not say we can build or create eigenface of a particular image face which is in the image space. Eigen face actually is an imaginary face which is a combination of all the images with in a particular image space. Figure 2 presents eight eigenfaces from a sample image database in Figure 5.

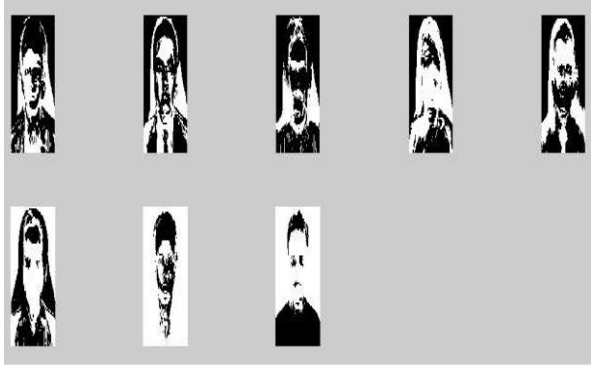


Fig2: Eigenfaces from the image database.

We present the mathematical formulation of eigenfaces below. More details of the formulation could be found elsewhere [5,6].

i. We obtain N training images I1, I2…….IN. Each of these images have dimension w×h. Convert these images into vector space by concatenation. After the concatenation a matrix is converted to a vector. An example of concatenation is given in the following page.

ii. Represent each image I_i with its corresponding vector λ_i .

$$\begin{bmatrix} B_{11} & B_{12} & . & B_{1h} \\ . & . & . & . \\ . & . & . & . \\ B_{w1} & B_{w2} & . & B_{wh} \end{bmatrix} \xrightarrow{\text{Concatenation}} \begin{bmatrix} B_{11} \\ . \\ . \\ B_{1h} \\ . \\ . \\ . \\ B \end{bmatrix} \equiv \lambda_i$$

iii. Calculate the mean face vector ϖ by the following Equation.

$$\varpi \equiv \frac{1}{N}\sum_{i=1}^{N}\lambda_i$$

iv. Subtract the mean face, ϖ from each face vector, λ_i to get a set of vectors, μ_i . The purpose of subtracting the mean image from each image vector is to keep only the distinguishing features from each face by removing the common information.

v. Find the covariance matrix C by the following equation:

$$C = A^T A \text{ where, } A = [\mu_1, \mu_2, \ldots\ldots\mu_N]$$

vi. Find the eigenvalues and eigenvectors for the covariance matrix, C. Sort the eigenvectors according to the eigenvalues. Take the first M eigenvectors that have higher eigenvalues. Now each eigenvector will have N×1 dimension. Let us name those eigenvectors as η_i for i=1,2,…M.

B. Projection Of New Face To Eigenfaces

When a new image is encountered, calculate the set of weights based on the new or input image and the M eigenfaces by projecting the input image onto each of the eigenfaces. The mathematical formulation is given below:

i. Let us consider the new image as I_{new}

ii. Find out the M eigenface components, ψ_l , by projecting the new image

iii.

$$\psi_l = \gamma_l^T (I_{new} - \varpi) \text{ for } l=1,2,\ldots M$$

where,

$$\gamma_l = \sum_{k=1}^{N} \eta_{lk}\mu_k \text{ for } l=1,2,..M$$

iv. Create a new feature vector, Ω_{new} for the new image by concatenating eigenface components, ψ_l

$$\Omega_{new} = [\psi_1 \quad \psi_2 \quad . \quad . \quad \psi_M]$$

C. *Face Recognition By Classification Algorithms*

The last step of the face recognition system is to identify the new face to be recognized or not recognized. If the face is recognized the system will tell the person's name for whom the face has been recognized. In the other word, if we have N persons in the image database, we say that there are N classes where each individual person representing a class. There are two algorithms used for classification, one is distanced based and the other one is neural network based classification.

The distance based classifier works in the following way:

i. For each image in the image database, find out the feature vector Ω_i for N persons where i=1,2,…N . The procedure will be same that is discussed for the new image in the earlier section.

ii. Classification is performed by comparing the feature vector of new image, Ω_{new}, with the feature vector of the images in the image database.

iii. Comparison is done by the Euclidian distance between two features, Ω_{new} and Ω_i, if the distance is less than some predefined threshold, t, we say that the image is recognized.

iv. The class of the new image will be one that has the least Euclidian distance with the new image, providing this distance is less than the threshold.

We also use Neural Network (NN) for classification of the new image to the image database. A fully connected, layered, feed-forward network is depicted in Figure 3, where x_i, h_i, o_i represent unit activation levels of input, hidden, and output units, respectively. Weights on the connections between the input and hidden layers are denoted by $w1_{ij}$, while weights on connections between the hidden and output layers are denoted by $w2_{ij}$. The neurons marked with "1" are threshold neurons and their activation value is set to 1. Figure 3's network has three layers, although it is possible and sometimes useful to have more layers. Each unit in one layer is connected in the forward direction to

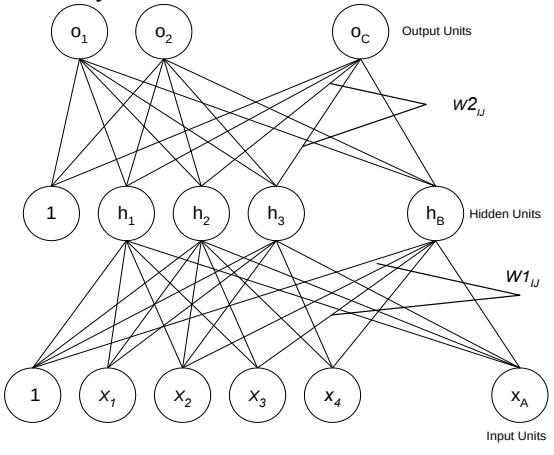


Fig. 3: A Multilayer Feed-Forward Neural Network

every unit in the next layer. Activations flow from the input layer, through the hidden layer, to the output layer. The knowledge of the network is encoded in the weights on connections between units. A backpropagation network typically is initialized with a random set of weights. The network adjusts its weights each time it processes an input-output pair. More details about neural network and its training algorithm could be found elsewhere [13,14].

Our neural network algorithm for classification work in the following way:

i. Randomize the weights to small arbitrary values and initialize the activations neurons. The neurons marked as "1" in Figure 3 are activation neurons. Their values are set to 1 in this step.

ii. Select a training pair from the training set. A training pair consists of an input and output vectors. The input will be the projected feature vector Ω_i, and the output will be the class label. For Ω_i input vector class label will be i. In other word this projected vector represents person i in the image database.

iii. Apply the input vector to the network input neurons.

iv. Propagate the activations from the input neurons of the input layer to the hidden neurons of the hidden layer using the activation function.

v. Propagate the activations from the neurons of the hidden layer to the neurons of the output layer.

vi. Calculate the error, the difference between the network output, and the desired output. The desired output is the output vector from the training pair and the network output is calculated by activation of output neurons. These errors are the errors of the neurons in the output layer.

vii. Compute the errors of the neurons in the hidden layer

viii. Adjust the weights of the network between the hidden layer and output layer.

ix. Adjust the weights between the input layer and the hidden layer. The error adjustment in steps 8-9 use the gradient decent method [13].

x. Repeat Steps 2-9 for each pair of input-output vectors (we have N images so N vectors) in the training set until the error for the entire system is acceptably low.

When the network is trained the weights are adjusted accordingly. If we now present the testing image data I_{new} through its projected vector Ω_{new} to the input unit of the neural network, at most one of the output neurons, for example, i neuron will fire that corresponds to the highest match to the input image. We classify the image to be recognized as person i .

III. SYSTEM DESCRIPTION

We have developed our system by using MATLAB 2008a (version7.6).because we found that the performance of MATLAB 2008a (version 7.6) [4] is better than other programming language. Besides, MATLAB is a high-level

language for technical computing development environment for managing code, files, and data interactive tools for iterative exploration, design, and problem solving. It also supports Mathematical functions for linear algebra, statistics, Fourier analysis, filtering, optimization, and numerical integration 2-D and 3-D graphics functions [2,3,9] for visualizing data. It has also tools for building custom graphical user interfaces and neural network toolbox that we use for classification. For our system we have used certain method and techniques that are offered explicitly by MATLAB. These are as follows:

A. Image Reading

MATLAB can easily read an image and convert the image in a certain matrix. Later on we can use the image matrix for our related work. MATLAB can read an image of 8 bit up to 32 bit .

B. Image conversion from RGB to GRAYSCALE

MATLAB can convert an image from RGB to GRAYSCALE. This computational task can be done by MATLAB command. If RGB image is 34 bit it represent RED for 8 bit, GREEN for 8 bit, BLUE for 8 bit.

C. Image Resize

MATLAB command can be used to resize a certain image in to any size that MATLAB allow.

D. Convert MATRIX to 1 dimensional VECTOR

We can use certain technique to convert a matrix to 1 dimensional vector which helps us to compute the desired out put.

E. Matrix Transpose

In MATLAB we can easily transpose a certain matrix.
In our system, at first of the system take the images. These.

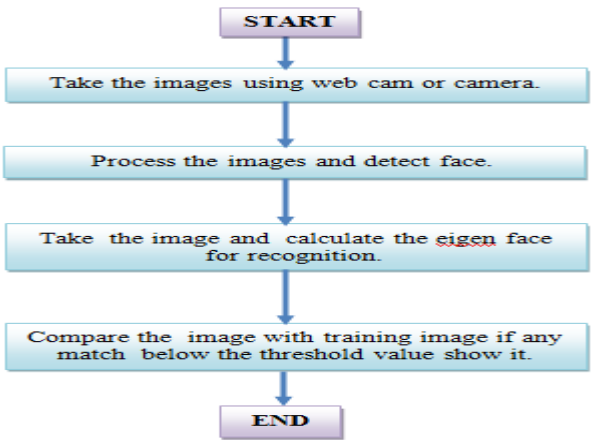


Figure 4: Basic flow diagram of the face recognition system

images are captured by webcam or other image capturing source. We calculate the mean image and eigenfaces Then, we take the input image that will be detected later on. We process the Input image and compare the input image with training image set if any match below the threshold value than we can say it is recognized other wise not. The detail flow of execution is given in Figure 5

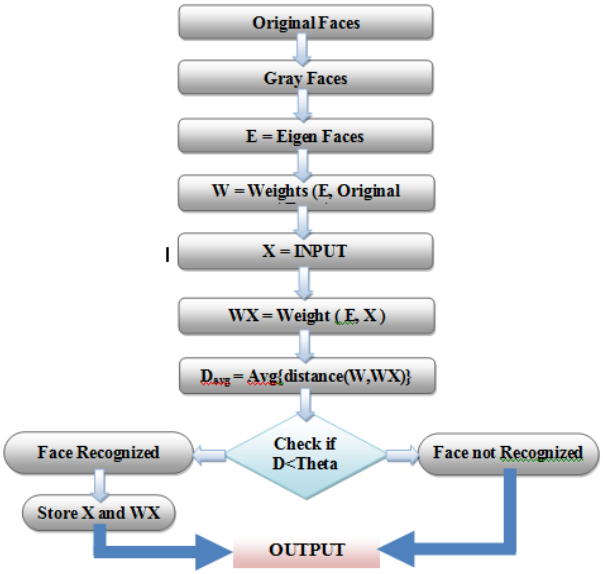


Figure 5: Detail flow of execution in the system

IV. Results and Analysis of Results

In total, six test cases were chosen to test the performance of face recognition system. All the images were in same size. The training set images were captured such that the person faces were frontal with minimal head tilt and had decent varied expressions. Tests were taken in well illuminate area to recognize, at least, the known images. We have to do 2 types of testing

i. Face image with single facial expression.
ii. Face image with different facial expression

Figure 5 and Figure 6 presents some sample image for single and differential facial expressions respectively.

Fig 5. Sample dataset for single facial expression

Fig 6. Sample dataset for differential facial expression

The system performs relatively well in all test cases. We apply our algorithm for six test cases. The test cases are described below:

Test Case 1- Test case 1 is measuring the impact of having large number of training images of a small group of persons. Altogether 30 images of 3 persons (10 images from each) were taken to create the training set. Different facial expression of each person is taken for this case.

Test Case 2- Test case 2 is measuring the impact of having large number of training images from a large group of persons. Altogether 50 images of 50 persons were taken to create the training set. The images are for single facial expression.

Test Case 3- Test case 3 is measuring the impact of having small number of persons in the training set. In total, 6 images from each of the 2 persons were considered to create the training set. So the total number of images is 12.

Test Case 4-Test case 4 is measuring the impact of having small number of images in the training set. In total, 20 images from 20 persons were considered to create the training set.

Test Case 5-Test case 5 is measuring the impact of having very small number of images in the training set. Altogether 3 images of 3 persons (per person image no is 1) were taken to create the training set.

Test Case 6-Test case 6 is measuring the impact of having small number of images in the training set. Altogether 3 images of 10 persons (in total 30 images) were taken to create the training set.

Table 1 summarizes all the test cases and the number of images in total used for recognition system. The accuracy reported here is the accuracy for distance based classifier.

Table 1: Accuracy of Recognition

Test Case	Image/ person	Number of People	Total Images	Total Testing Images	Accuracy
Case 1	3	10	30	50	80%
Case 2	1	50	50	70	100%
Case 3	6	2	12	24	37.5%
Case 4	1	20	20	40	80%
Case 5	1	3	3	5	95%
Case 6	3	10	30	36	83%

Figure 7 compares the accuracy between neural network and distanced based classifier. For most of the test cases, neural network outperforms the distance based classifer.

For the distance based classifier, the Euclidian distance between the new image and all the images in image database are calculated. Then, this distance has to be within two threshold values. The first threshold value is used to screen this image to be a valid face image or not. For example, if the image is for a flower or house then the Euclidian distance will be higher between flower and other facial images in the database.

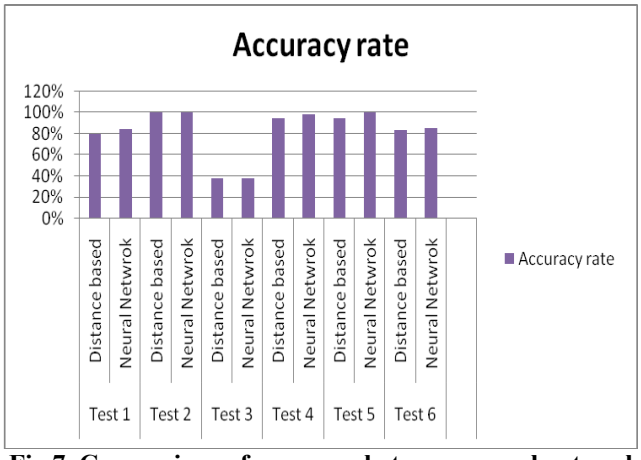


Fig 7. Comparison of accuracy between neural network and distance based classifier.

The system will let user know that this is not a valid face image as the distance is greater than the threshold. The next threshold is used to determine whether the image falls under the images in the database. The second threshold is chosen by trial and error method. Table 2 presents different threshold values and corresponding mismatch rates. If we increase the threshold value we could increase the probability of recognition as a face image. However, we should be careful about not to increase this value to much such that misclassification will increase due to non facial image classification to a facial image.

Table 2: Mismatch vs. Threshold

Mismatch	69%	32%	12%	1%
Threshold	0.3	0.4	0.5	0.6

We also analyze the impact of number of eigenvectors on the accuracy. We select the test case 5 for this purpose. Figure 8 depicts this situation. increasing number of eigenvectors increases the performance of the classifier but up to an extent. If we plot accuracy on accuracy on y axis against the number of eigenvectors on x axis, we will get the following graph.

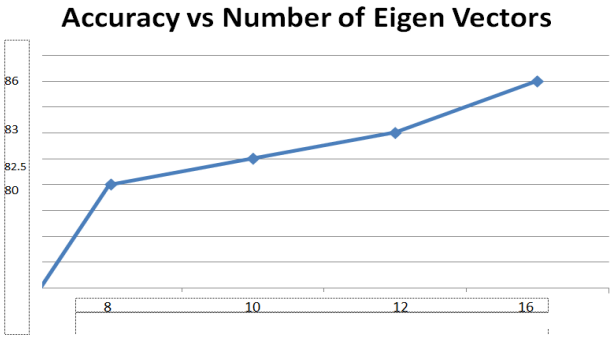


Fig 8: Accuracy vs. Number of eigenvectors.

V. Conclusion

From the tests and analyses performed in this research, we conclude with the following remarks. Training set and test images need to be taken in good, comfortable illumination settings and need to be frontal faces with minimal head tilt. Number of images in the training set is a significant factor. It impacts on defining the correct threshold value for accepting true matches and rejecting false matches. The system performs relatively well with larger training sets and reflects similar behavior irrespective of the number of persons present in those larger training set. However, increase in the training set size increase the system performance and varies the acceptance or rejection rate depending on the person group size. Neural network performs better than the distance based classifier in most of the test cases studied in this research.

VI. References

1) Face Recognition Website. Retrieved on 28th. June, 2010. http://www.face-rec.org/
2) Rafael C. Gonzalez, Richard E. Woods. Digital image processing, 2nd edition. Pearson Education, 2002.
3) Rafael C. Gonzalez, Digital Image Processing by MATLAB, 3rd edition Pearson Education, 2004.
4) MATLAB User guide, Retrieved on 28th. June, 2010. http://www.mathworks.com/access/helpdesk/help/techdoc/.
5) Delac and Mislav Grgic. Face Recognition Edited by Kresimir, I-Tech Education and Publishing, Vienna, Austria, 2007.
6) Santiago Sereno, Master of Science (M.Sc) Thesis, Drexel University, Retrieved on 28th. June, 2010. http://www.pages.drexel.edu/~sis26/Computer%20Vision.htm.
7) Belhumuer P, Hespanha J. Kreigman D, Eigenface Vs Fisherface using class specific linear projection IEEE transaction on pattern analysis and machine intelligence.19 711-720 , 1997.
8) Brunelli R, Poggio T . Face Recognition Features Vs Templates. IEEE transaction on pattern analysis and machine intelligence.15 :1042-1052, 1993.
9) Andrews H.C. Computer Technique in Image Processing, Academic Press, New York.
10) S. Belongie, J. Malik, and C. Fuh. Matching shapes. In 8th International Conference on Computer Vision, volume 1, pages 454-463. IEEE Computer Society Press, July 2001.
11) Bajcsy and A. Kovacic. Multiresolution elastic matching. Computer Graphics and Image Processing 46:1-21, 1989.
12) Baumberg and D. Hogg. An adaptive eigenshape model. In D. Pycock, editor, 6th British Machine Vision Conference, pages 87-96. BMVA Press, Sept. 1995.
13) Ham F. M and Kostanic I. Principles of Neurocomputing for Science and Engineering, McGraw Hill Higher Education, New York, NY, 2001.
14) Fausett, L., Fundamentals of Neural Network: Architecture, Algorithms, and Applications. Prentice Hall, Upper Saddle River, NJ, 1994.

Tsunami Early Warning System Using VIPO

Rudi Sunarno

GJCST Computing Classification
J.2, H.4.3

***Abstract*- Tsunami Early Warning System (TEWS) suddenly popular since the tsunami which hit the city of Banda Aceh and surrounding provinces of Nanggroe Aceh and Pangandaran in West Java and surrounding areas who have swallowed losses of property and lives of so many.Tsunami early warning system using VIPO, as one distribution channel distribution of information, be prepared to assist the government in the Climatology Meteorology and Geophysics Agency (BMKG) in the event of disseminating information to communities around the tsunami disaster areas before the storm arrived.**

VIPO (Virtual Phone) is a tsunami information dissemination system laid by the tsunami of information obtained from BMKG on the radio transmitter and received by a device equipped with an FM radio receiver LCD screen and supports the Subsidiary Communications Authorization (SCA), so information can be received at the terminal VIPO distributed to tsunami-prone communities.

In experiments performed in Pangandaran Beach and its surroundings obtained satisfactory results, where the tsunami disaster information can be sent in a short amount of time to the receiving device, although using a simple system. Field trials have been successfully conducted in Pangandaran Beach and its surrounding areas with excellent results despite the use devices mini transmitters, the main constraint on the soil surface around Pangandaraan hill cause to be obstructed signal reception.

I. PRELIMINARY

Tsunami in Aceh and Pangandaran some time ago leaving a deep wound for the Indonesian nation, especially to victims of disaster. These events not only claimed hundreds to hundreds of thousands of lives, but also leaves a deep anguish for victims who are still alive because they lost a lot of things that not a few even lost entirely from the treasure to the family and relatives.This condition triggered the government and several research institutes from home and abroad to build and develop a system that can detect and provide warning information about the possibility of a potential earthquake or a tsunami caused by earthquakes or other natural events. This is so that people can always be vigilant and take immediate measures to be undertaken that might be done as soon as possible. One study conducted by the Research and Development Center (RDC), PT. Telkom (Telkom-RDC) is to develop a tsunami early warning system or also called virtual phone (VIPO), which laid the information on radio transmitters. This development was done in collaboration between TELKOM-RDC with Climatology Meteorology and Geophysics Agency (BMKG), and the Directorate General of Post and Telecommunications (DG Postel), which facilitated supported also by the Office of Pangandaran STO as the location for placement of radio transmitters and other support systems.

Tsunami warning system is designed with the intention of utilizing VIPO as one disaster information dissemination system, especially the tsunami disaster, to the public directly, where the system is designed for easy to implement, cheap in maintenance costs as well as an affordable device price by the community.

II. SYSTEM OVERVIEW

A. System Description

Tsunami early warning system utilizing VIPO is a system that provides early warning for residents, especially in tsunami-prone areas such as coastal areas, where indicated the possibility of tsunamis arising. This system is connected with the main control system as a detector BMKG tsunami and access system for dissemination of information. In the system developed, the information will be disseminated to the 'ride' the FM radio channel used by the radio broadcaster, known as SCA (subsidiary carrier authorization). With this method, a person who was listening to radio broadcasts, will get information about the earthquake and tsunami hazards there is any possibility in real time (less than 1.5 minutes), after getting information from BMKG system.

VIPO is a terminology that is used to represent a device that functions as a virtual phone to communicate in one direction, whereby the device will be placed in the home and brought the user. VIPO device will have a unique identity number, so that the transmission of information can be directed to a particular identity without acceptable by other devices with different identities. Although there are methods that enable the transmission of information broadcast to all devices simultaneously. Information delivery system will utilize the SCA channel is superimposed on an FM radio broadcast. In this study, the device is used as the receiving device VIPO tsunami early warning.

B. System Configuration

VIPO system connected with the Tsunami disaster information resources available in the office of Climatology Meteorology and Geophysics Agency (BMKG) Indonesia. Tsunami disaster information is then disseminated through various media, including this VIPO system.

Once information is received by the server VIPO the information disseminated in a way modulated the information signal with an FM radio that reaches the target area for dissemination of information carried by the tsunami.

Rudi Sunarno / Pt. Telekomunikasi Indonesia

Furthermore, through the radio terminal will be equipped with modules VIPO will display tsunami information on the LCD screen mounted on VIPO receiving terminals, as shown in the following figures:

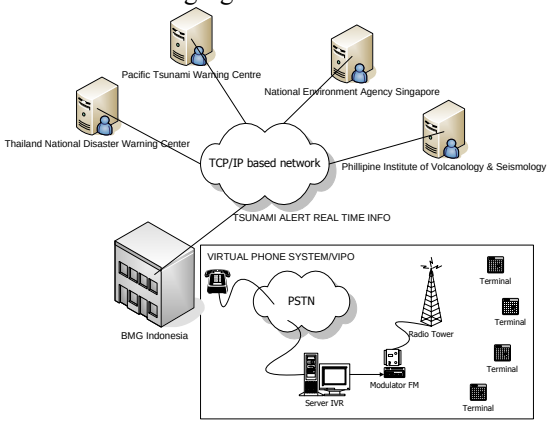


Figure 1. System Configuration VIPO.

Each terminal has a terminal ID is useful for addressing the information to be provided. Information can be directed to the terminal, the terminal groups or terminal depending on the needs of individuals who will be given an information dissemination through VIPO server.

C. Element System

VIPO system consists of two (two) parts, namely the transmitter and the receiver. Transmitter section consists of: SCA Modulator, SCA Generator, and Software Applications. While part of SCA Receiver integrated with the broadcast radio. The function of each part is as follows:

i. The data modulator to modulate the function of existing data on the server VIPO or from the Internet that contains specific information about the tsunami early warning. The appliance is connected to the SCA Injector for transmission to the receiver terminal.

ii. Subsidiary Communications Authorization (SCA) Injector function to generate the carrier frequency to carry data from results of the data unit is injected into the FM modulator for Mixer prior to transmission.

iii. SCA receiver functions to receive siganal SCA to didisplaykan on LED monitors installed in the terminal, even this signal as a triger for siren or loudspeaker.

iv. Software Applications VIPO consisting of: data grabber to "grab" data from online BMKG, Text format converter to convert XML data into binary format and a modem connector as a gateway to the injector through a connection SCA R232. Modem connector serves also set connection and data transmission and error correction.

Figure 2. Transmitter VIPO

Figure 3. Prototype Receiver VIPO

III. PROTOTYPE VIPO

A. Terminal

Terminal created as a common radio channel receiver combined with a device SCA receivers. The terminal is equipped with LCD display for displaying information received and lighting / speaker indicator if there is any information from the incoming BMKG. The terminal also has a specific ID for privacy and powered by AC or DC power input so the terminal can easy and comfortable to carry anywhere.

Figure 4. Recipient VIPO

B. Display

VIPO terminal equipped with a display to show information such as time stamp, magnitude and event or location of earthquake and tsunami potential so that people can be more vigilant.

Figure 5. Recipient VIPO

C. Protocol Information

Protocol information is used as a 'communication' between the transmitter and receiver so that information can be received by the terminal tsunami VIPO as well. Structure of VIPO protocol informations are composed in the 'package' with the following composition:

Figure 6. VIPO Information Protocol

The communication protocol consists of:

i. Start early flags of the packet of information (length 1 byte).
ii. Data length (including the optional params, body message and terminator (FF)).
iii. Optional parameters namely sending information types such as: terminal ID, time stamp, alarm type and validity period.
iv. body of the message is the message content according to information released by BMKG.
v. Terminator i.e. the end of the message packet.
vi. Checksum as a error correction code for the validity of data.

Alarm type parameter can be adjusted in accordance with the type of information selected parameter values are:
no tsunami potential

i. tsunami potential
ii. news
iii. advertising

Examples of message content:
07-04-09 20:29 WIB
5.6 SR 5.6 217km Tenggara
Mentawai Sumbar
Tdk Potential TSUNAMI

IV. FIELD TRIAL

A. Location & Time Trial

Research development of tsunami early warning system using VIPO was implemented and tested starting in January 2008 - December 2009, which specifically for the development of applications made during the nine months from March to November 2009. After going through the stages of research, conducted field trials on Pangandaran beach - Tasikmalaya and surrounding areas. The test was conducted on January, 17-20, 2010.

Results of field trials in addition to getting a good appreciation of the internal team, also from BMKG, DG Postel and Telkom - Pangandaran, where this is proven by their direct participation in the measurements carried out.

Figure 7. Trial Location VIPO

B. Trial Configurations

The test pilot is to determine the performance and test the reliability and coverage of the system to send information to a device used by the user. To transmit data, the team uses a mini transmitters with high power 300 Watt and 40 meter antenna high at 88 MHz frequency. Data is sent every 1 minute by the server that contains information from BMKG about the earthquake and possible tsunami.

C. The Trial Results

With good cooperation among team members and assisted by colleagues TELKOM - Pangandaran, piloting a satisfactory result. Measurements made to more than 10 points from around Pangandaran beach, until the air the farthest distance is 18.59 km which is the location Batu Karas.

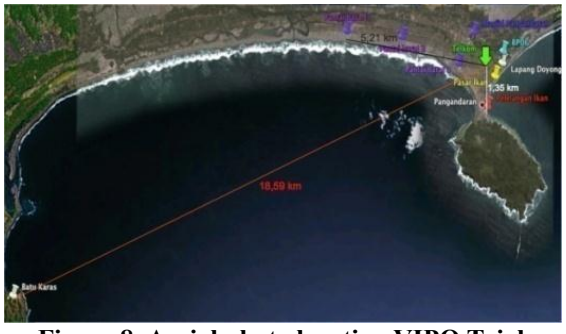


Figure 8. Aerial photo location VIPO Trial

Here is information on testing and test results VIPO:

i. Test Date: 17 - January 19, 2009
ii. Transmitter Frequency: 88.0 Mhz
iii. Transmitter Location: STO Pangandaran
iv. Deconvolution Power: 300 Watt
v. Radius reception: 7 KM (Non-LOS), 43 KM (LOS)
vi. The point of testing:

Pangandaran Beach:

i. Pangandaran highway roundabout mosque
ii. Prawn Breeding Pole Hall (BPUG)
iii. Inclined field
iv. Fish Market
v. Fish Auction

Batu karas

i. Beautiful beaches Batu Karas

D. Analysis Vipo System

From the VIPO trial, was obtained trial data as follows.

No.	Lokasi	Koordinat		Hasil
1	Bundaran mesjid Pangandaran	7° 41′ 02.08″ LS	108° 39′ 14.12″ BT	OK
2	BPUG	7° 41′ 20.22″ LS	108° 39′ 49.14″ BT	OK
3	Lapang Doyong	7° 41′ 30.91″ LS	108° 39′ 48.17″ BT	OK
4	Pasar Ikan	7° 41′ 43.18″ LS	108° 39′ 39.37″ BT	OK
5	Pelelangan Ikan	7° 42′ 11.70″ LS	108° 39′ 30.73″ BT	OK
6	Pantai Indah Batu Karas	7° 45′ 05.49″ LS	108° 30′ 03.23″ BT	OK

Table 1. VIPO trial results

i. Altitude test points ranging from 0.5 - 5 m above sea level so that all points of testing to receive broadcast / alarm information.
ii. Information disturbed the alarm if the level of acceptance in an extreme undulating ground surface in addition to the radio emission is low due to use temporary transmitters.
iii. Antenna heights are relatively quite utnuk reach the entire area due to antenna type used is the Omni so the radio beam evenly.

V. CONCLUSION

i. Tsunami early warning system utilizing VIPO dissemination is one of the ideal system, where information could be reached directly into the community without the need for high maintenance operating costs.
ii. Fore it is possible not only for tsunami early warning information that is sent, but can add other information such as weather information and other disasters.
iii. And not only that, the news information is also expected to be sent to enrich the information for the community., Particularly those living in remote coastal locations.
iv. Shipping signal tsunami information dissemination through the SCA did not affect the radio signals but currently not all radio stations have SCA input channel so that not all stations can provide this service.
v. SCA input dedicated to a radio station frequency and must remain, this may interfere with the terminal owner who wants to move the broadcast.

.

A Qos-Aware Context Construction And Discovery For Mobile Context Services In Next Generation Networks

GJCST Computing Classification
C.1.3 ,C.2.2, C.2,m

[1]R.Nandhagopal and [2]C.Chandrasekar

***Abstract*- The significant factors that contribute to the success of the mobile service delivery are the Quality of Service (QoS) and the mobile connectivity. The mobile service infrastructure must be QoS-aware in addition to context-aware. In this paper, we propose to develop a QoS-aware mobile service infrastructure for context construction and discovery. The choice of connectivity and adaptation of application protocol parameters can be intelligently made on the basis of the available information about these offered-QoS. We implement a QoS-predictions server as a part of the QoS and context aware service infrastructure to attain the objectives. In addition, we develop a QoS aware Context Construction protocol. As per the requirements of a specific mobile host mentioned in the context definition, for a set of mobile hosts which are parts of the context defined, a cost effective routing tree is constructed and maintained dynamically. In the given context, only the hosts are employed to carry out Context-sensitive operations through a cooperative effort. By simulation results, we show that our QoS-aware architecture attains ore throughput with low delay in acquiring the service.**

I. INTRODUCTION

A. *Providing Services In Mobile Devices*

Networks, the developing field we can also say it as blooming field with a boom. As we know the network may be a fixed or wireless, both these networks network together aiming high speed of wired networks and the wider coverage of wireless networks. The development of hardware and protocols of wireless networks pushes higher-level applications and lift the services in integrated networks to a hike mainly when wireless LAN (WLAN) technology becomes increasingly popular for providing IP connectivity and 3G is undergoing deployment stage.With regards to computing power, periphery and storage space, mobile devices are restricted. Hence, they depend on the external sources, just to acquire information or functionality. Static connectivity to resources of particular type won‘t sound reasonable to the mobile devices which operate in changing and dynamic environments. Hence there are possibilities for suitable resource connection at runtime, dynamically and automatically, which are transparent to the user. This can be done by service oriented computing. This approach provides functionalities as stand-alone services. These stand alone services are illustrated by a service offer, it is then published and are automatically discovered and selected by comparing the service request with the existing offers [1].

IT industry usually switches from making products to providing services to provide more flexibility and adaptability. The participants of a service are service provider and consumer. There will be a request from the customer for the service to be provided. If the interface of the service is public, an entity can request for a service. A single entity can provide service to rest of the entities by combining the numerous services. The architecture which relies on the services is known as Service Oriented Architecture (SOA).

The following tasks are performed by the service industry, presently [2]

Service description-Description of the service from the service developer point of view and from the service consumer point of view.

Service discovery-Discovery of new services, matching the user‘s needs.

Service monitoring-Monitoring service behavior, in order to ensure that it matches service specification

Service composition-composition a new service on the basis of several existing services in order to create a new service

B. *Mobile Context Services*

The context is the information which may be clearly or absolutely obtained. It can be used to distinguish certain aspects involved in a specific application or network service of an entity. A person, a place, a router, a 3G network gateway, a physical link, or a virtual object such as IP sec tunnel, SNMP agent anything in the list may be called as an entity. In order to respond to the highly changing computational environments a context-aware service should be more flexible and self-directed.

The traditional services are delivered by the mobile devices such as mobile phones or PDAs. When there is a change in the mobile user‘s requirements, the mobile services can be modified unambiguously. A context-aware mobile service adapts the current situation of the user. The main intent of the context-aware service is to provide right service at the right time to the user. It is an independent service such that it does not require too much of interaction with a computing device, to help the user.

[1]*R.Nandhagopal -Assistant Professor & Head, Department of Computer Applications, Annai Mathammal Sheela Engineering College, Namakkal-637013, Tamil Nadu, India. (Email - nandhagopal67@gmail.com)*
[2]*C.Chandrasekar*- Associate Professor, Department of Computer Science, Periyar University, Salem – 636006, Tamil Nadu, India.
(Email -ccsekar@gmail.com)

There are various types of devices which may range from a wall display to mobile phones. They must be provided with mobile and context-aware services. There are also many variety in their resources like screen size, memory etc. We must also know the services that adapt the host of the resources. The services that are provided should know the user‘s current context and should adjust themselves when the context changes.

C. Challenges Of Mobile Context Services

Suppose a user on his travel needs to send his work through GPRS mobile to his workplace. Of course it is impossible due to the lack of bandwidth of the mobile network. So the user will be in need of a service which can detect a faster wireless network and dynamically change the connection to that network.

In heterogeneous networking environment different providers provide different wireless ok and wired networks. Usually mobile service works in a similar kind of environment. Therefore, a part of the ‗end-to-end‘ communication path between a mobile user and an application server is the responsibility of every network.

Quality of Service (QoS) and the mobile connectivity are the two main significant factors that pay for the success of service delivery. The success can be measured by satisfying the QoS of the user. The current mobile services, will assume that the required-QoS to be static, without bothering the changes in user‘s context, during the service delivery [3].

D. Proposed Work

We propose to develop a mobile service infrastructure with the following objectives:

i. The mobile service infrastructure must be QoS-aware in addition to context-aware. (i.e.) aware of the user‘s required-QoS and the QoS offered by the various networks in user‘s context (location, time, etc).
ii. The choice of connectivity and adaptation of application protocol parameters can be intelligently made on the basis of the available information about these offered-QoS.

II. EXISTING WORK

Iris Braun, et al,. have proposed an approach for context and QoS aware discovery namely ConQo. Their approach is the enhancement of semantic service discovery and selection by taking quality of service and contextual information into the account [4].

Shudong Chen, et al,. have presented a context-aware resource management middleware namely VICSDA, for service oriented applications which aims to handle the inherent dynamics of services and the network [5].

Pravin Pawar, et al,. have presented a Mobile Service platform middleware which enables the resource constrained, handheld mobile devices to offer the hosted Nomadic Mobile services in the Internet based on the principles of service Oriented Architecture [6].

Kun Yang and Alex Galis, proposed an all-policy based context-aware service method for next generation networks (NGN).In their approach a thorough consistency was expected to be achieved where policies are well planned to cover from context representation through services down to the underlying networks [7].

Katarzyna Wac, focused on a QoS and context-aware service infrastructure for supporting the development of mobile applications in a heterogeneous network environment. Also they argued that the use of context information helps to better capture the user's required QoS and also improves the delivered QoS [8].

Youngkon Lee [9] has presented the design principle for incorporating quality management on Web service registry developed in UDDI specification and Web service quality management system (WSQMS). They adopted the WSQDL (Web Service Quality Description Language), for representing Web service quality information which is published by WSQM technical committee in OASIS. For quality data and to modify the necessary data structure of the registry, they have also presented the scheme to compose the classification scheme.

Licia Capra, Stefanos Zachariadis and Cecilia Mascolo [10] have presented Q-CAD, a resource discovery framework. In order to discover and select the best satisfying resources which the user requires, it takes the current execution context and quality-of service (QoS} requirements into account which enables persistent computing applications. Only those suitable to the current execution context of the application will be considered and hence the available resources are screened initially. Against the QoS needs of the application the shortlisted resources are evaluated and a binding is established to the best available.

Jose Antonio Parejo, Pablo Fernandez, Antonio Ruiz Cort´es [11] have addressed the optimal QoS-aware selection in composite web services. They have proposed metaheuristic based algorithms such as hybrid genetic algorithm and tabu search.

Rashid J. Al-Ali¤, Omer F. Rana and David W. Walker [12] have extended the service abstraction in the Open Grid Services Architecture for Quality of Service (QoS) properties. Advance or on-demand reservation of resources varies in type and implementation, and independently controlled and monitored. Based on the particular QoS properties, their focus is on the application layer where a given service may indicate the QoS properties it can offer, or where a service may search for other services.

III. SYSTEM REQUIREMENTS

A. Requirements On A Qos Aware Mobile Service Infrastructure

In a context-aware service infrastructure, in order to implement QoS-awareness, the infrastructure should satisfy the following set of requirements

i. The context-aware mapping of user's QoS requirements into the requirements on end-to end offered-QoS
ii. Context-aware service infrastructure should support
   a) User's QoS requirements specification
   b) The service delivery adaptation to the end-to-end offered-QoS, to satisfy the user's required-QoS.
   c) QoS monitoring and real-time measurements of end-to-end offered-QoS and record those measurements to the QoS-predictions server.

In heterogeneous networking environment, the QoS-predictions server is responsible for the predictions of the end-to-end QoS offered.
The QoS-predictions server has these functional requirements

i. The server should obtain and combine the results of real-time end-to-end offered-QoS measurements performed by the mobile users. The quality of the obtained information is estimated by the server.
ii. The obtained data should be transformed into a meaningful form by the server for gathering further QoS information.
iii. The QoS-predictions server must provide generic predictions of end-to-end offered QoS to mobile users. This can be done with the estimation of the quality of the prediction

B. *Protocol Fundamentals*

In QACCD, QPS specifies the operating context. The details of the hosts in the context are not necessary to be known earlier. Hence, the computation of the context can operate in a purely distributed fashion. In which the requests are responded by sending the reply in the same path in which it came. QACCD is also on-demand process in which, when a QPS send a request then only a shortest path tree is built. Along with this message, the context specification and the information necessary for its computation are embedded.
The Components of a request

Request, req

Req_QPS_id	*QoS-predictions server's id*
Req_seq_num	*application sequence number of req*
Req_forwarder	*sender of this copy of req*
Req_QPS_dist	*distance from the QPS to Req_forwarder*
Req_client_dist	*distance from the QPS to the requesting host*
Req_bound	*cost function bound*
Req_cost	*cost function*
Req_data	*application level data associated with this request*

i. Protocol State Information

Client_id	*host's unique identifier*
Seq_num	*application sequence number*
Dist	*distance from QPS*
Parent	*host's parent in the tree*
Parent_dist	*parent's distance (or cost) from QPS*
bound	*cost function bound*
Cost	*cost function*
P	*set of connected neighbors*
P_s	*a subset of P containing the connected neighbors*

The list of all connected neighbors of a client is included in set P. From the client, each neighbor has a link to it and the weight of that link is stored in P. It is referred to as Wp for some p Є P. If a client receives a request from any nodes in p then it would give it a cost distp < bound. Thus it does not use as its shortest path and it also remembers p's cost in P. In finding a new shortest path quickly and to replace a invalid path, this information is much useful.

ii. Context Building

In a request, the protocol maintains no global state instead all the information which are essential arrives in it for computing a context. The context specification and the request are bundled in a request from the application and forwarded to all its neighbors. The Req_bound and Req_cost together define the context. When a request arrives at a server, it includes these and also the cost to this server.
Before sending the request to a node the sending node calculates the destination node's cost hence it is guaranteed to be within the context's bound. The messages are sent to the neighbors that fall within that bound. Except for the request that offer a lower cost path, successive copies of the same request are ignored.
The following algorithm is used by a server to build the context tree, when a request is received.

Algorithm

```
1. If a request req is received then
   1.1 if req. Seq_num = Seq_num + 1 then
          1.1.1 Cost = Req_cost,
                bound = Req_bound
          1.1.2 remove P
          1.1.3 dist = Req_client_dist,
                Parent = Req_forwarder,
                Parent_dist = Req_QPS_dist
       1.1.4 for each p
                1.1.4.1 if Cost(dist,wp) < bound then
                        send the request to p
                     Req_client_dist = Cost(dist,wp)
                     Req_forwarder = Client_id
                     update Ps
                1.1.4.2 End if
          1.1.5 End for
       1.1.6. process the data message of the request
       1.1.7   Seq_num = Req_seq_num
1.2. else if Req_client_dist < dist   then
          1.2.1 dist := Req_client_dist,
        Parent := Req_forwarder,
       Parent_dist := Req_QPS_dist)
          1.2.2 repeat from 1.1.4 to 1.1.5
1.3 end if
2.dist Req_forwarder:= Req_QPS_dist
```

In QACCD, when a shorter cost path is found, the cost of the new path, the new parent, and the new parent's cost are all stored in QPS. The request is broadcasted to non-parent neighbors also. The distance of these non-parent neighbors will maintain them inside the context.

Let the non-parent neighbor be np, QPS applies the cost function to its own distance and the weight of the link to np. QPS broadcasts the request to np if the cost of np is less than the bound. This shorter path may allow additional downstream hence a node should broadcast a request with a lower cost even if its application has already processed it from a previous parent. QPS adds the information about the parent to the set P after receiving any request.

When QPS receives a request which is not seen before, the application automatically processes it not considering whether or not it arrived on the currently stored shortest path.

It is possible that the path via the parent may not exists hence QPS need not wait for more additional copies of a request to arrive only from its parent. When QPS receives a new request and if the path does still exist and is still the shortest path, then the request will ultimately come along that path. This may cause the cost to be updated and the effects to be passed to the children. The QPS host sends the data portion of the request to the application for processing.

a. QoS-Predictions Server (QPS)

Consider there are n nodes in which n1 nodes are clients and n2 nodes are servers. Assume that each server contains m services of which k of them with same service. The QoS-predictions server (QPS) maintains the following structures which contain the information of all the clients, servers and the services which are running currently.

The client structure includes

i. client_id
ii. bw (bandwidth)
iii. power
iv. speed
v. capacity

The server structure includes

i. server_Id
ii. servno (Service number)
iii. req_bw (Required Bandwidth)
iv. req_power (Required Power)
v. req_speed (Required Speed)

The service structure includes

i. servno (Service number)
ii. desc (Description)
iii. lifetime

A service request from the client will be of the following form

serv_req (client_id, servno)

Algorithm

1. Client C1 sends serv_req (client_id, servno) to QPS
2. For each server {Si}, (i=1, 2…N2)
2.1 If (servno = Si.servno) Then
3.1.1 Add Si into the set M.
2.2 End If
3. End For
4. If M <> NULL Then
4.1 For each server {Sk}, (k=1,2…M)
4.1.1 If (C1.bw >= Sk.req_bw) and
(C1.power >= Sk.req_power) and
(C1.speed >= Sk.req_speed) Then
4.1.1.1 return the server Sk as the QOS aware server
4.1.2 End If
4.2 End For
5. End If
6. QPS forward the req (client_id, servno) to Sk
7 Sk processes the request and sends reply to C1.

The client sends a service request with its id and service number to QPS. Then QPS searches the service number in the servers. QPS then add the matching servers in a set M. Then QPS find the QOS aware server with matching QOS constraints bandwidth, power and speed. Then QPS forwards the request to QOS aware server. Finally QOS aware server process the service request and sends reply to client.

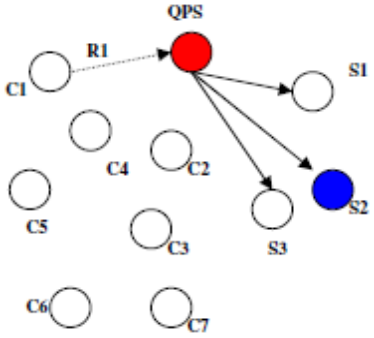


Fig 1.a

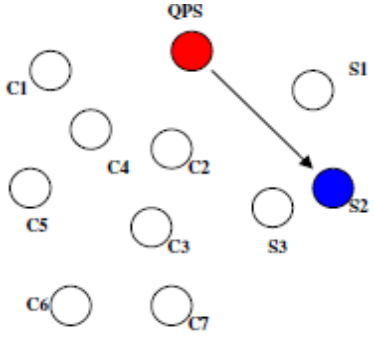


Fig 1.b

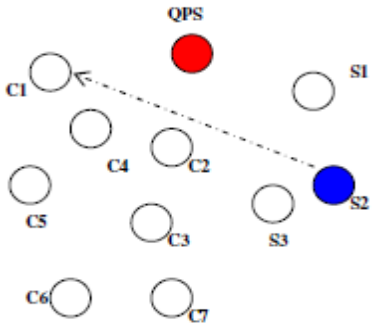


Fig 1.c

In diagram 1.a, the client c1 sends a service request to QPS. QPS forwards the request to all the servers (s1, s2, s3). In diagram 1.b, the service request will be forwarded to QOS aware server s2. In diagram 1.c, the server s2 send reply to client c1.

IV. Simulation Results

A. Simulation Model And Parameters

We use NS2 to simulate our proposed protocol. In our simulation, the channel capacity of mobile hosts is set to the same value: 2 Mbps. We use the distributed coordination function (DCF) of IEEE 802.11 for wireless LANs as the MAC layer protocol. It has the functionality to notify the network layer about link breakage

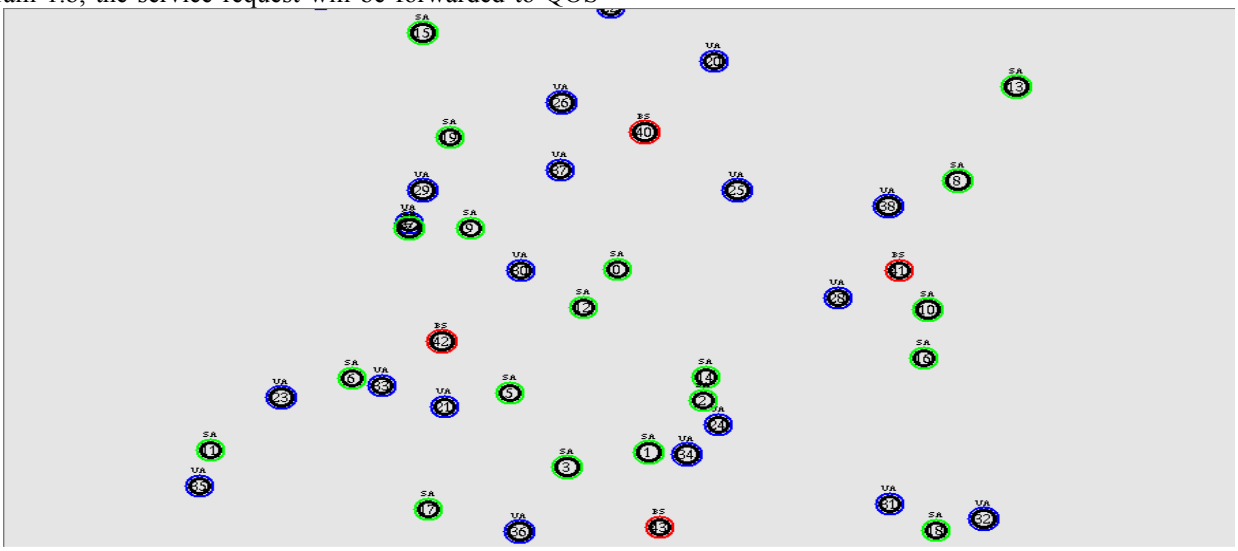


Fig 2: Network Topology

Figure 1 gives the sample network topology used in our simulation. In our simulation, 40 mobile nodes move in a 1000 meter x 1000 meter region for 100 seconds simulation time. Among the total 40 nodes, we treat 20 nodes as clients and 20 nodes as servers. There are 4 base stations to handle the queries of the clients.

We assume each node moves independently with the same average speed. All nodes have the same transmission range of 250 meters. In our simulation, the speed is 10 m/s. We have taken the Service Location Protocol (SLP) for service discovery. A SLP service agent is attached to the servers for providing the services and SLP user agent is attached to the clients for requesting the service. We have used the network abstraction model for the routing process.In our simulation, 4 clients send service requests to the server through the base station. Our simulation settings and parameters are summarized in table 1

Table1: Simulation Parameters

No. of Clients	20
No. of Servers	20
Base stations	4
Area Size	1000 X 1000
Mac	802.11
Radio Range	250m
Simulation Time	100 sec
Service Discovery Protocol	SLP
Server Application	SLPsa
Client Application	SLPua
Speed	10m/s
clients	1,2,3 and 4
Routing Protocol	NETABS

B. Performance Metrics

We compare our QoS-aware service infrastructure with a non-QoS aware service infrastructure. We evaluate mainly the performance according to the following metrics.

Average Delay: It is measured as the average delay occurred for each client while getting the requested service.

Average Throughput: It is measured as the throughput for each client in terms of Mb/sec.

C. Results

i. Based On Rate

In our first experiment we vary the requested traffic rate as 100 to 500 kb and measure the throughput and delay for each client.

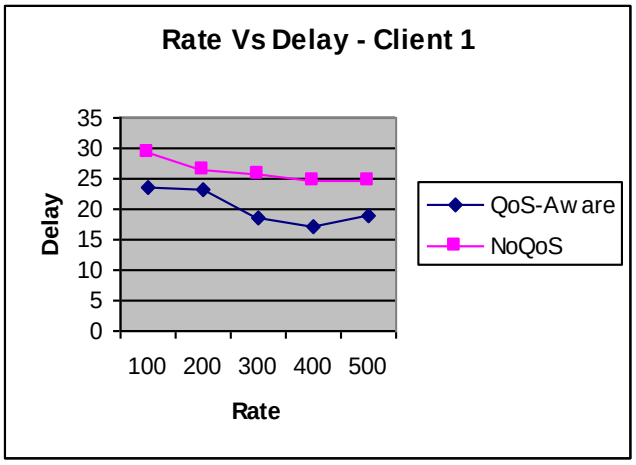


Fig 3: Rate Vs Delay – Client 1

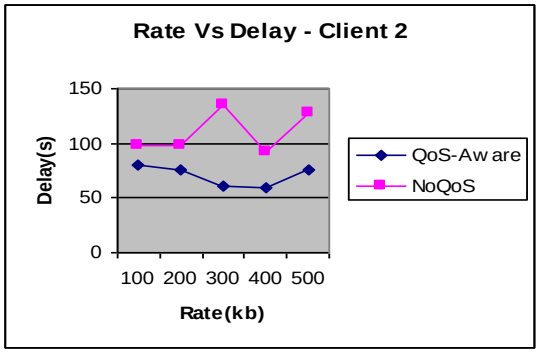


Fig 4: Rate Vs Delay – Client 2

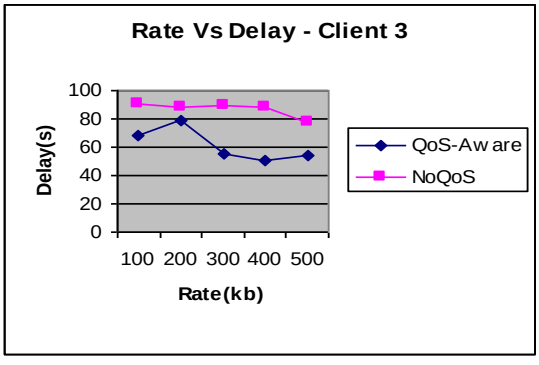


Fig 5: Rate Vs Delay – Client 3

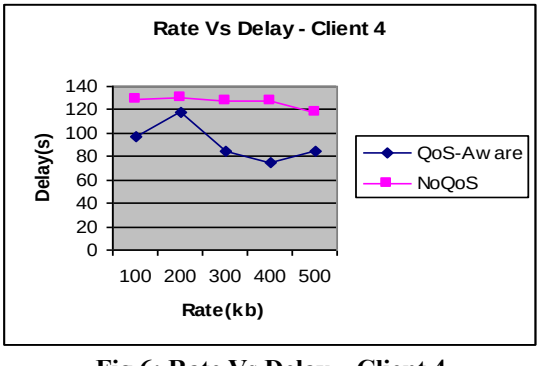


Fig 6: Rate Vs Delay – Client 4

Figures 3 to 6 show the delay occurred for the clients 1 to 4 using the QoS-Aware and NonQos-Aware schemes respectively. As we can see from the figure, the delay is significantly less for QoS-Aware, when compared to NonQos-Aware.

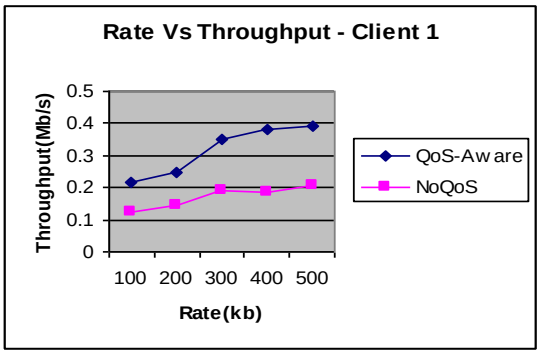


Fig 7: Rate Vs Throughput – Client 1

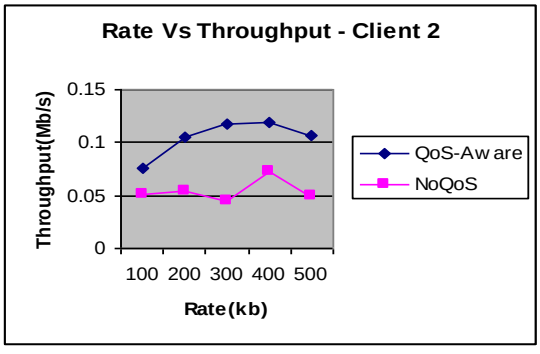


Fig 8: Rate Vs Throughput – Client 2

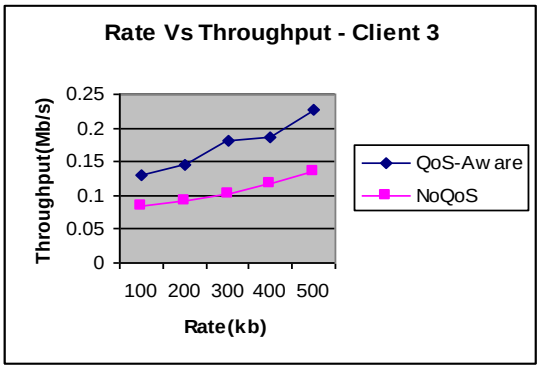


Fig 9: Rate Vs Throughput – Client 3

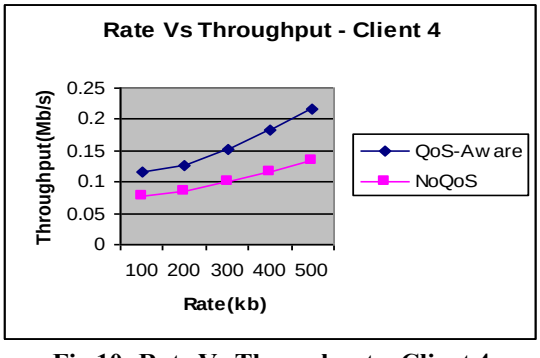


Fig 10: Rate Vs Throughput – Client 4

Figures 7 to 10 show the throughput achieved by the clients 1 to 4 using the QoS-Aware and NonQos-Aware schemes respectively. As we can see from the figure, the throughput

is significantly high for QoS-Aware, when compared to NonQos-Aware.

ii. Packet Size

In our second experiment we vary the requested query size as 100 to 500 bytes and measure the throughput and delay for each client

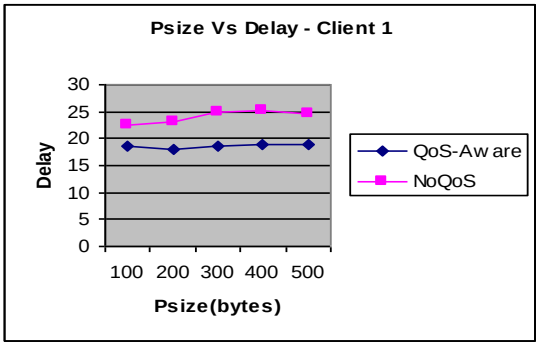


Fig 11: Psize Vs Delay – Client 1

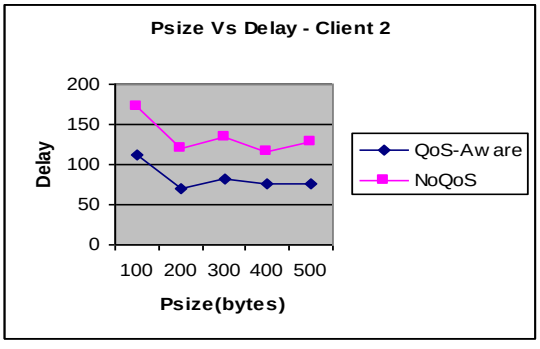


Fig 12: Psize Vs Delay – Client 2

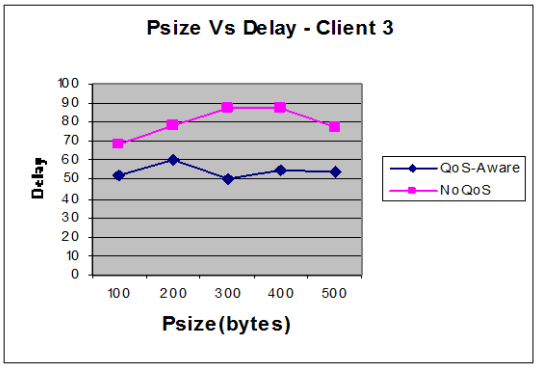


Fig 13: Psize Vs Delay – Client 3

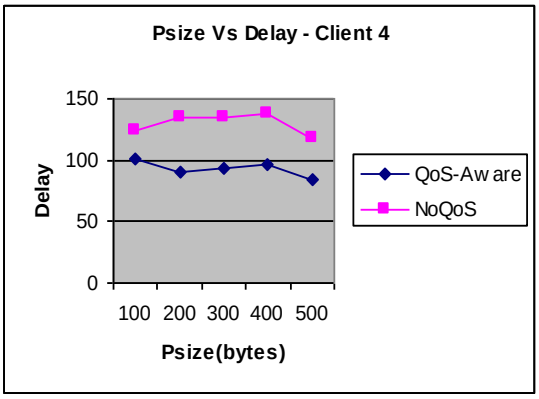


Fig 14: Psize Vs Delay – Client 4

Figures 11 to 14 show the delay occurred for the clients 1 to 4 using the QoS-Aware and NonQos-Aware schemes respectively. As we can see from the figure, the delay is significantly less for QoS-Aware, when compared to NonQos-Aware.

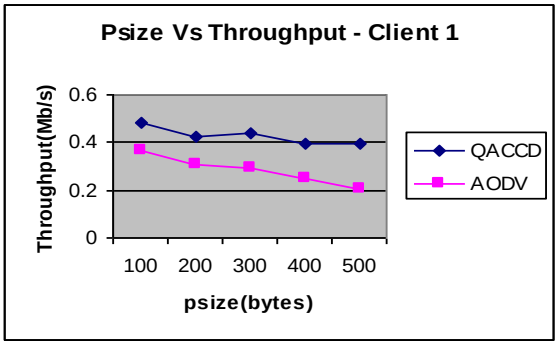


Fig 15: Psize Vs Throughput – Client 1

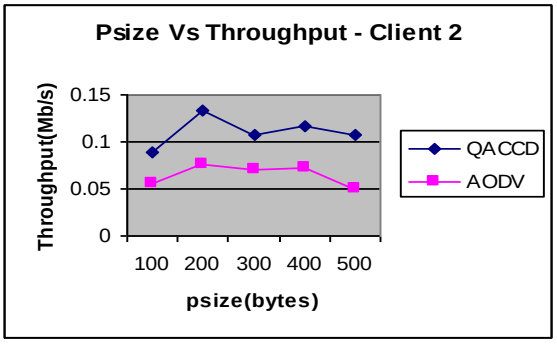


Fig 16: Psize Vs Throughput – Client 2

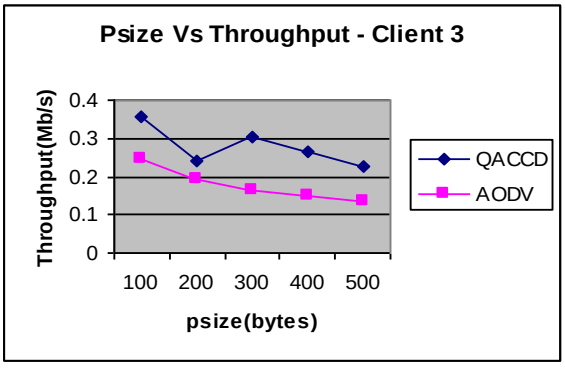


Fig 17: Psize Vs Throughput – Client 3

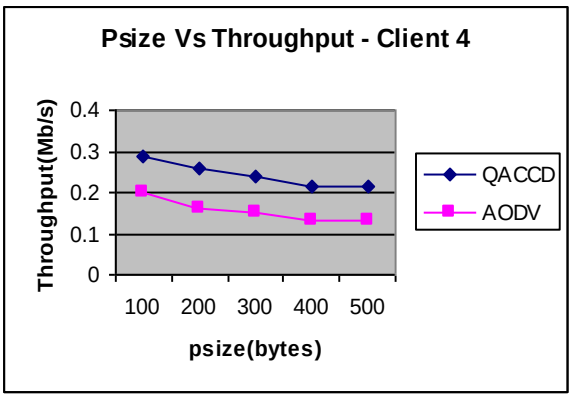


Fig 18: Psize Vs Throughput – Client 4

Figures 15 to 18 show the throughput achieved by the clients 1 to 4 using the QoS-Aware and NonQoS-Aware schemes respectively. As we can see from the figure, the throughput is significantly high for QoS-Aware, when compared to NonQos-Aware.

V. Conclusion

In this paper, we have proposed a mobile service infrastructure which is a QoS-aware in addition to context-aware. The choice of connectivity and adaptation of application protocol parameters can be intelligently made on the basis of the available information about these offered-QoS. We have implemented *QoS-predictions Scheme* as a part of the QoS and context aware service infrastructure to attain the objectives. On the basis of historical end-to-end offered-QoS, generic predictions on the offered-QoS are provided by this scheme. In addition, we have developed a QoS aware Context Construction protocol. As per the requirements of a specific mobile host mentioned in the context definition, for a set of mobile hosts which are parts of the context defined, a cost effective routing tree is constructed and maintained dynamically. In the given context, only the hosts are employed to carry out Context-sensitive operations through a cooperative effort. The source node only needs to broadcast the message over the nodes in the routing tree in order to send a message to only the members of the context. By simulation results, we have shown that our proposed architecture has less service delay and attains more throughput.

VI. References

1) Friederike Klan," Context-aware service discovery, selection and usage", in proc. of 18th GI-Workshop on the Foundations of Databases, pp: 95-99, June 2006
2) Julien Pauty, Davy Preuveneers, Peter Rigole, Yolande Berbers, "Research Challenges in Mobile and Context-Aware Service Development", in proc. of workshop on Research Challenges in Mobile and Context-Aware Service Development,pp:141-148,April 2006
3) Katarzyna Wac, Aart van Halteren, and Dimitri Konstantas,"QoS-Predictions Service: Infrastructural Support for Proactive QoS- and Context-Aware Mobile Services", in proc. of Workshop on Context Aware Mobile Systems, vol.4278, pp: 1924-1933, November 2006
4) Iris Braun, Anja Strunk, Gergana Stoyanova, Bastian Buder,"ConQo – A Context- And QoS-Aware Service Discovery", in Proceedings of WWW/Internet, 2008
5) Shudong Chen, Johan Lukkien, Richard Verhoeven, Pim Vullers, Goran Petrovic,"Context-Aware Resource Management for End-to-End QoS Provision in Service Oriented Applications", in proc. of workshop on Service Discovery and Composition in Ubiquitous and Pervasive Environments,pp:1-6, Nov.30-Dec. 4 2008
6) Pravin Pawar, Bert-Jan van Beijnum, —Context-Aware Mobile Service Platform for Vertical Handover Support to Nomadic Mobile Services", in proc Of ACM SIGMOBILE 9th Workshop on Mobile Computing Systems and Applications (HotMobile 2008), February 2008
7) Kun Yang, Alex Galis, Chris Todd," Policy-driven Mobile Agents for Context-aware Service in Next Generation Networks", in proceedings of International Workshop on Mobile agents for Telecommunication Applications, vol.2881, pp:111-120, October 2003
8) Katarzyna Wac,"Towards QoS-awareness of Context-aware Mobile Applications and Services", in proc. of OTM Workshops, ph.D Symposium, 2005
9) Youngkon Lee," QoS Management for SOA by Synchronizing Quality Context in UDDI", International Journal of u- and e- Service, Science and Technology, June 2009
10) Licia Capra, Stefanos Zachariadis and Cecilia Mascolo," Q-CAD: QoS and Context Aware Discovery Framework for Mobile Systems", IEEE, 2005
11) Jose Antonio Parejo, Pablo Fernandez, Antonio Ruiz Cort´es," QoS-Aware Services composition using Tabu Search and Hybrid Genetic Algorithms", SISTEDES, 2008
12) Rashid J. Al-Ali¤, Omer F. Rana and David W. Walker," G-QoSM: Grid Service Discovery Using QoS Properties"
13) Le-Hung Vu, Fabio Porto, Karl Aberer, Manfred Hauswirth," An Extensible and Personalized Approach to QoS-enabled Service Discovery"
14) Dongyan Xu, Klara Nahrstedt, Duangdao Wichadakul," QoS-Aware Discovery of Wide-Area Distributed Services"

Evaluating Erp Usage Behavior Of Employees And Its Impact On Their Performance: A Case Of Telecom Sector

GJCST Computing Classification
H.4.3 , J.1 , H.3.4

[1]Saima Kanwal,[2]Dr Irfan Anjum Manarvi

Abstract- **Acceptance and use of an information system always remained an important concern for practitioners and theorist. This study explores the factor affecting user behavior toward use of ERP system in developed as well as developing country. Performance expectancy, effort expectancy, social influence, facilitating condition, project communication, training, top management support and self efficacy were found frequently used factors to investigate the ERP usage behavior of employees of a telecom, manufacturing, oil and gas, engineering and government service sector. industry. This study also investigates the impact of behavior toward ERP system usage on the employees' performance.**

A questionnaire based survey was administered personally on 300 employees of telecom using ERP in sector In response to the survey, 255 valid responses were received. The response rate was 85%. Among the respondents, 71% were male while 29% were female.

The finding indicates that the proposed model over all explains 71% variation in the behavior. While Performance Expectancy, Self Efficacy, Training, and Top Management Support were significant factor while explaining the user behavior toward use of ERP system. Further, the model explains that there is difference about opinion of performance expectancy and effort expectancy related to ERP usage by male and female while effort expectancy related to ERP usage also vary among age and experience groups. Mostly respondents agreed that they likey the idea of usinf ERP syste and they want to become professional in ERP system. Seventy percent respondents say that ERP system has improve their performance in work. behavior toward use of ERP system This research will help the telecom organization management to understand the factors responsible for the employee's.

Keywords- ERP, Telecom Industry, ERP usage behavior, Technology Acceptance.

I. INTRODUCTION

Information and Communication Technology (ICT) refers to electronic computer based technology where information is accessed and used in electronic format using computers (Angello & Wema, 2010). Accurate and timely information in electronic form is a precious asset for an organization.

Planning and decision making functions are highly dependent on these information.

With the growth in use of Information Technology by organizations, user acceptance of information system has becomes an important management issue for better management of resources and needs.Business management system integrates all business functions. Enterprise resource planning (ERP) is one of business management systems provide for centralized business processes and functions in a company. ERP implementation requires extensive financial and personnel resources. In some cases, organization fails to attain the benefits from the system. One common reason for its failures is the reluctance or unwillingness of users to accept ERP system. Therefore, a good understanding of end users acceptance of ERP systems is essential to ERP success (Sun., Bhattacherjee & Ma, 2009).ERP deployment in an organization itself does not bring any change to its functions. It is beneficial if the end user use it in their job. There are a number of potential factors that could influence the usage of an information system. If users are not using a technology or organization fails to achieve target level of use then the system is not considered successful. To implement ERP system, the goal of managers is to achieve the desired level of use of the system Organizations are recognizing that user satisfaction with information systems is one of the most important determinants of the success of those systems ((Gymph, 2007). The theorist and practitioners, exploring the acceptance of user toward information system have developed many models and theories (Ajzen, 1991; Davis, 1989; Venkatesh, Morris, Davis & Davis, 2003; Clarke, 1999; Thompson, Higgins, & Howell, 1994).Kazmi, (2008) found role of strategic IT planning, executive and managerial commitment, IT skills, business process skills, ERP training and learning are very important in successful ERP implementation. Further he stated that the gap between ERP users experience and skills is need to bridge by conducting successful training for employees.Organizational performance depends on individuals' task completion. With the rapid growth in use of computing in organizations, practitioners are constrained to investigate the impact of information technology acceptance on individuals' performance (Huang & Wang, 2009; Law & Ngai, 2007; Nah , Lau, & Kuang 2005;Ramayah &Lo, 2007; Vosburg & Kumar, 2001; and Igbari & Tan, 1997).

Technology acceptance has become a mature field. Scholars and practitioners have investigated ERP usage behavior in various sectors like telecommunication, oil and gas, manufacturing government sector and finance sector (Al-Jabri & Al-Hadab, 2008; Govindaraju et al. (2008); Anjum & Rehman, 2010; Arunthari & Hasan, 2005; Baray,

[1]*Saima Kanwal is MS leading to PhD Scholar in Department of Technology Management, International Islamic University Islamabad saimakanval@gmail.com*

[2]*Irfan Anjum Manarvi is Associate Professor, Head of the Department, Department of Mechanical Engineering, HITEC University, Taxila Cantt., Pakistan. irfanmanarvi@yahoo.com*

Hameed, & Badiil, 2006; Bueno & Salmeron, 2008; Calisira & Calisir, 2004; (Chang, Cheung, & Cheng, 2008; Chung & Snyder, 2000; Davenport, 1998; Govindaraju, Maathuis, Bruijn , 2008; Holland, Light, 1999; Gumussoy, Calisir, & Bayram, 2007; Gyampah, 2007; Law and Ngai (2007); Zhang, Huang, Zhang, Huang, 2005; Hossain, Patrick, Rashid, 2002; and Shih, 2006); Ramayah & Lo (2007). According to Rajapakse, & Seddon (2005) ERP adoption rate in developing countries of Asia is very low. The positive ERP usage behavior leads to job completion on time, which improves the employee's performance. According to Vankatesh et al (2003), little research has addressed the link between user acceptance and individual or organization usage outcomes. This study will explore a set of variables that have their influence on ERP system usage behavior in developed and developing environment. It will provide information as to which variable is more influential in using ERP system by

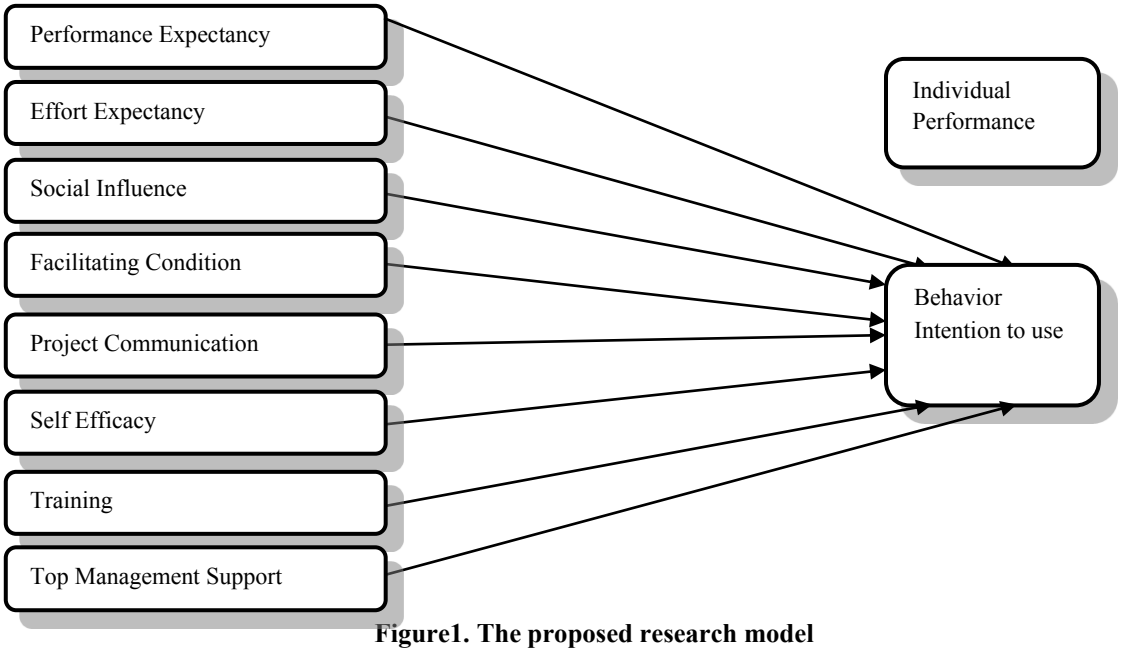


Figure1. The proposed research model

II. Research Hypotheses

In order to achieve the research objective, following research hypotheses are proposed.

H1-Performance expectancy has positive impact on employee behavior intention to use ERP system

H2-Effort expectancy has impact on employee behavior intention to use ERP system.

H3-social influence has positive impact on employee behavior intention to use ERP system.

H4-Facilitating Condition has positive impact on employee behavior intention to use ERP system.

H5-Self Efficacy has positive impact on employee behavior intention to use ERP system.

H6-Project Communication has positive impact on employee behavior intention to use ERP system.

the employees of telecom sector in Pakistan. More over the impact of behavior intention on employee's performance is also measure, the name of the telecom organization is kept confidential in the study and not mentioned here due to organizational polices.This research is expected to contribute to the understanding management to know factors responsible for an employee's behavior toward use of ERP system. The second part of the research is also important in providing information about impact of behavior intention to use system on performance of employees.Based on the factors explored from literature. a research model is proposed. As shown in figure 1 In this research model, performance expectancy, effort expectancy, social influence, facilitating condition, top management support, project communication, self efficacy and training are the independent variables which have their effect on the behavior intention to use ERP system (dependant variable). While behavior intention to use ERP system has its effect on Individual performance

H7-A relationship exists between top management support and employee behavior intention to use ERP system.

H8-Training has positive impact on behavior intention to use ERP system

H9-Employees performance is positively correlated with behavior intention to use ERP system

III. Methods

A. Respondents

Questionnaire was distributed among 300 employees working in telecom organization located in Islamabad-Pakistan. In response, 270 questionnaires were returned. Data of 255 completely filled questionnaires were entered in Statistical Package for Social Sciences (SPSS) for analysis. Inappropriately filled questionnaires, 15, were excluded from the data analysis. Therefore, the response rate was 85 %.The response shows that the sample represented

executive, team lead, assistant manager and managers. A pilot test was conducted to verify the various dimensions of the questionnaire such as language used, ease of completing the questionnaire and appropriateness of questions with relevance to usage behavior. Twenty ERP users working in organization were asked to fill in the questionnaire. Feedback was obtained about the clarity, wordings, interpretation, and appropriateness of the questions. The pilot test resulted in several small revisions to the primary instrument that included rewording of a few items No scaled item was dropped or added as a result of the pilot study.

B. *Measure*

A five point Likert type scale questionnaire based on items adapted from Davis (1989), Taylor & Todd (1995, Venkatesh et al. (2003) and Thompson et.al (1991)

IV. Results and Data Analysis

A. *Reliability Statistics*

To confirm the reliability of the questionnaire, Cronbach's Alpha reliability statistics analysis was conducted. In statistics the corbbach's Aplha value greater than .5 is considered to be a reliable scale.

Table1. Reliability Statistics

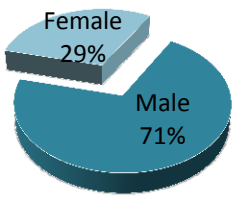
Cronbach's Alpha	N of Items
.967	28

Table 1 shows that reliability statistics of the questionnaire. The value .967 shows that the scale used in the questionnaire is highly reliable.

B. *Descriptive Statistics*

In order to explore the ERP user responses with respect to gender. A frequency statistics was made.

Figure 2. Descriptive data of gender wise response



The figure 2 shows the frequency distribution of the respondents. Out of 255 responses, 71% were male and 29% were female.

Figure 3. Age wise distribution of ERP users

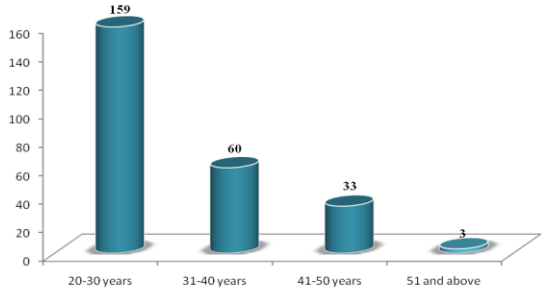


Figure 3 shows the variation in age of the respondents. Out of 255 respondents, majority lies in 20-30 age group (N=159), while 60 respondents are in 31-40 age group, 33 are in 41-50 and only 3 respondents are above 51.

Figure 4. Experience wise distribution of ERP users

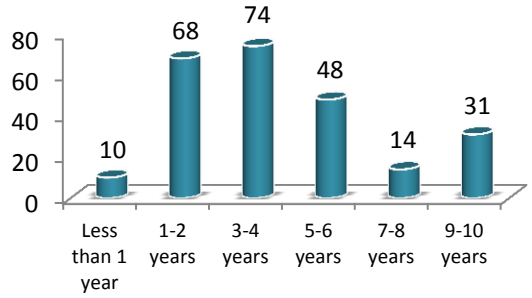


Figure 4 shows the respondents distribution according to their experience. Mostly respondents (N=74) have 3-4 year expereince, 68 respondents have 1-2 year experience, 31 are in 9-10 year experience block while 14 have 7-8 and only 10 have less than one year experience.

To find out the respondents' experience in using ERP system. The frequency distribuiton with respoct to their experience in ERP use is shown below.

Figure 5 ERP Usage Experience

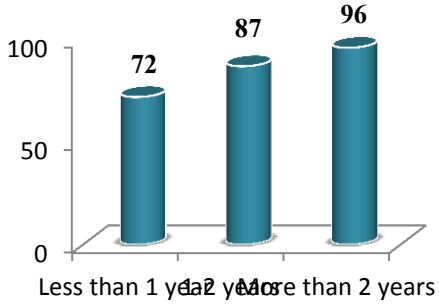


Figure 5 represent the ERP usage experiences possessed by the employees. Mostly respondents (N=96) had more than 2 year experience, 87 having 1-2 year experience while only 72 have less than one year experience in use of ERP system.

A. *Hypotheses results and analysis*

In table 2 the R2 (.71) value shows that the independent variable explains the 71% variation in the behavior intention

to use ERP system. Here we can say that our model best fits and it explain significant variation in the behavior.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.843(a)	.711	.702	.38816

Table2. Regression Analysis

Table 3: Detailed Regression Analysis

Dependant variables	R2	Independent variables	Beta	t	Sig
Behavior Intention	.71	Performance Expectancy	.322	4.602	.000*
		Effort Expectancy	-.358	-4.655	.000*
		Social Influence	.007	.094	.925
		Facilitating Condition	.063	.917	.360
		Self Efficacy	.505	6.873	.000*
		Project Communication	.036	.686	.494
		Training	.172	2.967	.003*
		Top Management Support	.189	2.967	.003*

Note. * Significant at .005 level

Table 3 shows the beta and significance value of each independent variable in regression model

The significance value (p=.000) in table 3 shows that performance and effort expectancy is significant in measuring the behavior intention. The Beta value, B=.322 of performance expectancy show that performance expectancy contribute to .322 variation the behavior intention.The negative beta and t value indicate that this variable is not positively associated with the behavior intention.. Here we will accept the H1 and H2.

Table 3 shows the regression analysis, the p value (p>.005) shows that Social influence is not a significant variable in measuring the behavior intention . Hence, we reject H3.

The p value (p=.360) in table 3 shows that facilitating condition is not a significant variable in measuring the behavior intention. Here we reject H4.

Table 3 shows the regression analysis, the value (B =.505) shows that the variable Social influence strongest predictor in measuring the behavior intention. The p value (p=.000) also shows that Self efficacy is a significant variable in measuring the behavior intention. Here we accept H5.

In table3, project communication is a not significant variable in measuring the behavior intention. Here we reject H6.

The significance value (p=.003) in table 3 depict that top management support and training contributes significant value while predicting the behavior intention to use ERP system. Here we accept H7 and H8.

To test this hypothesis 9 the regression analysis for model 2 will be implied.

Table 5.4: Regression Model II

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
2	.738(a)	.545	.543	.60026

The table 4 shows that the behavior intention toward use of ERP system can bring 54 % variation in the performance of the employee. Hence, here we accept the H9.

In order to analyze the effect of demographic characteristics on performance expectancy and effort expectancy of the ERP usage behavior as explained in literature review, the constructs of measuring the behavior intention were analyzed with demographic statistics.

Table 5 shows the t-test values of performance expectancy with respect to gender. The significance value p>005 (p=.010) show that there is difference among the opinion of male and female about the performance expectancy associated with the ERP system use.

Table 5: t-test-performance expectancy of ERP system with respect to gender

Statement	N	Gender	Mean	t	Sig.
Performance Expectancy	180 75	Male Female	4.01 4.41	-3.47	**.010***

Note. * Significant at .005

Table 6 shows the ANOVA values of performance expectancy with respect to age The significance value p<.005 shows that these is no significant difference between the responses against behavior intention to use ERP system among different age groups.

Table 6: ANOVA Statistics About Performance Expectancy of ERP System Usage With Respect to Age

Statement	Mean				F	Sig.
Performance Expectancy	20-30 years	31-40 years	41-50 years	51 and above		
	4.02	4.36	4.50	3.91	6.06	.001

Note. * Significant at .005

Table 7 shows the ANOVA analysis of three groups of ERP usage experience with respect to the performance expectancy of ERP system use. The significance value, p=.005, shows that there is no difference among the opinion of respondents of different experience group about the performance expectancy of ERP system.

Table 7: ANOVA Statistics About Performance Expectancy of ERP System Usage With Respect to ERP Usage Experience.

Statement	Mean			F	Sig.
Performance Expectancy	< 1year	1-2 years	> 2 years		
	3.88	4.29	4.15	5.32	.005*

Note. * Significant at .005

The table 8 shows the t-test value of effort expectancy with respect to gender significance value in table 5.16 p <.005 (p=.003) shows that there is no difference between the perception of respondents with respect to their gender about the effort expectancy of ERP system use.

Table 8: T Test Regarding Effort Expectancy of ERP System With Respect to Gender

Statement	N	Gender	Mean	t	Sig.
Effort Expectancy	180	Male	4.00	-3.814	.003
	75	Female	4.41		

Note. * Significant at .005

The table 9 shows ANOVA test regarding the perception of different age group about the effort expectancy of ERP system use. Significance value p>005 (p=.009) in table 5.17 shows that there is significant difference between the responses about effort expectancy of different age group against behavior intention to use ERP system

Table 9 ANOVA Test Regarding Effort Expectancy of ERP System With Respect to Age

Statement	Mean				F	Sig.
Effort Expectancy	20-30 years	31-40 years	41-50 years	51 and above		
	4.03	4.15	4.54	4.00	3.958	.009

Note. * Significant at .005

Table 10 shows the ANOVA analysis of three groups of ERP usage experience with respect to the effort expectancy of ERP system use. The significance value, p=.000, shows that there is no difference among the opinion of respondents of different experience group about the effort expectancy of ERP system.

Table 10: ANOVA Statistics About Performance Expectancy of ERP System Usage With Respect to ERP Usage Experience.

Statement	Mean			F	Sig.
Performance Expectancy	Less than one years	1-2 years	More than 2 years		
	3.84	4.27	4.30	10.120	.000*

Note. * Significant at .005

V. Findings

The results of correlation analysis shows that performance expectancy, facilitating condition, self efficacy, training and top management support are strongly correlated with the behavior intention to use ERP system. While effort expectancy, social influence, project communication have medium level correlation with behavior intention to use ERP system.

The R square value (.71) shows that the over all independent variable explains 71% variation in the behavior intention to use ERP system. Here we can say that the model best fits and it explain significant variation in the behavior.

While exploring all variables individually, the variable performance is significant in measuring the behavior intention. The Beta value, β=.322, show that performance expectancy is stronger predictor of the behavior intention.

Effort expectancy is significant while explaining behavior intention to use ERP system The negative beta and t value indicate that this variable is not positively associated with the behavior intention. The possible explanation of this negative relation can be the employee‘s perception as this is the case of mandatory technology (ERP) and even if the system if difficult to use employees has to use the system.

The p value of social influence (p=.925) shows insignificant variable in measuring the behavior intention.

The p value (p=.360) in the regression analysis of the facilitating condition shows that this is not a significant variable in measuring the behavior intention.

The regression analysis shows that Social influence is a strongest predictor in measuring the behavior intention while project communication is a not significant variable in measuring the behavior intention.

Top management support and training were also found the significant contributor toward measuring behavior intention toward use of ERP system.

The t-test value shows that

i. There is no difference among the opinion of male and female about the performance expectancy associated with the ERP system use.
ii. There is no difference between the perception of respondents with respect to their gender about the ease of use of ERP system.

The ANOVA statistics shows that

i. There is no significant difference between the responses against behavior intention to use ERP system among different age groups
ii. There is significant difference between the responses about effort expectancy of different age group against behavior intention to use ERP system
iii. There is significant difference between the responses against behavior intention to use ERP system about effort expectancy among different level of experience groups.

The regression model 2 shows that behavior intention toward use of ERP system can bring 54 % variation in the performance of the employee

VI. Conclusions

This study was centered on the ERP usage behavior. The aim of this study was to investigate factors affecting the employee behavior toward use of ERP system and its impact on their performance.

The first part of the study explores the literature related to implementation and acceptance issues of ERP. Use of different models and theories of individual behavior of ERP acceptance in different sectors especially in the telecom sector of developed and developing countries have also been discussed.

During the exploration of literature, many factors were identified for measuring the behavior intention toward use of ERP system. Based on the most repeated and influential factors a theoretical model was proposed.The population of this study was telecom organizations‘ employees who were using ERP system in the company. The sample size of 300 respondents was selected, however, 255 out of 300 selected participants responded. A questionnaire based survey was administered personally on 300 employees who were using ERP system. In response to the survey, 255 valid responses were received. The response rate was 85%. Among the respondents, 71% were male while 29% were female. The findings of this study indicate that the proposed model over all explains 71% variation in the behavior. While Performance Expectancy, Self Efficacy, Training and Top Management Support were significant factors while explaining the user behavior toward use of ERP system. Further, the model explains that there is difference about the opinion on performance expectancy by male and female. Effort expectancy related to ERP usage among different age groups and different experience group differ. Majority of the respondents agreed that they like to use ERP system. They want to become professional in ERP system use and intend to use ERP for future professional assignments. Seventy percent respondents say that ERP system has improve their performance in work. This research will help the management to understand the factors responsible for the employees behavior toward use of ERP system.

VII. Recommendations

Based on the findings of this study, the following recommendations are given to increase behavior intention toward use of ERP system.

i. There is need to ensure the employees that by using ERP system they will become a valuable assets to the organization.
ii. The technical support should be provided to employees to use ERP system.
iii. Frequent ERP training program should be launched for employees for using ERP.
iv. The training program for ERP should be easy.
v. Top management should communicate well in time the employees about the implementation of sub modules of the ERP system.
vi. The senior management should encourage ERP usage.
vii. Organization should provided necessary resources to use ERP System.
viii. Employees should be given guideline for using builtin help in ERP system.

VIII. Future Research Direction

Current research explains only 71% variation in the behavior of employees. The 29% portion of behavior intention is unmeasured. While this research can only measure the 54% variation in the performance of employees. There is a need of future research which explores the further variables to measures the leftover portion of behavior which was not measured in this research.

IX. References

1) Ajzen, I. (1991). The Theory of Planned Behavior. organizational behavior and human decision processes , 50, 179-221.
2) Al-Jabri, I. M., & Al-Hadab, A. (2008). End User Adoption of ERP Systems:Investigation of Four Beliefs. Americas Conference on Information Systems (AMCIS) (pp. 1-8). Toronto: Association for Information Systems.
3) Angello, C., & Wema, E. (2010). Availability and usage of ICTs and e-resources by livestock researchers in Tanzania: Challenges and ways

forward. International Journal of Education and Development using Information and Communication Technology , 6 (1).

4) Anjum, M. J., & Rehman, I. U. (n.d.). CRITICAL SUCCESS FACTORS OF ERP IMPLEMENTATION IN PAKISTAN (A Case of Telecom, Engineering, Oil and Government Sector Organization.
5) Arunthari, S., & Hasan, H. (2005). Beliefs and Attitudes Associated with ERP Adoption Behaviours: A Grounded Theory Study from IT managers and End-users Perspective . Pacific Asia Conference on Information System (pp. 1492-1497). Bangkok: National Sun Yat-sen University.
6) Baray, S., Hameed, S., & Badii, A. (2006). Analysing the effectiveness of implementing enterprise resource planning systems in the printing industry. European and Mediterreanean Conference on Information Systems, (pp. 1-20). Costa Blanca, Alicante , Spain.
7) Bueno, S., & Salmeron, J. L. (2008). TAM-based success modeling in ERP. Interacting with Computers 5–523 , 20, 515-523.
8) Calisira, F., & Calisir, F. (2004). The relation of interface usability characteristics,perceived usefulness, and perceived ease of use to end-user satisfaction with enterprise resource plnning systems. Computers in Human Behaviour , 20, 505-515.
9) Chang, M. K., Cheung, W., Cheng, C. H., & Yeung, J. H. (2008). Understanding ERP system adoption from the user‘s perspective. International Journal of Production Economic , 113, 928-942.
10) Davenport, T.H. (1998). Putting the enterprise into the enterprise system. Harvard Business Review, (July–August), 121–131.
11) Davis, F. D. (1989). Perceived Udefulness,Perceived Ease of USe, and User Acceptance of Information Technology. MIS Quarterly , 13 (3), 319-335.
12) Govindaraju, R., Maathuis, S. J., & de Bruijn, E. J. (2008). Individual And Organizational Factors Influencing The Behavioral Intention To Use ERP Systems. Proceedings of the 9th Asia Pasific Industrial Engineering & Management Systems Conference, (pp. 858-864). Nusa Dua, Bali – INDONESIA.
13) Gumussoy, C. A., Calisir, F., & Bayram, A. (2007). Understanding the behavioral intention to use ERP systems: An extended technology acceptance model. (pp. 2024-2027). **: IEEM.
14) Gyampah, K. A. (2007). Perceived usefulness, user involvement and behavioral intention: an empirical study of ERP implementation. Computers in Human Behavior , 23, 1232-1248.
15) Holland, C.P., Light, B., 1999. A critical success factors model for ERP Implementation. Software IEEE 16 (3), 30–36.
16) Hossain, L., Patrick, J. D., & Rashid, M. A. (2002). Enterprise Resource Planning: Global Opportunities & Challenges. United State of America: Idea Group Publishing.
17) Huang, S., & Wang, X. (2009). Influence of Organizational System to End-Users‘ Acceptance of ERP System in Chinese Enterprises. Ninth International Conference on Hybrid Intelligent Systems (pp. 160-164). IEEE computer society.
18) Igbaria, M., & Tan, M. (1997). The consequences of information technology acceptance on subsequent individual performance. *Information & Managemen , 32*, 113-121.
19) Kazmi, S. N. (2008). Competitive Constructs of ERP Implementation Public Sector in Pakistan. *International Conference on Business and Management* (pp. 1-11). Bangkok: nternational Colloquium on Business & Management .
20) Law, C. C., & Ngai, E. W. (2007). ERP systems adoption: An exploratory study of the organizational factors and impacts of ERP success. Information & Managemen , 44, 418-432.
21) Nah, F. F., Lau, J. L. & Kuang, J. (2001). Critical Factors for Successful Implementation of Enterprise Systems. Business Process Management, 7(3), 285–296.
22) Nah, F. F.-H., Tan, X., & Beethe, M. (2005). End-users‘ Acceptance of Enterprise Resource Planning (ERP) Systems: An Investigation Using Grounded Theory Approach. Americas Conference on Information Systems (AMCIS) (pp. 2053-2057). Omaha, NE, USA: Association for Information Systems.
23) Rajapakse, J. and Seddon, P. (2005b), ―ERP adoption in developing countries in Asia: a cultural misfit‖, paper presented at the 28th Information Systems Seminar in Scandinavia, Kirstiansand, 6-9 August, available at: www.hia.no/iris28/Docs/IRIS2028-1028.pdf (accessed 3 March 2010).
24) Ramayah, T., & Lo, M. C. (2007). Impact of shared beliefs on―perceived usefulness‘‘ and ―easeof use‘‘ in the implementation ofan enterprise resource. Management Research News , 30 (6), 420-431.
25) SHIH, Y. Y. (2006). The effect of computer self-efficacy on enterprise resource planning usage. Behaviour & Information Technology , 25 (5), 407-411.
26) Sun. Y, Bhattacherjee. A, Ma.Q,(2009) Extending technology usage to work setting: The role of perceived work compatibility in ERP implementation, Journal of Information and Management,46,pp 351-356
27) Taylor, S., & Todd, P. (1995). Assessing IT Usage: The Role of Prior Experience. MIS Quarterly , 19 (4), 561-570.
28) Thompson, R. L., Higgins, C. A., & Howell, J. M. (1994). Influence of experience on personal

computer utilization: testing a conceptual model. Journal of Management Information Systems , 11, 167-187.

29) Thompson, R. L., Higgins, C. A., & M, J. (1991). Personal Computing: Toward a Conceptual Model of Utilization. MIS Quarterly , 15 (1), 125-143.
30) Venkatesh, V., Morris, M., Davis, G., & Davis, F.(2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27, 425.
31) Vosburg, J., & Kumar, A. (2001). Managing Dirty data in Organization using ERP:Lessons from a case study. Industrial Management & Data Systems , 101 (1), 21-31.

Active Queue Management in TCP NetworksBased on fuzzy-PID Controller

GJCST Computing Classification
C.2.5 , I.5.1

[1]Hamed Moradi Pour, [2]Hossein Ashtiani,[3] Mohsen Nikpour

***Abstract*- We introduce a novel and robust active queue management (AQM) scheme based on a fuzzy controller, called hybrid fuzzy-PID controller. In the TCP network, AQM is important to regulate the queue length by passing or dropping the packets at the intermediate routers. RED, PI, and PID algorithms have been used for AQM. But these algorithms show weaknesses in the detection and control of congestion under dynamically changing network situations. In this paper a novel Fuzzy-based proportional-integral derivative (PID) controller, which acts as an active queue manager (AQM) for Internet routers, is proposed. These controllers are used to reduce packet loss and improve network utilization in TCP/IP networks. A new hybrid controller is proposed and compared with traditional RED based controller. Simulations are carried out to demonstrate the effectiveness of the proposed method and show that, the new hybrid fuzzy PID controller provides better performance than random early detection (RED) and PID controllers.**

Keywords- AQM, Fuzzy controller, Fuzzy PID controllers, PID controllers, adaptive hybrid controllers.

I. INTRODUCTION

Internet is the fundamental part for running many new applications such as Web, multimedia, etc. However, due to unpredictability in interference and number of users who may access Internet in a given time congestion may result. This brings about long delays in data transmission and frequently makes the queue length in the buffer of intermediate routers to overflow, and even may lead to total network collapse. An active queue management (AQM) scheme is one of the efficient tools which detects inceptive congestion and gives early notice of current Internet situation by dropping (or marking) the incoming packets before router queues become full. Recently, many active queue management (AQM) schemes have been proposed to increase network utilization by regulating queues at the bottleneck links in TCP/IP networks, including Random early detection (RED) [1], Adaptive RED (A-RED) [2], proportional-integral (PI) controller [3], and Random exponential marking (REM)[4]. Random early detection (RED) algorithm, the earliest well-known AQM scheme, was proposed and is now used in the Internet routers for reducing the flow synchronization problem and calming the traffic load via the measurement of average queue length..

[1]*Hamed Moradi Pour, Electrical Engineering Department, Zanjan University , H_moradi25@yahoo.com*
[2]*Hossein Ashtiani, Electrical Engineering Department , Zanjan University , hashtiani79@yahoo.com*
[3]*Mohsen Nikpour, Electrical Engineering DepartmIslamic Azad University-Babol Branch, Babol, Iran, mhsnnikpour@yahoo.com*

Unfortunately, RED causes oscillations and instability in the network due to the parameter variationsTherefore, some modified RED schemes have been introduced in the literature. Proportional-integral-derivative (PID) feedback control is a practical and simple control approach for controller design. This approach has been used to design and analyze various present AQM schemes in Internet congestion control.Based on linearized fluid TCP/AQM model, a proportional-integral (PI) controller was developed [5] to regulate the queue length, round trip time and packet loss. The virtual rate control (VRC) algorithm for AQM in TCP networks has been proposed in [6]. Additionally, a saturated nonlinearity of the control input usually exists in such a control problem due to the property of packet-dropping probability. Therefore, the effect of a saturated actuator should also be considered; otherwise it may cause serious degradation and instability of the network especially in large-scale, complex ones.In this paper, a PID controller for a time-delayed TCP/IP network with input saturation is developed to achieve a stable and desired queue length, low packet loss and high link utilization. Numerical simulations show that the proposed scheme is of good stability and is robust with respect to variations in number of TCP sessions.

II. AQM MECHANISMS

AQM mechanisms aim to provide high link utilization with low loss rate and queuing delay while reacting to load changes quickly. Several schemes have been proposed to provide congestion control in TCP/IP networks. RED, which was the first proposed AQM algorithm, simply sets some minimum and maximum marking thresholds in the router queues. In case the average queue size exceeds the minimum threshold, RED starts randomly marking packets based on a probability which depends on the average queue length, whereas if it exceeds the maximum threshold every packet will be dropped. The properties of RED have been extensively studied in the past few years. Here issues of concern are: problems with performance of RED under different scenarios of operation and loading conditions [7], the correct tuning of RED parameters which implies a —global" parameterization which itself is a very difficult job if not impossible, as is shown in [9]. Some researchers have advocated against using RED, in part because of its tuning difficulty [8]. Linearity of the dropping function has also been questioned by some number of researchers.

A. Random Early Detection: RED

The RED active queue management algorithm allows network operators to simultaneously achieve high throughput and low average delay. However, the resulting average queue length is quite sensitive to the level of congestion and to the RED parameter settings, and is therefore not predictable in advance. Delay as a major component of the quality of service delivered to customers, is used by network operators to give a rough a-priori estimate of the congestion in routers. To achieve such predictable average delays, RED would require constant tuning of the parameters to current traffic conditions. Since the RED-based algorithms control the macroscopic behavior of the queue length (average), they often cause sluggish response and fluctuations in the instantaneous queue length. As a result, an important change in end to end delays is observed. As a consequence, RED and its variants provoke dramatic consequences on sensitive flows.

B. Control-Theoretic to Flow Control

The control theoretic techniques have been lately introduced in flow control and congestion avoidance. These recently developed mechanisms outperform the existing works by presenting formal proofs. Whereas classical approaches are rather informal, they give the necessary means to explicitly specify multi criterion performance. These alternative methods model the AQM algorithm as a feedback control system that tunes the router queue length as a plant variable.

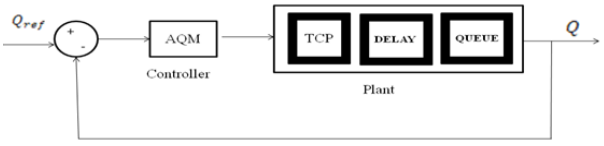


Figure 1. Feedback control modeling of congestion control with AQM

In the system above, Qref represents the reference value of the queue length. The *Plant* represents a combination of subsystems such as TCP sources, TCP receivers, routers ... and others.

Among the contributors to control-theoretic flow control, [10],[11] proposed use of the PID controller which is widely used in automatic control systems. In [12] a feedback compensation element was introduced in order to provide a more robust controller under time-varying network conditions. Beside that, in a variable structure based control scheme was used to take into account the model uncertainties and the number of active TCP connections. These controllers permit better performance in compared with REDs. For instance, the PID controller adds to the integral control, a proportional control for a faster response, and a derivative control for anticipated congestion avoidance.

The theoretic control in the network management domain brings some new alternatives which permit good performance in router throughput and better management in queue length. Note that the plant model is strongly uncertain and nonlinear. This leads to some parameterization problems under realistic traffic environment which is characterized by its intrinsic bursty nature and time-variable structure. Based on this fact, it is much more suitable to deploy an adaptive/auto-tuned control to tackle the changing network conditions

III. Fuzzy – Based PID Controller Structure

PID is regarded as the standard control structures of the classical control theory, and fuzzy controllers have positioned themselves as a counterpart of classical PID controllers with the same dominant role at the knowledge rich spectrum. PID controllers are designed for linear systems and provide a preferable cost/benefit ratio. However, the presences of nonlinear effects limit their performances. On the other hand Fuzzy controllers are successfully applied to non-linear systems because of their knowledge based nonlinear structural characteristics. Hybridization of these two controllers brings to ones mind immediately to exploit the advantages of both categories.

The objective of this study is to design a new hybrid fuzzy PID controller for congestion control in TCP/IP networks so as a further improvement can be achieved in response performance , both the transient and steady states, in compared with the system response obtained from each individually [13],[14].

A. Proposed Controller

The proposed hybrid controller which is shown in Figure 2 consists of two main parts: the classical PID controller and fuzzy PID controller.

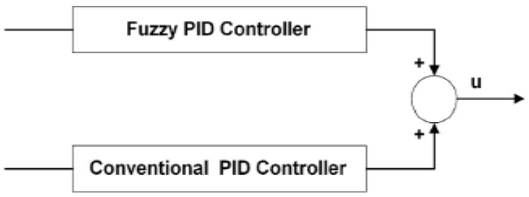


Figure 2. Block diagram of hybrid type fuzzy PID controller

Transfer function of a PID is generally written as

$$Gpid(s) = K(1 + \frac{1}{T_I s} + T_D s)$$

where K is the proportional gain, K_I the integral gain, K_D the derivative gain, T_I the integral time constant and T_D the derivative time constant[15].

The structure of the fuzzy PID controller, which has two inputs and one rule base, is shown in Figure 3.

The inputs are the conventional error (e) and the rate of the change of error ($\dot{e}$).

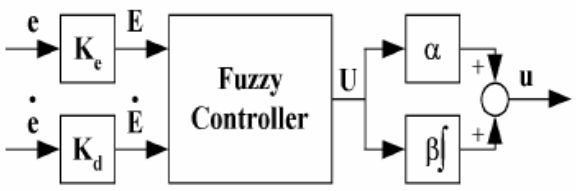


Figure 3. The Fuzzy PID Controller structure

Triangular membership functions are used for input variables and the output as is shown in Figures 4 and 5.

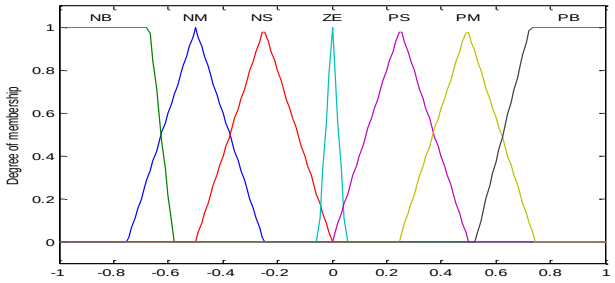


Figure 4. The membership functions of e and $\dot{e}$.

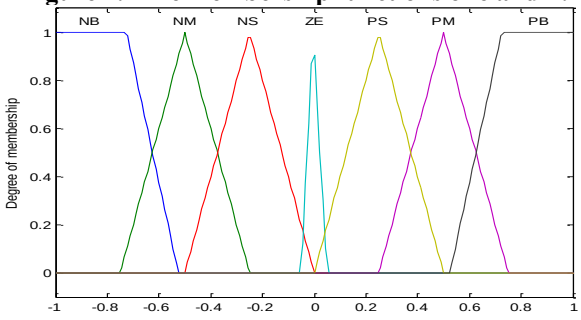


Figure 5. The membership functions of u.

In this paper, the classical PID and fuzzy PID controller are combined by a blending mechanism which depends on a certain function of actuating error [16]. Moreover, an intelligent switching scheme is induced on the blending mechanism that makes a decision based on the priority of the two controller parts. The Matlab/Simulink simulation model of the proposed intelligent hybrid PID controller is shown in Figure 6. The parameters of the PID controller are denoted by K, T_I, and T_D. These stand for proportional gain, integral and derivative time constants respectively. The parameters of the fuzzy controller are defined as K_e, K_d, α, and β.

Figure 6. Due to lack of space, it is shown At the end of the paper.

The fuzzy PID controller rule base is composed of 49 (7x7) rules as shown in Table 1.

Table 1: PID type Fuzzy Controller Rule Base

E / CE	NB	NM	NS	Z	PB	PB	PB
NB	NB	NB	NB	NM	NS	NVS	Z
NM	NB	NB	NM	NS	NVS	Z	PVS
NS	NB	NM	NS	NVS	Z	PVS	PS
Z	NM	NS	NVS	Z	PVS	PS	PM
PS	NS	NVS	Z	PVS	PS	PM	PB
PM	NVS	Z	PVS	PS	PM	PB	PB
PB	Z	PVS	PS	PM	PB	PB	PB

IV. Performance Evaluation

A. Simulation Model

In order to evaluate the performance and robustness of our proposal, we performed some simulations using the topology depicted in figure 7. The bottleneck is located at the central router level which implements one of the three AQM algorithms: RED, PID or Adaptive Hybrid Fuzzy (our proposal). Its capacity is 10 Mbps (2621 packets/s, default packet size is 500 bytes), and delay is 20 ms. Each queue may contain up to 400 packets.[17]

However, the most performing queue length is of 80 packets.

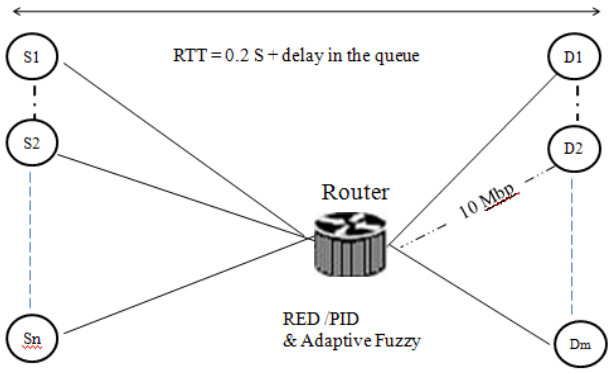


Figure 7. Simulation Network Topology

Note that the parameters of RED and PID are set as recommended respectively in [18] and in [19]. The controller parameters of the classical PID controller are set to K=2, T_I = 0.25, T_D = 0.025 in order to have a small rise time. On the other hand, the fuzzy PID controller has the following parameters: α = 0.05, β = 4.5, Ke = 1, Kd = 0.56.

In our simulation, the number of TCP connections changes in accordance with the curve in Figure 8.

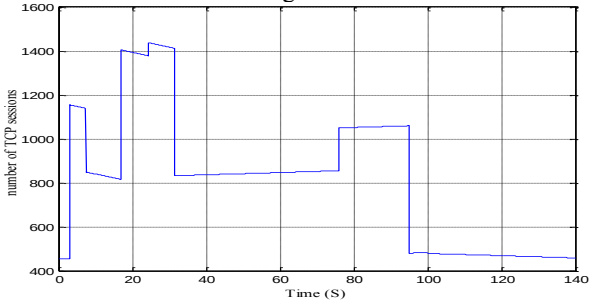


Figure 8. Variations in time of TCP sessions number

B. Experimental Results

Figure 9 presents the queue evolution, which shows that the Hybrid Fuzzy controller regulates the queue to the reference value very quickly in spite of the work-load variations. Contrary to Hybrid Fuzzy , the PID controller converges slowly and is very disturbed with the variations of the number of TCP sessions. The statistics in Table 2 show that, in the case of the PID controller, the static error decreases

and the queue length converges to the referenced value asymptotically with the stabilization of the work-load. Figure 9 shows that RED doesn't guarantee the stability of the queue; it only permits good performance with respect to the throughput (see Table 2). As It can be seen from Figure 10, all three controllers permit good performance. However, Hybrid Fuzzy gives a constant and maximal throughput value by stabilizing of the queue size as is shown in Figure 9.

Note that the maximal throughput of the router is 10 Mbps, which means that any excessive traffic will cause a buffering at the router queue, and even higher drop rate when the buffer queue length is reached (figure 9).

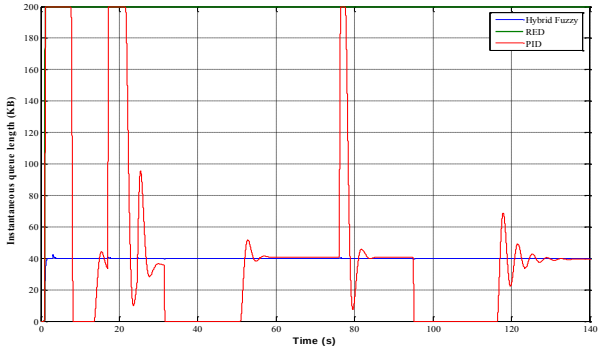


Figure 9. Instantaneous queue length

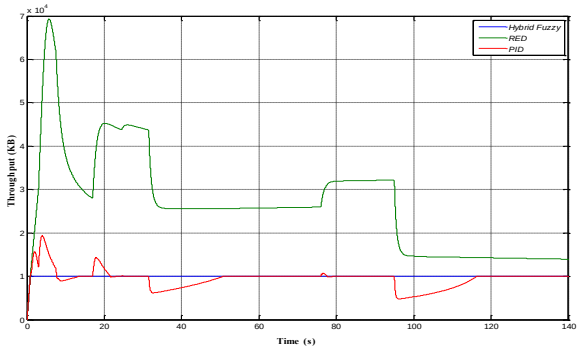


Figure 10. *Throughput*

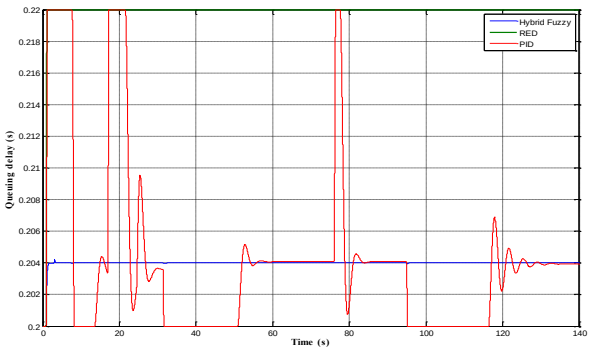


Figure 11. Queuing delay

From figure 10 and 11, it can clearly seen that queuing delays in Hybrid Fuzzy are constant. RED and PID controllers are sensitive to the number of session's variations. In addition, fluctuations in queuing delay implies a deep deterioration of multimedia communications which have strict requirements in term of end to end delay. So, in order to provide quality of service (QoS) guarantees, the elimination of the queuing delay fluctuation is a must. Our controller achieves this objective.

Furthermore, from the statistics tables, it can be seen that Hybrid Fuzzy induces a maximum delay of 0.004 s, while the RED and PID controllers have 0.0200 s each.

By looking at figure 12, it can be observed that the dropping probability in Hybrid fuzzy is not regular. Moreover, it is clearly seen that the dropping probability value in Hybrid Fuzzy is better than PID.

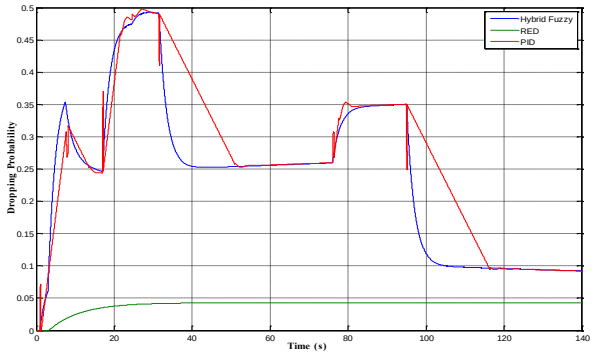


Figure 12. Dropping probability

The Hybrid Fuzzy algorithm shows much better performance than RED and PID controllers. In fact, the performances of RED and PID are sensitive to network fluctuations. In addition, RED and PID can make a buffer overflow in a transient period which itself can result in packets loss. All these remarks are summarized in Table 2, which briefly describes advantages and disadvantages of each of the AQM algorithm considered in this paper.

Table 2. A classification of AQM algorithms in terms of objectives.

Characteristics	Hybrid fuzzy	RED	PID
Router Throughput	good	good	good
Sensitivity to network parameters	no	yes	yes
Response	very fast	slow	fast
Queue stability	stable	not stable	stable
Delay	small	very large	variable
Buffer overflow/underflow	no/no	yes/yes	yes/yes

V. Conclusions

We have presented a new AQM scheme using fuzzy logic techniques, which we refer to as Hybrid Fuzzy PID controller. It can be implemented in TCP/IP networks to provide effective congestion control for high utilization, low losses and delays which are very important for multimedia applications.

Hybrid Fuzzy controllers behave better than other AQM schemes in terms of queue fluctuations and delay, loss and link utilization of packets in TCP/IP networks. Obtained

results show that Hybrid Fuzzy Control methodology will offer significant improvements on controlling congestion in TCP/IP networks. It permits a very fast response in compared with the classical adaptive controllers, RED and PID.

VI. References

1) S. Floyd, V. Jacobson, —Random Early Detection gateways for congestion avoidance", IEEE/ACM Trans. On Networking, Aug. 1993.
2) S. Floyd, R. Gummadi, S. Shenker, —Adaptive RED: An Algorithm for Increasing the Robustness of RED's Active Queue Management", Technical report, ICSI, August 2001.
3) Hollot, V. Misra, D. Towsley, W.-B. Gong, —Analysis and Design of Controllers for AQM Routers Supporting TCP Flows" IEEE Transactions on Automatic Control, vol. 47, no. 6, pp. 945-959, June 2002.
4) S. Athuraliya, V. H. Li, S. H. Low, Q. Yin, —REM: Active Queue Management", IEEE Network Magazine, 15(3), pp. 48-53, May 2001.
5) Hollot C V, Misra V, Towsley D and Gong W B 2001 On Designing Improved Controllers for AQM Routers Supporting TCP Flows in: Proc. of IEEE INFOCOM (Alaska, USA) pp.1726.
6) H. Lim, K.J. Park, E.C. Park, and C.H. Choi, "Virtual rate control algorithm for active queue management in TCP networks," IEEE Electron. Lett., vol.38, no.16, pp.873–874, Aug. 2002.
7) Iannaccon G, Brandauer C, Ziegler T, Diot C, Fdida S, —Tai Drop and Active Queue Management Performance for bulk-data and Web-like Internet Traffic", 6th IEEE Symposium on Computers and Communications, Hammamet, 2001.
8) M. May, T. Bonald, and J.C. Bolot, —Analytic Evaluation of RED Performance", Tel Aviv, IEEE Infocom 2000.
9) Yanfei, and F.L. Chuang: —Design a PID Controller for Active Queue Management". ISCC 2003 (July 2003)
10) Yanfei, and Chuang: —Design a PID Controller for Active Queue Management". ISCC 2003 (July 2003)
11) Ryu, S., Rump, C., and Qiao, C.: —Predictive and Robust Active Queue Management for Internet Congestion Control". ISCC 2003 (June 2003).
12) Z. Heying, L. Baohong, and D. Wenhua: —Design of Robust Active Queue Management Algorithm Based on Feedback Compensation". ACM/SIGCOMM 2003 (August 2003).
13) I Erenoglu, I. Eksin, E. Yesil, M. Guzelkaya, —AN INTELLIGENT HYBRID FUZZY PID CONTROLLER " , European Conference on Modelling and Simulation, Wolfgang Borutzky, Alessandra Orsoni, Richard Zobel © ECMS, 2006
14) Akbiyik, B., Eksin I., Guzelkaya M., Yesil, E., 2005. "Evaluation of The Performance of Various Fuzzy PID Controller Structures on Benchmark Systems", ELECO '2005, 4rd International Conf. on Electrical and Electronics Engineering, Bursa, Turkey, 388-393.
15) N. Matsunaga,S. Kawaji, 1991. —Fuzzy hybrid control for DC servomotor." Trans. Inst. Electrical Eng. Japan D (111), 195-200.
16) Yesil, M. Guzelkaya,I. Eksin, 2003. "Fuzzy PID Controllers: An Overview." The Third Triennial ETAI International Conference on Applied Automatic Systems, Skopje, Macedonia, 105-112.
17) Y.H.Aoul, A. Nafaa, D. Negru, A. Mehaoua, —FAFC: fast adaptive fuzzy AQM controller for TCP/IP networks" , Global Telecommunications Conference, 2004. GLOBECOM apos;04. IEEE, Dec. 2004
18) S. Keshav: —Congestion Control in Computer Networks". PhD Thesis, published as UC Berkeley TR-654 (September 1991)
19) Bitorika, A., and al. —A Comparative Study of Active Queue Management Schemes". ICC 2004 (June 2004).

N-Square Approach For Lossless Image Compression And Decompression

G. Srinivas* Prasad Reddy P.V.G.D* K. Gayatri Devi*

GJCST Computing Classification I.4.2

***Abstract*- There are several lossy and lossless coding techniques developed all through the last two decades. Although very high compression can be achieved with lossy compression techniques, they are deficient in obtaining the original image. While lossless compression technique recovers the image exactly. In applications related to medical imaging lossless techniques are required, as the loss of information is deplorable. The objective of image compression is to symbolize an image with a handful number of bits as possible while preserving the quality required for the given application. In this paper we are introducing a new lossless encoding and decoding technique which even better reduces the entropy there by reducing the average number of bits with the utility of Non Binary Huffman coding through the use of N-Square approach and fasten the process of searching for a codeword in a N-Square tree, we exploit the property of the encoded image pixels, and propose a memory efficient data structure to represent a decoding N-Square tree. Our extensive experimental results demonstrate that the proposed scheme is very competitive and this addresses the limitations of D value in the existing system by proposing a pattern called N-Square approach for it. The newly proposed algorithm provides a good means for lossless image compression and decompression.**

Keywords- Minimum Redundancy Code, Image compression, Non Binary Huffman, Redundancy, N^2, Interval generation, N2 Tree.

I. INTRODUCTION

In many military and security applications, there is a need to compress image data to reduce the bandwidth requirements of wireless and networked-based systems. By reducing the amount of data associated with these images, the amount of storage space required can also be minimized, allowing long image sequences to be stored on a single disk. With the advanced development in Internet, teleconferencing, multimedia and high definition television technologies, the amount of information that is handled by computers has grown exponentially over the past decades. Hence, storage and transmission of the digital image component of multimedia systems is a major problem. The amount of data required to present images at an acceptable level of quality is extremely large. High quality image data requires large amounts of storage space and transmission band width [1]. One of the possible solutions is to compress the information so that the storage space and transmission time can be reduced. Image compression address the problem of reducing the amount of data required to represent a digital image with no significant loss of information. The goal of image compression is to represent an image with a few number of bits as possible while preserving the quality required for the given application [2].

In this paper, as per the conventions of Huffman, the sequence of symbols associated with a given message will be called the ―message code". The entire number of messages to be transmitted will be called the ―message ensemble". The amiability between the sender and receiver about the meaning of the code for each message of the ensemble will be called the ―ensemble code". In the way to formulate the requirements of an ensemble code, the coding symbols will be represented by numbers. Therefore, if there are D different types of symbols to be used in coding, they will be represented by the digits 0, 1, 2... D-1 [4].

The number of messages in the ensemble is denoted by N. Let P(i) be the probability of the its message. Then

$$\sum P(i) = 1 \qquad (3)$$

The length of message, L(i) is the number of coding digits assigned to it. Thus the average length is

$$L.avg=\sum P(i)\, L(i) \qquad (4)$$

This paper presents a new array data structure to represent the N-Square tree. The memory required in the proposed data structure is $nd = O(n)$, which is less than the previous ones. We then address an efficient N-Square decoding algorithm based on the proposed data structure; given a N-Square code, the search time for finding the source symbol is $O(\log n)$ [5].In section 2 of the paper, we present the overview of proposed data structure. In section 3, we propose a N-Square approach for compression. In section 4, we present the decoding technique and elucidate how the proposed approach is better than the existing one. In section 5, we conclude our paper and summarize our recommendation.

II. THE PROPOSED DATA STRUCTURE

The occurrences of probability of pixels $p = \{p_0, p_1, \ldots, p_{n-1}\}$ with frequencies $f = \{f_0, f_1, \ldots, f_{n-1}\}$ for $f_0 \geq f_1 \geq \ldots \geq f_{n-1}$, where fi is the frequency of probability pi [5]. Using the N-Square algorithm to erect the N-Square tree T, to obtain the codeword ci, $0 \leq i \leq n-1$, for probability pi determined by traversing the path from root to the leaf node allied with probability pi, where the left branch is corresponding to ‗0' and the right one is corresponding to

G. SRINIVAS PRASAD REDDY P.V.G.D** K. GAYATRI DEVI* *Dept. of IT, GITAM University, AP, srinivas.gitam@gmail.com, gayatridevi07@gmail.com **Dept. of CS & SE, Andhra University, AP.*

_1‘,_2‘,etc. Let the level of the root be zero and the level of the other node is equal to summing up its parent‘s level and one. Codeword length l_i for p_i can be known as the level of p_i. Then the weighted external path length $\sum_{i=0}^{N-1} fili$ is minimum. For example, the N-Square tree corresponding to the pixels $\{p_0, p_1, \ldots, p_{22}\}$ with the probabilities {0.3, 0.2, 0.1, 0.1,0.065, 0.05,0.04, 0.02, 0.02, 0.02, 0.01, 0.01, 0.01,0.01, 0.009, 0.008, 0.007, 0.006, 0.005, 0.004, 0.003, 0.002, 0.001} is shown in Fig. 1. The codeword set C $\{c_0, c_2, \ldots, c_{22}\}$ is derived as {0,2,3,4,11,12,14,101,102,103,104,130,131,132,133,134,1000,1001,1002,1003,1004,10000,10011}, where the length set L$\{l_0, l_1,\ldots..l_{22}\}$ is {1, 1, 1,1, 2, 2, 2,3, 3, 3, 3, 3, 3, 3, 3, 3, 4,4,4,4,4,5,5}. In addition, the codeword set composes of a space with 2^d addresses, where d = 5 is the depth of the N-Square tree [5].

III. N-SQUARE ENCODING

The erraticism of Non binary Huffman coding stems from the fact that there is no definite pattern for the value of D. For the same number of samples the method shows different results for different D values. There are no limitations expressed on the value of D to be taken. Therefore one can take the maximum value of D for which the message can be sent in a one step auxiliary ensemble which is quite disgusting as the burden of adding large number of messages is an overhead to the encoder. So in our approach we follow a pattern or structure for fixing the value of D and eradicating the erraticism of the existing approach. The number of initial messages to be added has also been used differently which ultimately reduces the average no of bits per message.

A. N-Square Approach For D Value

Here the number of samples N plays a major role i.e. the value of D will be fixed depending on the number of samples.

Step 1-Find the range of N-Square in which our N lies where N takes values from 1 through n.

Step2-Assign D to the upper boundary of N of the range.

Let us examine an example, for suppose our N=20. Then N lies between 4^2 and 5^2. So the range is [4 5]. The upper boundary of the range is 5. Therefore the value of D=5.

Table 2: shows N values.

N^2	N
1^2	1
2^2	4
3^2	9
4^2	16
5^2	25
6^2	36
.	.
N^2	N

Thus in this way we are following a pattern instead of taking arbitrary values for D. As we got all the required elements i.e. N, D lets start our ensemble coding. And as usual the first step will be finding the initial no of messages to be added in the original ensemble (say k) and is given by the following formula.

J=2 + [(N-2) mod (D-1)] K=J+D-1 (7)

Here K indicates the least number of messages above D to be added in the initial step in order to make the penultimate ensemble to contain D messages. And from the later steps we keep on adding D least probable messages to form a composite message. It will be noted that the terminating auxiliary ensemble always has one unity probability message. Each preceding ensemble is increased in number by D-1 until the fist auxiliary ensemble is reached. Therefore if N1 is the number of messages in the first auxiliary ensemble, then (N1-1) / (D-1) must be an integer. However N1=N-K+1.Therefore K must be of such a value that (N-K) / (D-1) is an integer. We are calculating K and all these just to ensure that in the penultimate ensemble the number of messages should be equal to D. If in our problem when J=D then no need to calculate K. We continue by adding D bits in each pass. The decoding is done by assigning bits 0 to D-1 in order, to each of the brackets. But since the initial number of messages to be added i.e. K is more than D we repeat the cycle 0 to D-1 again for the remaining extra K-D messages. The whole method would get cleared with the following illustration of an example.

Table 1. An Example of N-Square Encoding

Gray level value	Probability P_i	Code word C_i
G4	0.3	0
G7	0.2	2
G1	0.1	3
G17	0.1	4
G6	0.065	11
G2	0.05	12
G8	0.04	14
G14	0.02	101
G20	0.02	102
G11	0.02	103
G22	0.01	104
G13	0.01	130
G9	0.01	131
G15	0.01	132
G16	0.009	133
G5	0.008	134
G18	0.007	1000
G25	0.006	1001
G10	0.005	1002
G21	0.004	1003
G12	0.003	1004
G3	0.002	10000
G19	0.001	10011

The codeword‘s generated from the N-Square tree could be treated as suffixes, which obey the suffix property, i.e., no codeword is the suffix, or start of the code for another codeword. For example, c_0 could be treated as an interval from address 00000 to 11110, and c_{20} starts from 00102 to 11102. Since there is no empty branch in the N-Square tree,

each address is occupied by exactly one interval of the suffix [5].

IV. DECODING ALGORITHM

We can achieve the decoding procedure by N-square tree. The detailed algorithm is listed below. The time complexity for decoding is O(log n)[5].

Step 1-Interval Generation Algorithm:

Table 3 : Neighboring Interval Representation of symbols in Table 1 [5]

Interval	Gray level value	Starting Address
int_o	G4	00000
Int_1	G7	00002
Int_2	G1	00003
Int_3	G17	00004
Int_4	G6	00011
Int_5	G2	00012
Int_6	G8	00014
Int_7	G14	00101
Int_8	G20	00102
Int_9	G11	00103
Int_{10}	G22	00104
Int_{11}	G13	00130
Int_{12}	G9	00131
Int_{13}	G15	00132
Int_{14}	G16	00133
Int_{15}	G5	00134
Int_{16}	G18	01000
Int_{17}	G25	01001
Int_{18}	G10	01002
Int_{19}	G21	01003
Int_{20}	G12	01004
Int_{21}	G3	010000
Int_{22}	G19	010011

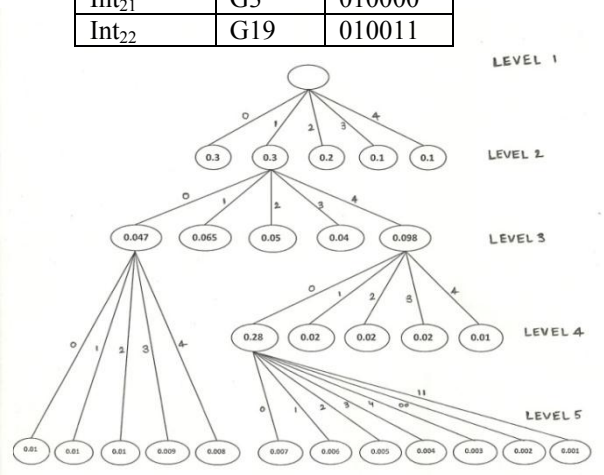


Figure 1: An Example of the N-Square tree

Input-Construct d level N-Square tree

Output-sort N intervals.

Step 2- Generate a tree equivalent to the encoding tree.

Step 3-Scan the input characters or string from left to right and go left on a 0 and go to right sub-tree on reading 1, 2, and 3 until a leaf is reached.

Step 4-Output the probability of pixels encoded in the leaf node and return to root.

Step 5-Continue with step2 until the input is empty.

V. CONCLUSION

In this paper we introduced a method for lossless compression of images as an efficient one. We saw that the proposed approach eliminates the erraticism of non binary Huffman coding regarding the value of D. It introduced a pattern for finding the D value as per the value of N through N-Square approach. We also saw that the proposed method outperformed the existing one in terms of entropy as well. The implication of the N-Square is due to its acknowledged use in image and data compression. The main rationale of this paper is to simplify the representation of the N-Square tree and the decoding procedure by the property obscure in the N-Square tree.

VI. REFERENCES

1) —Digital Image Processing", S.Jayaram, S.Esakkirajan, T.VeeraKumar, Tata Mc grawHill Publications.
2) —Digital Image processing", Second Edition, Rafel c. Gonzalez, Richard E.Woods, 6Th Impression 2009, ISBN 978-81-7758-168-3.
3) —Fundamentals of Digital Image processing", S. Annadurai, R.ShanmughaLakshmi, 3rd Impression ,ISBN 978-7758-479-0.
4) DAVID A. HUFFMAN, ASSOCIATE, IRE," A Method for the Construction of Minimum-Redundancy Codes", 1952.
5) Pi-Chung Wang, Yuan-Rung Yang, Chun-Liang Lee, Hung-Yi Chang, —A Memory-efficient Huffman Decoding Algorithm", (AINA'05), 1550-445X/05, 2005 IEEE
6) Cebrail TAŞKIN, Serdar Kürşat SARIKOZ," An Overview of Image Compression Approaches", ICDT.2008.
7) Lu G., —Advances in digital image compression techniques", 1993.
8) Bei Chen, Hong-Cai Zhang, Wen-Lun CAO, Jian-Hu Feng, —Huffman Coding Method based on Number Character", ICMLC 19-22 Aug-2007 IEEE.
9) Reza Hashemian ," Direct Huffman Coding and Decoding using the table of Code-Lengths",ITCC'03 IEEE.
10) Jia-Yu Lin, Ying Liu and Ke-Chu Yi," Balance of 0,1 Bits for Huffman and Reversible Variable-engt Coding", IEEE Transactions on Communications, VOL.52, NO.3, March 2004

An Algorithm To Reduce The Size Of Cipher Text

[1]Mani Arora , Dr. Derick Angles[2]

GJCST Computing Classification
E.3 , E.4

Abstract- **Globalization has increased the degree of connectivity of communication networks. Organizations are spending a huge amount for securing their data and information. The rising competition in the global market has increased the threats to access the sensitive data. Whether data is at rest or mobile data, it is the need of the hour to secure it from unauthorized access. Over the time a number of techniques and algorithms have been proposed for its security. A technique, which is good for one type of application, may not be effective and efficient for other applications. However, all the existing techniques suffer from one common drawback that is the size of the Cipher text, which is always same or more than the size of the Plain text. We in this research paper, qualitatively, emphasize the need of securing the data but keeping the size of Cipher text under check. We propose a new technique and algorithm to encrypt the Plaintext into a reduced size Cipher text.**

Keywords- Algorithm, reduce, cipher, size, text, security.

I. Introduction

Globalization has increased the degree of connectivity of communication networks. Organizations are spending a huge amount for securing their data and information. The rising competition in the global market has increased the threats to access the sensitive data. Unauthorisedly. People are busy in developing [1] new and advance Cryptographic techniques to secure their data. The Cryptography has a history over 4000 years. A single security technique cannot be the best for all the applications. So far in the history of Cryptography only one point is kept in mind while developing a Cryptographic protocol that is —to provide adequate security —The purpose of our paper is to point out the need of modern time Cryptography. This paper emphasizes the requirement of new techniques of Cryptography, which with the basic goal of providing adequate security must also be efficient in reducing the size of Encrypted Text referred as Cipher Text. We have used the term —reduced size cipher text —to highlight the main aim of this proposed technique.In order to develop such a technique one must take care of the fact that Cryptography is an effective and the only practical way to ensure secure communication over open communication networks. While keeping in check the size of cipher text we can't compromise with security of data and information. Further there must be no loss of text in the whole process of conversion of plaintext into reduced size cipher text.There is a reason for this and that is the fact that we can't afford to miss even a single number or character value while dealing with the text.

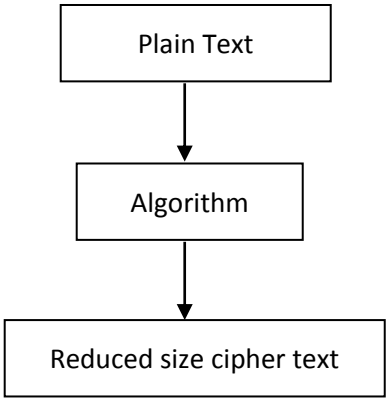
[1]Mani Arora Assistant Professor ,Khalsa College Amritsar (Mobile:9814088687,mani_mcain@yahoo.com)
[2]Dr.Derick Angles Professor,Guru Nanak Dev University (Mobile :9815952913,derickengles@yahoo.com)

II. Need of Proposed Technique

In general every cryptographic algorithm produces a cipher text, which has got the size equal to or more than the size of original plaintext. At the same time, it must be noted that when the cipher is decrypted then we will get the same plaintext, as it was original. But there is need for us to focus at cipher text. We can use it for when, we are sending or receiving it for a secured communication with other parties. For fast communication the use of reduced size cipher with the same message or information with the original text is decisively preferable. Moreover it will help in reducing the communication cost involved as well as of storage. When we are dealing with text each & every bit counted carefully, the loss of even a single bit can change the meaning of the text subsequently causing it misleading.

III. Introduction to Proposed Technique

The proposed technique intends to encrypt the plain text with a prechosen mathematical function along with the objective of reduced size cipher text.



The proposed algorithm works as in figure. In the technique we have maintained two dictionaries.

A. Primary Dictionary

The dictionary contains words, which are probably most frequently used along with the codes*(which will be explained later). The dictionary will be fixed in size & the codes too. Even if someone cracks the dictionary and codes, still our technique is going to work because it is based on both primary and secondary dictionary and secondary dictionary is not fixed.To maintain this dictionary plain text file is used and for processing we read the file into memory as an associative arrays. Associative arrays maps arbitrarily typed objects to arbitrarily typed objects. Data structures used to represent associative array when initialized in

memory will be linked list. For searching a word in the dictionary, simple linear search is used.

B. *Secondary Dictionary*

The dictionary is not fixed. This dictionary will be empty when initialized. Every time the algorithm come across the string in pass 2,it will add it to dictionary and assign a code to it. Starting code for first string in secondary dictionary will be fixed. The next codes can be obtained by doing increment of one step. As this dictionary will be created during runtime so it's difficult to crack the encryption. This dictionary will also be stored in same file that for primary dictionary is used.

IV. ENCRYPTION ALGORITHM

The algorithm will start with initializing a primary dictionary and variable S, which will be initially empty. We read the file containing dictionary into memory in linked list structure. The variable S will read plain text data file word by word and comparisons in dictionary will be made with help of this variable.

In the algorithm each word of plain text will be first searched in primary dictionary, this is done using linear search. If the word is present in primary dictionary then it will be replaced by corresponding code assigned to it and stored in encoded output file. Codes are in binary as binary take less memory space than any other data type. Codes are fixed for each data word so it can't be changed later on. Size of code allocated to each string will be 12 bits that is any dictionary can contain maximum 4096 entries

If the word is not found in primary dictionary then it is searched in secondary dictionary. If the word is present in secondary dictionary its corresponding code will be substituted to encoded output file, if not present then that word is substituted in secondary dictionary and the new code is generated to it by incrementing the code by 1 of last entry in secondary dictionary. The code for first entry in secondary dictionary will be fixed and rest codes will be obtained by incrementing each code by 1.Secondary dictionary will be of variable size.

After these pass we will get encoded output file. Further in the algorithm we use a predetermined mathematical function XOR to further encrypt the encoded output. As code assigned to each word is of 12 bits, so in pass3 each 12 bit block will be XOR with a secret key to get final output which will be a cipher text in reduced size then plain text.

So In Whole The Algorithm For Proposed Technique Is

i. Initialize S as an empty string
ii. Read primary dictionary in memory
iii. Do till EOF Read the next word from the file If this word is in primary dictionary get corresponding code from primary dictionary.Write the code in the output file else Read secondary dictionary in file

If this word is in secondary dictionary see the corresponding code from secondary dictionary Write the code in the output file else add this word to secondary dictionary

assign the next code obtained by incrementing previous code by 1 used to substitute the new code in the output file

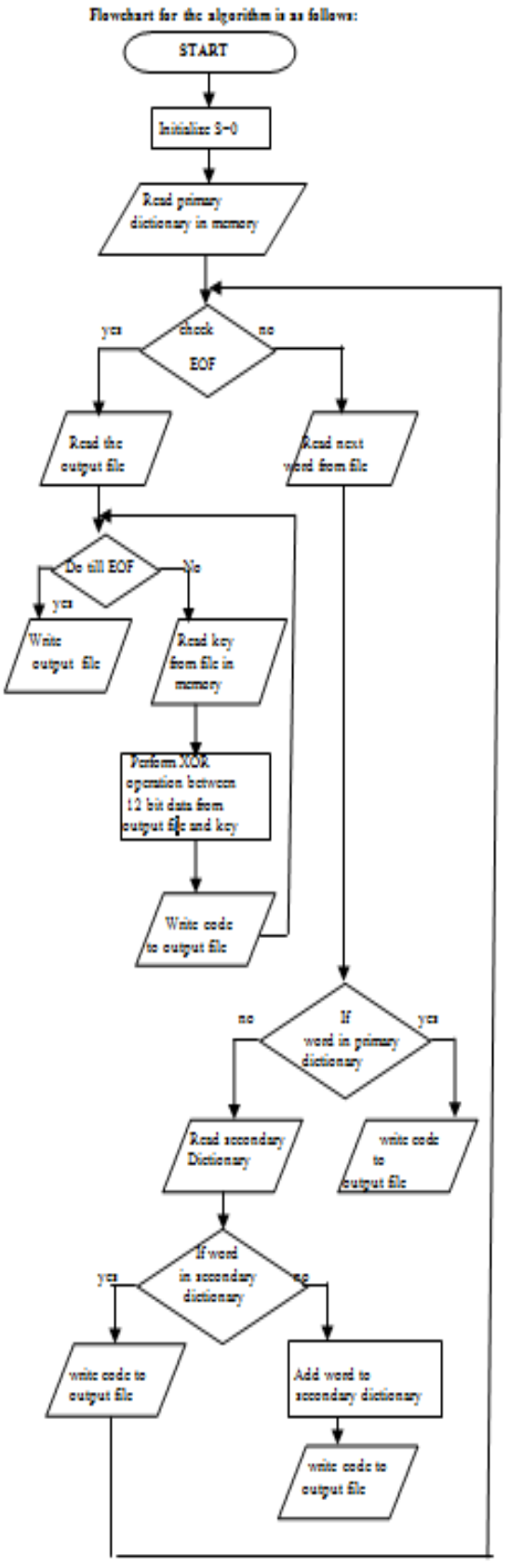
iv. Read the output file
v. Do till EOF Read key from file in memory Read next 12 -bit block from file Perform XOR operation between key and 12-bit block Write the code in final output file

Algorithm for Primary dictionary

Primary dictionary is fixed so sender and receiver both before transmission of cipher text will know it.

Algorithm to create primary dictionary

i. Start with prechosen code and assign first word to it.
ii. Add new word to dictionary.
iii. Assign code to new word by incrementing previous code by 1



Example-

Example of this technique is given below.

Plain text that is to be encrypted is-

The information cannot be understood by anyone for whom it was unintended. Confidentially is the protection of transmitted data from passive attacks. The other aspect of confidentiality is the protection of traffic flow from analysis. This requires that an attacker not be able to observe the source & destination frequency length or other characteristics of the traffic on a communication facility:

Pass1

000000010001	000100000001
000000110011	000000001011
000100000010	000000001010
000100000011	000000010100
000000101110	000000001100
000000010010	000100000100
000000111001	000100000101
000000000100	000000010001
000100000110	000000001000
000100000111	000100001000
000000100010	000100001001
000100001010	000000111001
000000010001	000000000100
000100001011	000000001000
000100000101	000000000100
000000010001	000100000110
000000001000	000100001100
000000100010	000100001101
000000111001	000000100001
000100001110	000000100000
000000000011	000100001111
000000010101	000000001011
000000101001	000000001101
000100010000	000000010001
000100010001	000000001110
000100010010	000100010011
000000001111	000100010100
000000000111	000000101111
000100010101	000000001000
000000010001	000100010110
000000001001	000000000001
000100010111	000100011000

Pass2 In this pass each 12-bit block will be XOR with key, which is 111100001111

111100011110	111000001110
111100111100	111100001100
111000001101	111100000101
111000001100	111100011011
111100000001	111100000011
111100011101	111000001011
111100110110	111000001010
111100001011	111100011110
111000001001	111100000111
111000001000	111000000111
111100101101	111000000110
111000000101	111100110110
111100011110	111100001011
111000000100	111100000111
111000001010	111100001011
111100011110	111000001001

111100000111	111000000011
111100101101	111000000010
111100110110	111100101110
111000000001	111100101111
111100001100	111000000000
111100011010	111100000100
111100100110	111100000010
111000011111	111100011110
111000011110	111100000001
111000011101	111000011100
111100000000	111000011011
111100001000	111100100000
111000011010	111100000111
111100011110	111000011001
111100000110	111100001110
111000011000	111000010111

Contents Of Key File

Primary Dictionary

Code	String
0000 0000 0001	a
0000 0000 0010	as
0000 0000 0011	an
0000 0000 0100	is
0000 0000 0101	in
0000 0000 0110	if
0000 0000 0111	or
0000 0000 1000	of
0000 0000 1001	on
0000 0000 1010	by
0000 0000 1011	be
0000 0000 1100	it
0000 0000 1101	to
0000 0000 1110	&
0000 0000 1111	,
0000 0001 0000	can
0000 0001 0001	the
0000 0001 0010	was
0000 0001 0011	are
0000 0001 0100	for
0000 0001 0101	not
0000 0001 0110	and
0000 0001 0111	has
0000 0001 1000	one
0000 0001 1001	two
0000 0001 1010	too
0000 0001 1011	its
0000 0001 1100	who
0000 0001 1101	his
0000 0001 1110	her
0000 0001 1111	were
0000 0001 0000	that
0000 0010 0001	this
0000 0001 0010	from
0000 0001 0011	into
0000 0001 0100	with
0000 0001 0101	most
0000 0001 0110	here
0000 0001 0111	back
0000 0001 1000	form
0000 0001 1001	able
0000 0001 1010	must
0000 0001 1011	where
0000 0001 1100	which
0000 0001 1101	there
0000 0001 1110	whom
0000 0001 1111	other
0000 0011 0000	did
0000 0011 0001	should
0000 0011 0010	could
0000 0011 0011	cannot
0000 0011 0100	does
0000 0011 0101	he
0000 0011 0110	she
0000 0011 0111	—
0000 0011 1000	‘‘
0000 0011 1001	.
0000 0011 1010	always
0000 0011 1011	almost
0000 0011 1100	also
0000 0011 1101	national
0000 0011 1110	International
0000 0011 1111	like

Secondary Dictionary

0001 0000 0001	Information
0001 0000 0010	understood
0001 0000 0011	anyone
0001 0000 0100	unintended
0001 0000 0101	confidentially
0001 0000 0110	protection
0001 0000 0111	transmitted
0001 0000 1000	data
0001 0000 1001	passive
0001 0000 1010	attacks
0001 0000 1011	aspects
0001 0000 1100	flow
0001 0000 1101	analysis
0001 0000 1110	requires
0001 0000 1111	attacker
0001 0001 0000	observe
0001 0001 0001	source
0001 0001 0010	destination
0001 0001 0011	frequency
0001 0001 0100	length
0001 0001 0101	characteristics
0001 0001 0110	traffic
0001 0001 0111	communication
0001 0001 1000	facility
1111 0000 1111	Key

V. Conclusion

In this paper we have presented a completely new encryption algorithm that focuses on reducing the size of cipher text. Such technique will be helpful in saving the communication cost as well as storage space required in computer.

VI. REFERENCES

1) Mc Conell, Mike. & Harmittion, Booz Allen . (2002) —Information Assurance in twenty-first century — IEEE Security and Privacy, pp. 16 -18.
2) Adams, C.M.& Tavares, S.E. (1993) —Designing S-Boxes for Ciphers Resistant to Differential Cryptanalysis —proceedings of 3rd Symposium on state and Progress of Research in Cryptography, pp.181-190.
3) D.K. Branstad, D.K., Gait, J., and Katzke, S. (1976) —Report on the Workshop in support of computer security" in National Bureau of standards, p77-91.
4) Rijndael (1998) "Rijndael AES proposal" National institute of science and technology. Available [online] http://www. Csrc.nist.gov/encryption/aes/.
5) 5.Rivest, R.L. , Shamir, A. & Adlenman, L.M. (1978) —A method for obtaining digital signatures and Public-Key Cryptosystem —, Communications of the ACM, Vol.21, No.2, pp.120 –126 .
6) Schneier, B. (2001) –Applied Cryptography", (John Wiley and Sons) p11-13.

Two Way Clock Scheme In Pipeline To Minimize The Clock Skew

GJCST Computing Classification
C.1.3, C.1.1

[1] N. Suresh Kumar, [2]D.V.Rama KotiReddy, [3]B. Pramod Kumar, [4]A. Chaitanya Prabhu

Abstract- **In most of the digital systems the clock skew decreases the performance of the digital systems in terms of providing good sensitivity or to maintain data synchronization. In the conventional pipeline system it is facing problems due to improper synchronization of clock pulses. This is a universal problem in all the digital systems mostly called jitter or skew. The propagation of information in the digital systems mainly controlled on the basis of clock pulses. In most of the digital systems the clock skew decreases the performance of the digital systems. Here a new system is implemented in the path of the clock to reduce the clock skew.**

Keywords- Clock skew, Pipeline, Microcontroller.

I. INTRODUCTION

In conventional Pipeline technology the clock signal arrive at the first register take the same time as the data takes time to arrive the first register. But the clock signal

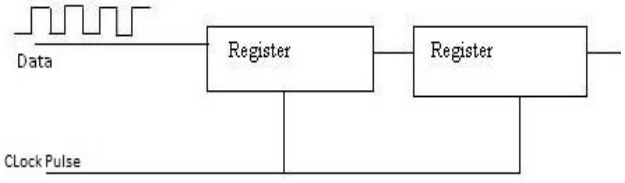


Figure 1 Conventional pipeline System

may take more time than the data to reach the second stage of the pipeline. So there may be a chance of data overlapping at the second stage of the pipeline.But the conventional pipeline system is facing problems due to improper synchronization of clock pulses. This is a universal problem in all the digital systems mostly called jitter or skew. In conventional pipeline systems the clock signal is derived as[2]

$$T_{clk\_conv} \geq D_{max} + D_r + T_s + \Delta_{clk} \quad (1)$$

Where D_{max} is the maximum propagation Deay D_r is clock-to-output delay of the pipeline register T_s, T_h is pipeline register setup and hold time In the present work a new system is proposed in the path of the clock to remove or reduce the clock skew. There are already few methods effectively working on clock skew such as wave-pipelining [5] and Mesychronous pipeline [1] methods.The idea of wave-pipelining [5] was originally introduced by Cotten [7], who named it *maximum rate pipelining* Cotton observed that the rate at which logic can propagate through the circuit depends not on the longest path delay but on the difference between the longest and the shortest path delays. As a result, several computation —waves," i.e., logic signals related to different clock cycles, can propagate through the logic simultaneously. The system clocking must be such that the output data is clocked *after* the latest data has arrived at the outputs and *before* the earliest data from the *next* clock cycle arrives at the outputs. Critical speed-limiting factors in wave-pipelining [5] are the uncontrolled clock-skew, the sampling time of registers, and the worst case transition time at the logic outputs. While the minimization of these factors has been a major challenge in the design of conventional high-speed pipelined systems as well, the equalization of path delays comes as a new challenge for the design of wave-pipelined systems. Different clock signal paths can have different delays for a variety of reasons [8]. Differences in delays of any active buffers within the clock distribution network may cause un-synchronization of data and clock in wave pipeline method.

Smaller clock periods are achieved in wave pipelining [2][5] by reducing the maximum propagation delay (D_{max}) by splitting the stages into number of stages. The clock signal is derived in the wave pipelining is

$$T_{clk_w} \geq (D_{max} - D_{min}) + t_h + T_s + 2\Delta_{clk} \quad (2)$$

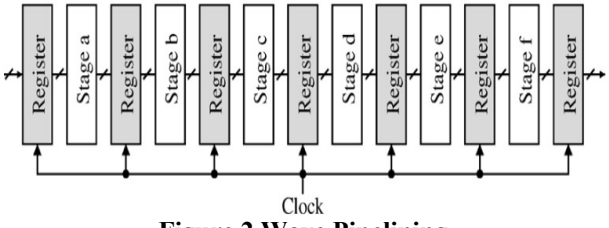


Figure 2 Wave Pipelining

And further the propagation delay in is reduced and the clock synchronization is controlled by introducing a delay element in the path of clock signal of Mesynchronous pipelining [1]. This delay will be equal to the delay created by the pulse passed from one stage to other stage of the pipeline. The system is clocked such that a pipeline stage is operating on more than one data wave simultaneously. At any given time, multiple waves can be present in a stage and the waves are separated based on physical properties of internal nodes in the logic stage. The clock signal is derived in the Mesynchronous pipelining is

$$T_{clk\_m} \geq \left(D_{\max(j)} - D_{\min(j)}\right) + t_h + T_s + 2\Delta_{clk} \quad (3)$$

[1] N. Suresh Kumar, [2]D.V.Rama KotiReddy, [3]B. Pramod Kumar, [4]A. Chaitanya Prabhu
[1] Department of IT, GITAM University, Visakhapatnam, India
[2] College of Engineering, Andhra University, Visakhapatnam, India
[3] Venkateswara college of Engineering, Srikakulam, India
[4] Sri Sivani Institute of Technology, Srikakulam, India
1nskgitam2009@gmail.com

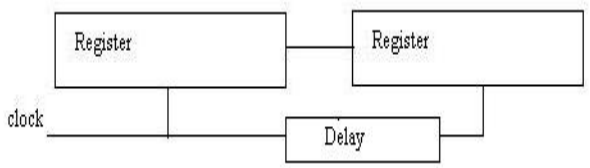


Figure 3 Mesynchronous pipelining

From equation 1, 2, 3 it implies that

$$T_{clk_m} \leq T_{clk_w} \leq T_{clk_{conv}} \quad (4)$$

It derives that mesochronous pipelining delivers an improved performance compared to conventional pipeline scheme.

II. ENHANCED METHOD

In the present work an external pipeline is developed using external registers and the working of these registers are controlled and observed by a microcontroller. The present system performance is almost similar to mesynchronous pipeline performance which will produce,

$$T_{2way\ clk\_p} \leq T_{clk\_w} \leq T_{clk\_conv}$$

When a single clock pulse is applied to manage the data transmission through the registers in the pipeline. But it will create a clock skew in the pipeline which will decrease the data speed from one stage to other stage. The pulses from the encoder are fed into the first register when clock pulse is applied to the first stage of the pipeline. The pulse will be passed to the next stage after applying the clock pulse to the next stage. The clock pulse path is directly given to the registers where the encoder pulses are passes from one stage to another stage through flip flops of the registers. This may create a problem of overlapping of pulses in the first stage before it enters into the next stage.

In the present work a new method is proposed to introducing clock pulses alternatively to the parallel pipeline stages. Here two pipelines alternatively work to carry the data pulses. The two pipelines are selected alternatively by activating the clock pulses alternatively. The clock pulses are supplied by port pins of microcontroller instead of external clock generator. The port pins alternatively supply the clock pulses to the pipelines. The alternative supply of clock pulse will create a delay in the data propagation. This will synchronize the data propagation. In the register hold time of the first pipeline the register set time will be enabling in the second pipeline. Mean while the data will be output through the first register. While the register holds time of second pipeline the third clock pulse will be given to first pipeline. Now the next pulse from the encoder enters into first stage of first pipeline. And the out from the first stage of first pipeline will enter into second stage of first pipeline. So the encoder pulses enter into pipeline alternatively by changing the clock pulses from first pipeline to second pipeline vice versa. The clock signal is derived in the present system pipeline method as,

$$T_{2way\ clk\_p} \geq \left(D_{\max(j)} - D_{\min(j)}\right) + t_h + T_s + 2\Delta_{clk}$$

The system is clocked such that pipeline stages are operating on more than one data pulse simultaneously. The present system at any given time, multiple data pulses can be present in a stage similar to the mesychronous pipelining.

III. HARDWARE

In the present work two pipelines are individually connected and controlled through the port pins of microcontroller. Each pipeline has integrated with two stages. Each stage is connected with a single pipeline. These pipeline stages are alternatively controlled by p1.0 and p1.1 as shown in figure5.

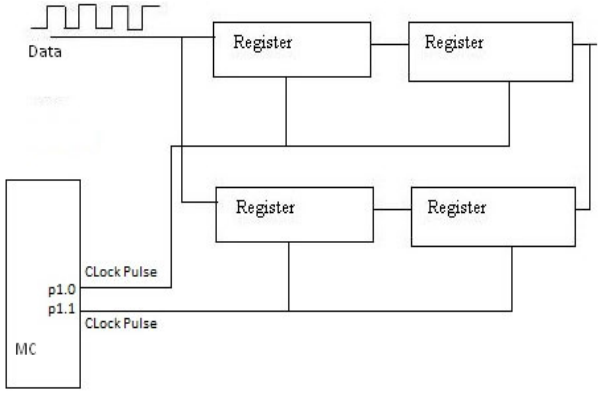


Figure 4 A New method to control parallel pipeline

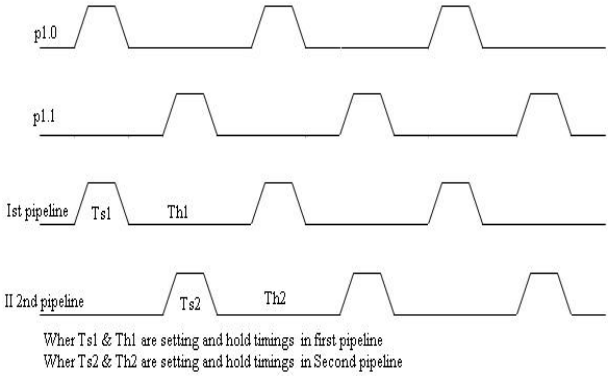


Figure 5 Data entering into first and second pipelines at alternate clock pulses applied through p1.0 and p1.1

IV. SOFTWARE ROUTINES

Step1-Send high and low signals alternatively through parallel port to enable clock input to the pipelines

Step2-Enable P1.0 to enable the first pipeline

Step3-Disable P1.0 and Enable P1.1 to enable second pipeline

Step4-Disable p1.1 and repeat step 2 to step 4 in the entire process.

V. CONCLUSIONS

The relative error of the frequency measurement almost zero when compare with other method measurements. The system can also measure the small readings which are nearer to zero value. The clock distribution becomes simpler by controlling clock signals by internal ports of microcontroller. Power consumption is small when compared to other methods. A simple control is implemented on the clock skew. Simple software logic is used to control the ports of microcontroller to generate the clock signals higher performance can be achieved using

smaller number of pipeline stages. The proposed pipeline scheme operates on more than one pulse simultaneously.

VI. References

1) Suryanarayana B. Tatapudi, Student Member, IEEE and José G. Delgado-Frias, Senior Member, IEEE,A Mesychronous high performance digital systems, VOL. 53, NO. 5, MAY 2006
2) Thomas gay, ―Timing constraints for wave pipelined systems" IEEE transactions on Computer aided design of integrated circuits, vol13, no.8, august 1994
3) Jabulani Nyathi, ―A high performance hybrid wave pipelined linear feedback shift register with skew tolerant clocks", IEEE, 1384- 1387, 2004
4) Mohammad Maymandi, ―A digital programmable delay element: Design and analysis", IEEE transaction VLSI systems, Vol.11, no.5, October 2003
5) David E. Duarte, ―A Clock Power Model to Evaluate Impact of Architectural and Technology Optimizations", IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS, VOL. 10, NO. 6, DECEMBER 2002
6) Wayne P. Burleson, ―Wave-Pipelining: A Tutorial and Research Survey", IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS, VOL. 6, NO. 3, SEPTEMBER 1998
7) L. Cotten, ―Maximum rate pipelined systems," in Proc. AFIPS Spring Joint Comput. Conf., 1969.
8) EBY G. FRIEDMAN, ―Clock Distribution Networks in Synchronous Digital Integrated Circuits", Invited paper, PROCEEDINGS OF THE IEEE, VOL. 89, NO. 5, MAY 2001 665
9) Douglas V hall, : Microprocessor Interfacing: Programming and hardware", 2nd edition, TMH.
10) G.A.Gibson, Micro Computer systems, The 8086/8088 family Architecture, PHI, 2nd edition
11) Barry B. Brey, ― The Intel Microprocessors, 8086/8088, 80186/80188,80286,80386,80486 and pentium processors, Architecture, programming interfacing"
12) Keneth J. Ayala, 8086 Microprocessor, Penram International/Thomson, 1995.
13) A.K. Ray , Burchandi, Advanced Microprocessors and Pheripherals", Second edition, Tata McGraw-Hill.

Grid Enabled Architecture For DWDM Network Design And Optimization Tool

[1]Bhupinder Singh, [2]Seema Bawa

GJCST Computing Classification
B.2.1 , C.2.4

Abstract- **High bandwidth networks are indispensable to support the present ever increasing demand for various services on internet and to cater highly bandwidth extensive application such as video streaming and multimedia conferences. Dense Wavelength Division Multiplexing (DWDM) technology based optical networks provide potentially large transmission capacity that has an obvious advantage from both technical and economic perspectives.DWDM network design and optimization tools are developed to aid the designing and deployment of networks. Design tools play an important role in facilitating routing and wavelength assignment, filter placement, DCM and amplifier placement with the aim of minimizing the overall cost of network, in minimum amount of time. Designing and optimizing large networks require lot more processing power than a single desktop machine can actually provide. This scarcity of processing power results in either sub-optimal or infeasible solutions. This paper presents the grid enabled architecture for DWDM network design and optimization tool, which aims to harness the processing power of existing idle resources in an organization to quickly provide the optimized design for huge networks.**

Keywords- DWDM network design, Grid Computing, optical network design, Grid applications.

I. INTRODUCTION

Need for the optical networking is driven primarily by three major technological advancements: 1. Dense Wavelength Division Multiplexing (DWDM) 2. Broadband optical amplification and 3. Wavelengthgranularity optical switching. Optical networks can span from hundreds to thousands of kilometres and involve thousands of different network elements [1]. Analysis and design of optical network require the solution of nonlinear equations from optical physics to combinatorial solution of NP- hard [2] optimization problems. Optical networks can also require significant investment, ranging from millions to hundreds of millions of dollars. Last but not the least,increased competition in the business also demands for an optimized design (both cost effective and energy efficient) as quickly as possible. All these reasons strongly push the demand for computer aided software tools to design, optimize, simulate, and operate DWDM networks. These tools address several aspects of optical network design, from helping sales personnel to provide optimized bids to clients,

[1]*Bhupinder Singh, Module Lead, Design Tools Department, Ciena India Pvt. Ltd. Gurgaon (Haryana), India.*
(bsaini@ciena.com)
[2]*Seema Bawa, Professor, Computer Science and Engineering Department, Thapar University Patiala. (Punjab), India .*
(seema@thapar.edu)

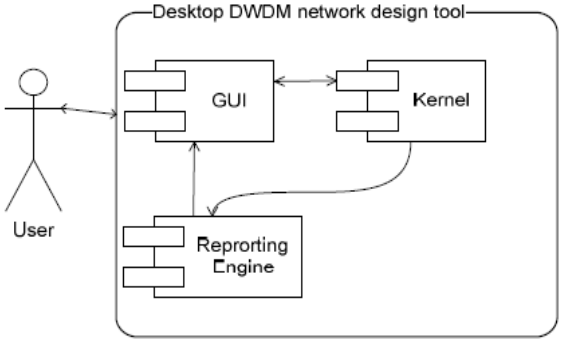


Fig. 1: High level architecture of a Desktop based DWDM design tool

to providing a detailed selection of optical components, to planning wavelength growth in operating networks. DWDM network design tools are basically Graphical User Interface (GUI) based desktop software as depicted in the Fig.1. Flow of DWDM network design process is depicted in the flow chart given in Fig. 2.

It mainly consists of the three components. First, GUI based canvas that facilitates the users to draw the physical layout of the network and specify the demands. Second major component is Kernel, which is the implementation of the set of algorithms that try to optimize the overall network design, so as to fulfil the requested demands, as well as minimize the overall cost of the network. Finally third component is the reporting engine, which provides the different type of reports and network diagrams that help the sales person to place their bids and also help the deployment engineers during the actual installation of the network.

Routing and wavelength assignment (RWA), optimized DCM placement and optimized amplifier (EDFA or Raman) placement, are compute intensive problems [3]-[5]. While designing very large networks, the need for computing resource and memory grows exponentially. This cannot be fulfilled by desktop machine due to very limited availability of the resources. This leads to either infeasible or suboptimal solutions which are not cost effective. Alternate solution is to transfer the processing logic to a high end,multi-processor server machine with parallel processing capability and huge memory. But this solution is accompanied by the increased cost of server hardware.

Grid computing [6] is an emerging trend for making easy access to computing resources like an electric power grid.Inspired by the electrical power grid‘s pervasiveness, ease of use, and reliability, computational grid provides an analogous infrastructure for wide-area parallel and distributed computing. A grid enables the sharing, selection, and aggregation of a wide variety of geographically

distributed resources including supercomputers, storage systems, data sources, and specialized devices owned by different organizations for solving large-scale resource

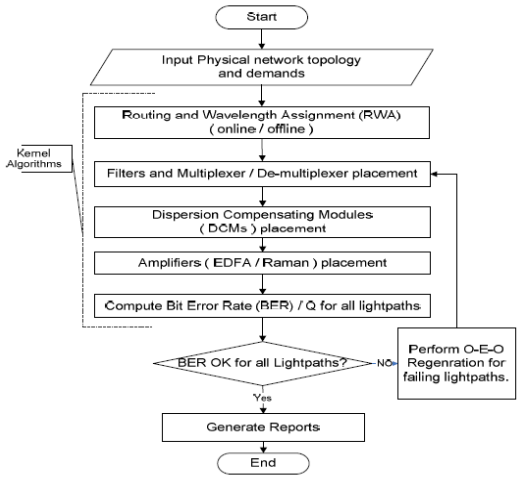


Fig.2: Steps of DWDM network design

intensive problems in science, engineering, and commerce. Since DWDM network design and optimization is also A resource intensive problem, hence it is a suitable candidate to be tackled in grid environment.This paper proposes a grid enabled architecture for optimized DWDM network design. Use of existing procedures and components makes this architecture practically realizable with little extra effort. The proposed architecture aims to harness the power of idle resources in an organization, and use it for DWDM network design and optimization.

II. PROPOSED GRID ENABLED ARCHITECTURE

The proposed grid enabled architecture is an extension to existing desktop based DWDM design tool. Desktop based software is enhanced to a grid application, where processing is seamlessly distributed to many idle computing resources that are registered with a grid. The results computed by different processing resources on the grid are combined by the grid scheduler and are sent back. Details are discussed in next subsections. Proposed grid enabled architecture is depicted in Fig. 3.

A. Pre-Requisites For Proposed Architecture

Before discussing any further details, first we need to see how DWDM network design and optimization problem qualifies as a suitable candidate to be executed as grid application. Since grid environment, by nature, is loosely coupled, with heterogeneous resources and high latency networked environment, so grid application should be such that their work units can be parallelized into a number of independent computation units and should have high computation vs. communication time ratio [7].First, is already supported by the discussion in the previous section that DWDM design and optimization algorithms (like RWA, DCMs and Amplifier placement) are compute intensive and require a lot of processing power. To support the second point we propose to formulate the compute intensive algorithms as a linear programming (LP) / mixed integer linear programming problem (MILP). Details on how to formulate RWA and amplifier and DCM placement into LP / MILP problem is discussed in the [3]-[5]. Providing solution to large MILP problems is again A compute intensive problem. Many commercial LP / MILP solvers like CPlex, GurObi, Xpress etc. and non commercial solvers like COIN-OR‘s CBC exist that support parallel processing. So network design and optimization application qualifies to be designed as a grid application.Next point of discussion is how to distribute the parallel threads over the grid. Most solvers can only harness the power of more than one processing elements in a single machine but cannot use the processing capabilities of idle processing elements in other machines. To achieve this General Algebraic Modelling System (GAMS) [8] is used, which is specifically designed for modelling linear, nonlinear and mixed integer optimization problems. GAMS is a high-level modelling system for mathematical programming and optimization. It consists of a language compiler and a stable of integrated high-performance solvers. GAMS is tailored for complex, large scale modelling applications. A large part of the time required to develop a model is involved in _data preparation and transformation‘ and report preparation. Each model requires many hours of analyst and programming time to organize the data and write the programs that would transform the data into the form required by the mathematical programming optimizers. Reference [9] describes the GAMS for modelling optimization problems on a grid computing environment. This framework is easy to adapt to multiple grid engines and can seamlessly integrate evolving mechanisms from particular computing platforms. So, by integrating with any of the existing MILP optimizers, GAMS provides a lightweight, portable and powerful framework for optimization on a grid.

B. Information And Control Flow In Proposed Architecture

It is clear from the discussion in the previous sub-section that required components / procedures already exist in one form or another. A slight modification in existing procedures and a little integration effort paves the way to A very powerful and a cost-effective grid enabled network design and optimization tool. Fig. 3 shows the proposed grid enabled architecture.

GUI- provides a canvas or other input interface to provide physical network layout and demands as input to the software. It is also used to display the reports and layout diagrams to the user, that are prepared by the reporting engine

Kernel- consists of a set of algorithms as discussed earlier assuming the compute – intensive algorithms like RWA, DCM placement and amplifier placement etc. are formulated as LP or MILP.

Reporting Engine is responsible for collecting the results from the kernel and prepare different type of reports (e.g. bills of materials, network design and deployment diagrams etc.), as required by user.

Kernel-GAMS- interface provides the way to map MILP based formulation of kernel algorithms to GAMS model (A form understandable by GAMS language complier) and calls GAMS APIs (application programming interface) to solve the model. Also it returns the solution back to the kernel.

GAMS- provides a platform that accepts the input model and converts it into a form understandable by the integrated optimizer that actually does the job of solving the GAMS model. As soon as a solve statement is encountered while executing a GAMS program, control is passed to the script.This script is responsible for running the optimizer on the problem instance and passing back the solution to the GAMS.

Grid Engine- manages a pool of connected computers available as a common computing resource. It provides the effective sharing and utilization of idle computing resources and massive parallel task execution. Scheduler (e.g. condor, sun grid engine, globus, etc.) handles all management tasks. Machine (M1-Mn) is a set of machines that are registered with grid engine to share the computing resources. Any machine in the organization can be a part of a grid environment. Grid scheduler utilizes the idle CPU cycles from these machines.

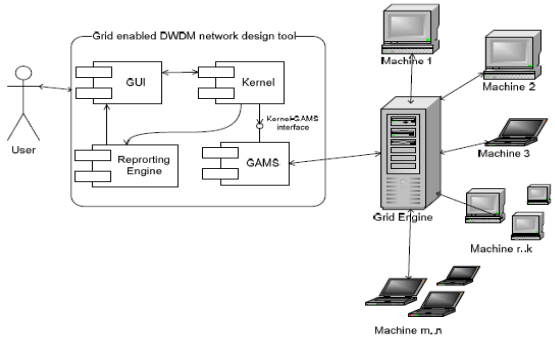


Fig. 3: Proposed grid enabled architecture

Information exchange and control flow among various components is as follows: GUI based canvas used to draw the physical layout of the network and provide it as input to the kernel along with the demands. Kernel executes a series of algorithms to design the networks. For compute intensive algorithms formulated as MILP, it prepares a GAMS understandable model and calls the GAMS APIs to solve this model via Kernel-GAMS interface. As soon as a solve statement is encountered while executing a GAMS program, control is passed to the script. This script is responsible for running the optimizer on the problem instance and passing back the solution to the GAMS. In a grid environment, GAMS use the file system to give each instance its own environment and its own directory. The script then schedules the solver execution. The submission script centralized all information required to tailor the system to A specific grid engine. Grid engine distributes the processing over the different machines, connected to the grid. The results are collected by the grid engine and returned back to the GAMS, which assembles the solution for the submitted model and returns to the kernel for further processing. As soon as the network design and optimization process is complete the designed network is used by the reporting engine to prepare different type of reports as required by the user.

III. Conclusion

A grid enabled architecture for DWDM network design and optimization tool is presented. This simplified architecture can be realized as an extension to existing desktop based network design software with slight modifications. The use of existing procedures and integration of existing components paves a way to the practical realization of very powerful and cost effective software with little extra cost and integration effort. Hence the proposed grid enabled architecture aims to harness the power of idle resources in an organization, and use it for DWDM network design and optimization. Effective utilization of the idle resources caters the exponentially growing need for computing resources for designing huge optical networks. Massive parallel processing capability provided by the computational grid cuts down the need for very high end and costly server hardware.

IV. References

1) Chekuri, C. et.al. (2006), Design tools for transparent optical networks, Bell Labs Technical Journal, vol. 11, Pp. 129-143.
2) Garey, M. R. and Johnson, D. S. (1979), Computers and Intractability: A Guide to the Theory of NP-Completeness,ed. W. H. Freeman, New York,NY.
3) Ozdaglar, A.E. and Bertsekas, D.P. (2003), Routing and wavelength assignement in optical networks", IEEE/ACM Transactions on Networking (TON) archive, vol. 11, Pp.259-272.
4) Ramamurthy, B., Iness, J. and Mukherjee, B. (1998), Optimizing amplifier placements in a multiwavelength optical LAN / MAN: the equally powered-wavelengths case, Journal of Lightwave Technology. vol. 16, Pp.1560-1569.
5) Minagar, A., Premaratne, M. and Tran, A. V. (2006), Cost optimal allocation of amplifiers and DCMs in WDM ring networks, Optics Express, vol. 14, Pp. 10278-10291.
6) Berstis, V. (2002), Fundamentals of Grid Computing Available: http://www.redbooks.ibm.com/abstracts/ redp3613.html.
7) Nadiminti, K., Luther A. and Buyya, R. (2005), User Guide for Alchemi 1.0 Available:http://www.gridbus.org/ ~alchemi/files/1.0.beta/docs/ AlchemiManualv.1.0.htm.
8) The GAMS System. Available: http://www.gams.com/docs/ intro.htm.
9) Bussieck, M. R., Ferris, M. C. and Meeraus, A. (2009), Grid-Enabled Optimization with GAMS, (2009) INFORMS Journal on Computing. Vol. 21 (3), Pp: 349-362.

Characteristics Analysis Of Network Non-Optimum Based On Self-Organization Theory

Ping He

GJCST Computing Classification
D.4.6, I.2.6

***Abstract*- This paper introduces some known non-optimum to the networks security, categorizes the non-optimum, and analyses protection mechanisms and techniques for countering the non-optimum. The non-optimum have been classified more so as definitions and then followed by the classifications of these non-optimum. Also mentioned are the protection mechanisms. The paper establishes the syndrome and empirical analysis based on the non-optimum category of the network system. At the same time, it also puts forward the non-optimum measurement of the networks system along with non-optimum tracing and self-learning of the networks systems.Besides, the various characteristics and functions of the network security can be measured from the non-optimum attributes. By summing the practice, this paper has also come at non-optimum analysis principle of the networks, established the conception of non-optimum thresholds and put forward three theorems about non-optimum parameters. Through the concept of extensionality networks function, it analyzes the actual significance of networks security based on non-optimum analysis. Based on the analysis from non-optimum to sub-optimum, it puts out the academic idea of extension networks optimal. Meanwhile, it discusses about the general framework of extension optimum. Finally, according to the previous practice of optimization, kind of method has been developed to learning approach the sub-optimum from non-optimum network.**

Keywords- non-optimum category; network security; non-optimum and security; extensionality networks function; self-learning approach.

I. Introduction

The future security of societies that depend increasingly on networks is contingent upon how our complex human and technical systems evolve. New network technologies including the Internet favor fragmentation into many loosely connected open and closed communities governed by many different principles. As the reach of today's networks has become global, they have become the focus of arguments over the values that should govern their development [1].

However due to the complexity of network's practice, there are numbers of unknown and uncertain factors, longitudinal and transverse relationship of things, people's networks behavior. Especially as the network systems heads to the orderly dynamic condition, some of the hidden troubles are not exposed, the achieved most optical modes are in unstable states. This implies that the recognition and practice of mankind is featured by the exploration and pursuit not only in an optimum category, but also, under many conditions, in a non-optimum category. That is to say when people are faced with urgent problems, they need not only to find out the most optimum mode or realize the most optimum aim, but also, more importantly, to get rid of the vicious influences of non-optimum accidents effectively as well as control the non-optimum factors of the network system [2].

The concept of non-optimum was introduced by He Ping in his classic paper [1]. Using the concept of non-optimum Literature [2] introduced the non-optimum analysis. Literature [3] introduced the notion of sub-optimum sets; Literature [4, 5, 6, and 7] studies the sub-optimum learning system. This approach provided a wide field for investigation in the area of system optimization and its applications. Continuing the work in [1-7], as an extension of concept presented in [6, 7], we will introduce network security based on the characteristics of non-optimum. And, we will establish their properties and relationships with other classes of early defined forms of non-optimum to sub-optimum.

This paper is structured as follows: The second section introduces the non-optimum concepts of network system, and related research reviews theoretical principles relevant to network system like characteristics of non-optimum, analysis on technology acceptance model (TAM). The third section network security architecture studies based on non-optimum analysis, at the same time, the conversion mechanism of non-optimum to sub-optimum mode. The fourth puts out the approach of extension networks optimal. Meanwhile, it discusses learning model the optimum from non-optimum network Finally the fifth section conclusion puts forward the discoveries of this research and future research direction.

II. Non-Optimum Analysis Of Network System

A. Basic Concept

The theory of non-optimum analysis on systems and the tracing of optimum modes are interrelated and inter-perforated, and stand reciprocally contradictory. The former expresses the escape from non-optimum category and the latter displays the exploration of the most optimal mode and its procedure. Based on the interrelationship of the two

Ping He
Department of Information Liaoning, Police Academy
Dalian, 116036, P.R. China e-mail:heping2000@163.com

research areas, the formation of non-optimum category and the constraint of non-optimum are the foundation to establish the optimum category. It means that only when man does the research out of the non-optimum category, can they be on their way to trace the most optimal modes. The concept of non-optimum is quite comprehensive in network system. From the viewpoint of network systems' software, non-optimum means unfeasible and unreasonable [2]; from the viewpoint of human' network behavior, it means non-trusted [3]; from the viewpoint of network systems' capacity, it means ineffective and abnormal [4]; from the viewpoint of network systems' change, it means obstacles, disturbance and influence [5]. There exists a serious of non-optimum problem from the entity of the network system to the change of the network system, which causes non-optimum category. As to every kind of networks security problems, there is the individual non-optimum category as well as the common non-optimum category. The so-called individual non-optimum category is decided by the characters of the networks relationship system, while the common non-optimum category is an objective entity of networks behavior. At present, most security analysis is designed manually based on past experience of their networks behavior. Since the number of possible optimization model very large for realistic applications of reasonable complexity, security analysis modeling designed manually may not work well when applied in new problem instances. Further, there is no systematic method to evaluate the effectiveness of security designed manually. For these reasons, a ―eooperative" method for discovering the proper security decision for a particular application is very desirable. This leads to the development of our method for extensionality security model (ESM) of non-optimum category.The security of the network system emerges and develops in non-optimum category. Every security attributes exists in a non-optimum category, and the real actions of the non-optimum tell its risk phenomena of networks systems. Generally speaking, these risk phenomena are included in the non-optimum category of network system, but since the network system is rather complex, it takes on certain unclear attributes under any condition. The unclear attributes are unknown things possessed by the risk system of network security, which are decided by the complexity of the risk system in numerical value. For example, the risk analysis of the network system is much more complicated than the physical system. Therefore the unknown things of an network system are much more than a physical system. These unclear attributes cause the risk factors of the system. Figure 1 shows the relationship between the non-optimum and the network security.

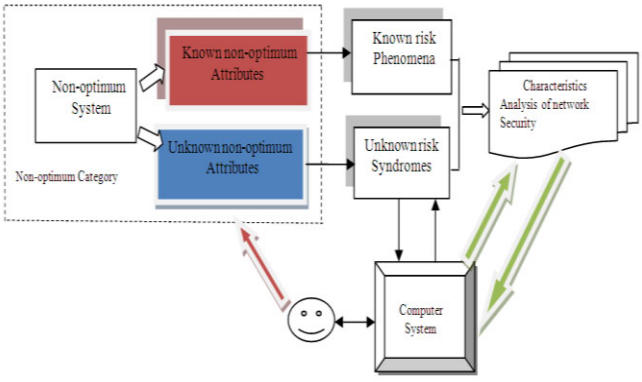


Fig.1. Characteristics analysis of networks security based on non-optimum

B. Experience analysis

Network's experience provides non-optimum syndrome for the network system. When the recognitions are different, the non-optimum syndromes are different as well. The tracing to the network's behavior and conditions of the past can propose a non-optimum syndrome. In an artificial network system, different people have different be saviors and stories, thus different experiences. Sometimes experiences are called a kind of recognitions; but as the level of recognition is different, the experience of the network is also different. The syndrome of the system is selected and decided by the experience of the network, and the reasonability of the experience's selection is also a meaningful question for discussion. For example, the increase of the function of the network can reduce the non-optimum category, and the changes of the network's behavior can cause new non-optimum factors, which change with the network's behavior. Thus the non-optimum category of the actual system is composed of non-optimum syndrome, the amount of non-optimum changes and the potential non-optimum factors. Under the prerequisites of the formation of the network's experience, there is a process of recognition to the non-optimum behavior, which is a self-learning and self-accustomed process. Natural non-optimum is an objective entity, which does not change with people's will. However, when people get hold of the basic characteristics of the non-optimum, they can set up certain functions to avoid the non-optimum, which is not the main subject of the non-optimum analysis theory of the system [3]. From the creation to the death of the network system, there is an overall running procedure. In fact, a whole, standard running condition does not exist, and also breaches the development regulation of things. From the viewpoint of the dialectic from recognition to entity, this also accords with the entity and recognition to the non-optimum. For example, as a decision-maker of a concern, one first needs to do a series of work related to the management of the concern and the strategic development objectives. That is to say, to find out what methods to take, what problems to solve and what difficulties to conquer. The key to finish this series of work is to correctly find out the non-optimum problem that exists meanwhile with the objectives. Of course, these non-optimum problems are formed by direct

experience, indirect experience and partial hypotheses. Mentioning hypotheses, people might ask: Can hypotheses be hypotheses? Can they be replaced? These doubts are unnecessary. The actual research shows that if there is no hypothesis, there is no affirmation; to accept a hypothesis is to confirm; the acceptable effect is in the direct ratio of the affirmation. Most of the chemical systems are set up on hypotheses, which importance is obvious. Mathematics is also the conclusions made by logistic reasoning inference discursiveness discussions based on hypotheses.

III. Self-Organization of Network System Security

A. *Self-Organization Of Network System*

In the research of the self-organization theory of systems, the transmission of order and non-order is a core question. The theory of dissipation structure, the theory of hyper circulation, synergetic theory and chaos theory contribute a great deal to it. In fact, their individual theories include non-optimum theory of the system. Because the major character of the self-organization of the system is to perfect the running of the system, develop its goals, they have to experience from non-optimum to optimum, and from optimum to non-optimum. If the system is not featured with this attribute, it doesn't need self-organization either. Analysis shows that systems always stay on the border of optimum and non-optimum, and the aim of self-organization is to bring the system from the border to the optical category. There is a time limitation on the system's stay in the optimal category. Within a certain time, because the system is stable, it stays in the optimal category. However, if the system is not stable, it will soon move from the optimal category to the new border and cause a sustained situation of the system. The sustained situation is neither a developing situation, nor an ideal situation. Of course, the actual angle of the system doesn't have the most optimal criteria, and it is also not necessary to make sure what is most optimal. As long as the system can shorten the time of moving from the non-optimum category to the border and from the border to the optimum category, the system is satisfactory.

If the system is able to realize the transit, it has a good self-organizing capacity. As is known from the self-organization theory, profound changes will not influence the system, and only the huge changes composed of profound changes might cause the evolution of the system. This conclusion can make the non-optimum control of the system effective, and they system will stay naturally in the optimum category, or on the border. People can achieve the self-organization function on the border, e.g. the organization is open, exchanges energies with the outside. Thus the function and behavior of the system change and new non-optimum control comes into being. Then, the system goes back to the optimal category. The self-organization through coordination and super-circulation can let the system replicate and consummate itself and reach the optimum category. (It still needs be emphasized that the optimum category shows the category that can be controlled by the system's non-optimum).

Researchers found out that non-optimum system should be set up at the same time with the optimum system. The non-optimum here refers to the one against the optimum. Non-optimum system is decided by all the incompatible problems and limits inside and out side the system, which influences directly or indirectly the system's executive process and final goal. Whether a system falls in the optimum category is also decided by the incompatibility and degree of limitation, which in turn decide the non-optimum degree of the system (non-optimum degree).

Only when the system has the full control and adjustment capacity on its incompatibility and degree of limitation, is the system on the tracing of the most optimum, and this is the new research for the self-organization theory. The bases of the non-optimum analysis theory are obtained from the hyper-circulation theory. The hyper-circulation theory can feed a lot of random effects back to its jumping-off point, which represents the start of the system's circulation, and make themselves a reason of the maximum. A highly orderly macroscopically functional organization can evolve through self-replicating and self-selecting. This kind of self-replication and self-selection is realized in the non-optimum and optimum hyper-circulation. The entity of non-optimum calls for the optimum category under certain demanding conditions. The measurement of optimum and non-optimum category is the core of system optimization.

There are two sides to everything, and the final direction of the network can be only achieved through practice and the transition of the two sides. The state of the network behavior decides its goal by choosing between optimum and non-optimum. Hence the non-optimum problem is illustrated through the following method:

Suppose N_o represents an optimal system of network behavior, N_{no} a non-optimal system of network behavior. No matter it is an optimal system or a non-optimal system, they are all composed of network system objective O, and G is the function subsets of the network system, E is the environment of the network system. As to the optimal system, if the structure Π (O, G, E) of the sub-optimum (cannot make sure of the condition of optimum and non-optimum) of the network system S composed of objective O, function G and environment E meets the following conditions:

The objective of the network system can be attained;

The function of the network system can be achieved;

The environment of the network system can be controlled.

Then S is called an optimal network system. The attainability of the objective of the system shows that the distance between the recognized goal of the system and the actual goal of the system is acceptable. The achievability of the function of the system refers that the actual functional resources is near to the objective-required resources. The controllability of the environment of the system refers to the self-organizing capacity or the order parameters achieving the permitted value. Suppose O_r is recognition goal of the network system, O_s acts as the actual goal of network system, α represents the difference of the value between O_r and O_s, which shows the degree of acceptance of the

goal of the system. G_s acts as the system's actual functional resources, G_r acts as the resources demanded by the system's objective, and β expresses the functional measurement value between G_r and G_s; The entropy of the actual system $e \leq \gamma$, and γ expresses the system's standard entropy. Thereby for the system's sub-optimum structure Π (O, G, E), if there are ε, δ, ε, (random minimal discrepancy can be accepted) causing $|\alpha - \alpha_0| \leqq \varepsilon$, $|\beta - \beta_0| \leqq \zeta$, $|\lambda - \lambda_0| \leqq \eta$ to hold at the same time, the system S is an optimal system. $\alpha_0, \beta_0, \lambda_0$ is the border value of the system's optimum and non-optimum, and thereby the gathered assemble Π ($\alpha_0 \beta_0 \lambda_0$) is the criteria of the system's non-optimum analysis. In the actual system analysis, under certain selected standards (ε, δ, ε is known), for α, β, λ, when man can't obtain α_0, β_0, λ_0, the system S is called non-optimum system.The above is the overall description of the non-optimum problem of the network system, which tells how to decide the overall frame-saw in the non-optimum system. However, different measurement and means have to be applied in different systems to solve actual problems. Proper quality and quantity determining methods are applied in the actual system analysis. Furthermore, artificial intelligence and expert system reasoning tools can play important roles in non-optimum system analysis. One of the emphases on the non-optimum analysis theory is to describe the borders of the optimum and non-optimum category quantitatively. Because the borders change with objective conditions and subjective desire of mankind and human being has different behavior parameters, they always appear as uncertain under dynamic. Meanwhile, because of the continuous progress of mankind's practice and recognition, under cooperating of the widely exchanged scientific information, the borders might become certain and describable during the dynamic changes. As to the judgment of the reasonability and accountability of the described borders, it is no a theoretical problem, but a problem of selecting the methods and checking the practice. In addition, when analyzing the problems of the network system through quantitative methods, a lot of relationship parameters need to be statistically analyzed and attributably appraised. In many aspects, the influences of the system's non-optimum are depended largely on the experience accumulated in the recognition of the system. That is to say, experiential analysis plays an important role in the non-optimum system analysis, which reflects the meaning and function of the combination of the nature and quantity evaluation.

B. Network Risk Analysis Based On Non-Optimum

There are two situations in risk analyzing of network system: the inherent non-optimum attribute under stable conditions of a network system is decided by the function of the system; the non-optimum attribute under unstable situations of the system is obtained through statistic analysis. That is to say, risk the process of a systems development, non-optimum factors effect on the system, which causes a relationship that does not exist when the system is stable, and it is called non-optimum-born relationship. Every system has to have a non-optimum-born relationship; otherwise, the system goes into risk when it is unstable. For instance, in a strategic decision-making of a large finance corporation, how to build up a non-optimum-born relationship is the key of the corporation's survival and development. It works as this: through the yearlong experience of the corporation, a stable non-optimum area is formed (according to certain experience-decision effect of each year), through which the reasons of unstable factors of the system can be reflected and non-optimum genes found. Of course, there is non-optimum genes everywhere in the system and what we need are the major genes, which are the major factors that cause the system risk to fluctuate to a certain extent. In the actual analysis of the system risk, some factors have direct relationship with the non-optimum genes, some indirect. More relative factors are more influenced by non-optimum genes. Therefore, the factors can be divided into the major non-optimum effect and the minor non-optimum effect. Minor non-optimum effect is influenced by other factors. The core of the tracing to the risk happened from building up non-optimum syndrome of risk and non-optimum cause of formation. The syndrome cannot really become the influence, and the actual non-optimum indeed influences the system, both of which come from non-optimum syndrome. There is a procedure of diagnosis from the syndrome to the cause of formation. The diagnosis happens when the behavior of the system finishes, and includes: the cause of foundation of non-optimum from the major syndrome; the cause of foundation of non-optimum from the minor syndrome, which is the overall framework of the tracing to non-optimum. Two types of mapping F_{I} and F_{II} can be established:

$$F_{\mathrm{I}} : S \to D, \quad F_{\mathrm{II}} : A \to K,$$

where $S=\{s_1,s_2,...,s_m\}$, which is the aggregate of the major syndrome, s_m $\{m=1,2,\ldots,g\}$ is every detailed major non-optimum aggregate, $D=\{d_1,d_2,...,d_n\}$ d_j $\{j=1,2,...,n\}$ acts as every detailed major non-optimum syndrome, $A=\{a_1,a_2,\ldots,a_r\}$ is the minor non-optimum syndrome aggregate, a_u $\{u=1,2,\ldots,r\}$ acts as every detailed minor non-optimum syndrome, $K=\{k_1,k_2,\ldots,k_v\}$ acts as the non-optimum aggregate of the system, ,k_l $\{l=1,2,\ldots,v\}$, is every detailed non-optimum of the system.

If system's non-optimum is shown as $W=\{S, A\}=\{W_1, W_2, \ldots, W_{g+r}\}$, when given a group of non-optimum input $W_i \subseteq W$, under the above two mapping effects, the relative sorts of the non-optimum of the system $d_j \in D$ and the output of the non-optimum factors $K_j \in K$ can be gotten.

Before deciding the characteristics of the two mappings, the non-optimum syndrome drawn from experiential material statistics need to be divided into major syndrome and minor syndrome. The principles of deviation are decided by how much information the non-optimum syndrome can provide, concerning the recognition of the system's non-optimum. Suppose:

$$P(d_j \mid w_i) = \frac{N_{d_j}}{N_{w_j}}$$

When syndrome W_i appears, non-optimum d_j causes the conditional probability of the system, where N_{wi} acts as the number of times that syndrome W_i arises, and N_{dj} acts as the number of times that non-optimum d_j occurs as non-optimum appearance when the syndrome W_i arises. More

$$P(w_i) = \{P(w_1), P(w_2), \cdots, P(w_{m+r})\}$$

is the probability distribution of syndrome W_i, that is

$$P(w_i) = \frac{N_{w_i}}{N_w},$$

Where N_w is the overall times that all the syndromes occur, N_{wi} acts as the number of times when syndrome i occurs. The above-mentioned N_{dj}, N_{wi}, N_w can be drawn from empirical statistics of resources. Thus, the entropy function of the relative $P(D_{wj})$ is

$$H(D \mid w_i) = -\sum_{j=1}^{n} P(d_j \mid w_i) \text{ lb } P(d_j \mid w_i)$$

The correspondent mean entropy is

$$H(D \mid W) = \sum_{i=1}^{m+r} P(w_i) H(D \mid w_i) \circ$$

Choosing dependable level β to make the syndrome Wi of

$$\frac{H(D \mid w_i)}{H(D \mid W)} \le \beta$$

the required major syndrome , it is decided that the non-optimum of the system d_i is the long symptom, and the others short symptom are all minor syndrome. As to relative syndrome w_i, w_g,

$$H(D \mid w_i) = H(D \mid w_g)(i \ne g),$$

both W_i and W_g belong to minor syndrome. The level of selection is decided by the utilization rate of the experiences. The higher the rate is, the greater is the level of selection. [4,5]

After the analysis of non-optimum symptoms of the system, non-optimum syndrome can be set up, which provide useful information for the real-time analysis of the system. In fact, there are different non-optimum symptoms in people's minds. As an excellent decision-maker, he has to possess good recognition capacity of non-optimum symptom; otherwise, he is simply not able to control the system.

When analyzing the non-optimum symptom of the system two aspects have further to considered: one is the non-optimum symptom within the system, and the other is the non-optimum symptom outside the system (the environment). As to a secluded system, it only has the non-optimum symptom within the system. Therefore, the non-optimum symptom will most probably influence the system, which is an alternative and interrelated process with the behavior of the system. According to the definite rules of the development of things, the number of times that the non-optimum occurs provides evaluating opportunities to the behavior of the system. For example, the traders in the markets face these situations when they do the business. How the four parties of A, B, C, and D agree to a bargain depends on whether they have had the experiences of trade, except for some basic conditions of deals. If trader A has experienced some kinds of unfeasible, unsatisfactory and unfavorable non-optimum situations, his result is relatively dependable. Obviously, if B doesn't have these kind of experiences, his result of trade cannot be as good as A's. In fact, it is impossible that all of the experiences are balanced, and as for the absolute balancedness, the non-optimal experiences are not taken into account. Actual analysis shows that the traders don't have the same non-optimum recognition, which is because every dealer has different objective, behavior attribute and environment. If they do possess the same non-optimum recognition and similar entity conditions, it can be balanced under certain conditions. Except for the analysis of the system of the past, the key is to analyze the system's present and future situation, where the dynamic and opening characteristics of the system have to be taken into account.

IV. Conclusion

This paper presents a technique for analyzing network security using non-optimum analysis on systems. The inputs to the system are suitable sub-optimum sets representing linguistic values for network security goals of congeniality, integrity and availability. The non-optimum analysis was constructed using the fuzzy reasoning in order to adequately analyze the inputs. It might also be necessary to use an adaptive non-optimum analysis for security risk analysis. We have been able to design a system that can be used to evaluate the security risk associated with the production of secure network systems. This will definitely help network system organizations meet up with the standard requirements. A technique for assessing security of network system before final deployment has been presented.

The result of this study shows that if the network producing companies will incorporate security risk analysis into the production of network system, the issue of insecurity of network system will be held to the minimum if not eliminated. This study has also revealed that if each of the network security goals can be increased to the maximum, then the level security will also be increased and the risk associated will be eliminated. Finally, security risk analysis is a path towards producing secure network and should be considered a sub-optimum activity by network systems self-organizations.

V. References

1) He Ping, Theories And Methods of Non-optimum Analysis on systems, Journal of Engineering Science, 2004, vol.1, no. 1, pp. 73-80.
2) He Ping, "A Guiding System for Non-optimum Recognition", Journal of Control and Decision, 1989, vol.3, no1, pp. 17-21.
3) Myagmar, A. J., S.and Lee and Yurcik,W. Threat modeling as basis for security requirements. In Symposium on Requirements Engineeing for Information Security (SREIS), 2005.
4) Noopur, D. and Samuel, T. R. Processes to produce secure software. A paper written for, Cyber Security Summit Taskforce Subgroup on Software Development Lifecycle, 2004.
5) Sheyner, O. andWing, J. Tools for generating and analyzing attack graphs. In Proceedings of formal methods for component and Objects, 2005.
6) He Ping, Method of system non-optimum analysis in crisis management, Proc. of Second International Conference of Information System for Crisis Response and Management, 2007, pp. 640-645.
7) He Ping, The Method of Non-optimum Analysis on Risk Management System, In: Bartel Van de Walle, ed, proc. of the Int'l conf ISCRAM, 2008, pp. 604-609.
8) He Ping, Methods of Systems Non-optimum Analysis Based on Intervenient Optimum, In: Qifan Wang, ed, proc. of the Int'l conf on System Science, Management Science and System Dynamic, 2007, pp. 661-670.
9) Smith, E. and Eloff, J. Cognitive fuzzy modeling for enhanced risk assessment in a health care institution. IEEE Intelligent systems & their applications, 15(2), 2000.
10) Smith, E. and Eloff, J. Transaction based risk analysis using cognitive fuzzy techniques. Advances in Information Security Management&Small Systems Security, 2001.
11) Sodiya, A. S., Onashoga, A., S., and Oladunjoye, B. A. Threat modeling using fuzzy logic paradigm. Journal of Issues in Informing Science and Information Technology.U. S. A, 15(4):53.61, 2007.
12) Sodiya, A. S., Onashoga, S. A., and Oladunjoye, O. B. Towards building secure software products. Journal of Issues in Informing Science and Information Technology, U. S. A, 13:635.646, 2006.
13) Zadeh, L. A. Fuzzy sets. Information and Control, 1965.
14) Ping He. The method of attribute analysis in non-optimum to optimum. Journal of Liaoning Normal University.2008, Vol.31, No.1, pp. 29-33.
15) Ping He. On theory and practice of reliable innovation system. Journal of Liaoning Normal University.2009, Vol.32, No.1, pp. 119-122.
16) Zengtang Qu, Ping He, Method of Non-optimum Analysis on Risk Control System, Journal of Software, 2009, Vol.4, No. 4, pp.374-381.
17) Libo Hou, Ping He. Approach of Non-optimum Analysis on Information Systems Security, IEEE Computer Society, IITA International Conference on Control, Automation and Systems Engineering, 2009, pp.225-228.
18) Zengtang Qu, Ping He, Intelligence Analysis Based on Intervenient Optimum learning Guide System , IEEE Computer Society, International Conference on Computational Intelligence and Natural Computing, 2009, pp. 336-339.
19) He Ping, The system is non-optimum distinguished the direction system. The dominates together with the decision, 1989, vol.3, no.4, pp. 18-21.
20) Miao Dongsheng, The system science pick. Beijing: People University of China press, 1989.
21) He Ping, Method of system non-optimum analysis in crisis management, Proc. of Second International Conference of Information System for Crisis Response and Management, 2007, pp. 640-645.
22) He Ping, Method of system non-optimum analysis based on Intervenient optimum, Proc. of ICSSMSSD , 2007.
23) Sebastiani, F, Machine learning in automated text categorization. In: ACM Computing Surveys, 2007, vol. 34, no 1, pp. 1-47

Function Optimization Using Genetic Algorithm By VHDL

GJCST Computing Classification
B.5.2 C.m

[1]D.C. Dhubkarya , [2]Deepak Nagariya, [3]Jay Kumar,

Abstract- **This paper presents the work regarding the synthesis and implementation of a hardware genetic algorithm utilizing very high speed integrated circuit hardware description language (VHDL) for programming FPGAs. Genetic Algorithms were invented to mimic some of the processes observed in natural evolution. The idea with GA is to use this power of evolution to solve optimization problems. They are based on the principles of the evolution via natural selection, employing a population of individuals that undergo selection in the presence of variation-inducing operators such as mutation and recombination (crossover). we solved the problem with the help of hardware description language so it's take less time to find a result as compare to GA's because of HDL solve the problem by parallel processing. The present work deals with implementation and optimization of De jong's first function.**

Genetic algorithms need large memory banks to store the intermediate results and this has made the hardware implementation of GAs very inefficient but by using FPGA our task become simpler. Field-Programmable Gate Arrays (FPGAs) are flexible circuits that can be easily reconfigured by the designer. The program is written in VHDL and compiled with 32-Bit Microsoft Windows and implemented on a Spartan-3A FPGA from Xilinx.

Keywords- Fitness function, Crossover, mutation, random no, population, FPGA.

I. Introduction

A Genetic Algorithm is a search/optimization technique inspired by biological processes such as Natural Selection and Evolution. This project explores the application of such techniques to Parameter optimization problems [17]. A software framework was developed to store and manipulate data representations. Each data representation is a candidate solution to the problem being examined.

These candidates Solutions are grouped together into a family/generation. Starting from generation zero (Initial data), the algorithm iteratively processes each family allowing solutions to ~~–b~~reed", thus creating new candidate solutions. By discarding weak solutions and favoring the reproduction of strong ones the algorithm progressively refines each generation, leading to successive generations containing stronger solutions [10] we are using here VHDL for solving the problem of population control and these populations are randomly generated. So we use random bitgenerator here [7]. After thebit generation we need optimal solution result and also need fitness function. If you have above requirement than we select the best two functions and crossover between them and find the best result after finding the result we follow mutation operation if you find the less value of result as compare to fitness function then good otherwise you can follow these operation again and again.The paper is organized as follows. In section 2, an overview on Genetic Algorithm discussed in Section 3 previous work in brief. Section 4, presents the Hardware model Section 5 Simulation results and the paper is concluded in section 6.

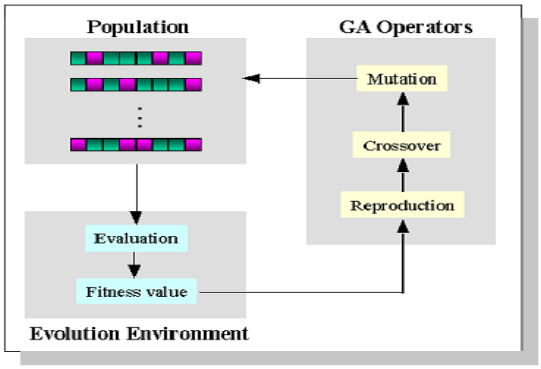


Figure (1) Evolution flow of genetic algorithm [13]

II. Overview of Genetic Algorithms

Genetic algorithms (GAs) were invented by John Holland in the 1960s and were developed by Holland and his students and colleagues at the University of Michigan in the 1960s and the 1970s. In contrast with evolution strategies and evolutionary programming, Holland's original goal was not to design algorithms to solve specific problems, but rather to formally study the phenomenon of adaptation as it occurs in nature and to develop ways in which the mechanisms of natural adaptation might be imported into computer systems [4]

Genetic Algorithms (GAs) are adaptive heuristic search algorithm based on the evolutionary ideas of natural selection and genetics. As such they represent an intelligent exploitation of a random search used to solve optimization problems. Although randomized, GAs are by no means random, instead they exploit historical information to direct the search into the region of better performance within the search space. The basic techniques of the GAs are designed to simulate processes in natural systems necessary for evolution, specially those follow the principles first laid

[1]*D.C. Dhubkarya , BIET Jhansi ,India , dcd3580@yahoo.com*
[2]*Deepak Nagariya, BIET Jhansi ,India , deepaknagaria@gmail.com*
[3]*Jay Kumar, FET RBSAgra,India, jaykumar_1981@yahoo.co.in*

down by Charles Darwin of "survival of the fittest.". Since in nature, competition among individuals for scanty resources results in the fittest individuals dominating over the weaker ones. [11]

GAs simulates the survival of the fittest among individuals over consecutive generation for solving a problem. Each generation consists of a population of character strings that are analogous to the chromosome that we see in our DNA. Each individual represents a point in a search space and a possible solution. The individuals in the population are then made to go through a process of evolution. GAs is based on an analogy with the genetic structure and behavior of chromosomes within a population of individuals using the following steps:

A. Random Number Generation

A linear feedback shift register (LFSR) is a shift register whose input bit is a linear function of its previous state. The only linear functions of single bits are xor and inverse-xor; thus it is a shift register whose input bit is driven by the exclusive-or of some bits of the overall shift register value. The initial value of the LFSR is Called the seed, and because the operation of the register is deterministic, the sequence of values produced by the register is completely determined by its current (or previous) state. Likewise, because the register has a finite number of possible states, it must eventually enter a repeating cycle. However, a LFSR with a well-chosen feedback function can produce a sequence of bits which appears random and which has a very long cycle. [7]

LFSR has two parts shift register & feedback function. A shift register is a device whose identifying function is to shift its contents into adjacent positions within the register. The feedback function is used in LFSR is XOR.[5]

B. Selection

Two chromosomes must be chosen from the population and recombined to produce a pair of new genomes in the new population for the next generation. Randomly choosing two chromosomes would be undesirable, as poor solutions would have an equal chance of being chosen as good solutions. Instead, a method such as roulette wheel selection is used. Roulette Wheel selection has been chosen for this application [14]. This means that the chance of an individual being chosen is proportional to its fitness. Individuals are not removed from the source population, so those with a high fitness will be chosen more times than those with a low fitness.

C. Crossover

With a crossover probability cross over the parents to form a new offspring (children). If no crossover was performed, offspring is an exact copy of parents.[3]

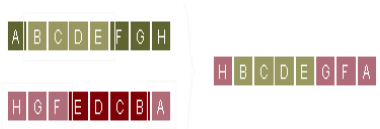


Figure 2 Crossover

D. Mutation

With a mutation probability mutate new offspring at each locus (position in chromosome)[3]

Figure 3 Mutation

III. RELATED WORK

There have been few reported studies on GA hardware implementations, one AHDL description has been announced in {10] but no performance estimations were made. In hardware description of GAPA system containing multiple FPGA chips and multiple digital signal processors was given. In our work we have implemented a GA in hardware using VHDL hardware description language and simulated the performance of the system on FPGA chip [10]

IV. HARDWARE IMPLEMENTATION

Our hardware consists of a Pentium microprocessor 4 which has high-speed PCI bus slots available. This is connected to Xilinx's 3 family of FPGA chip. The population resides in FPGA chip which are flexible RAM.

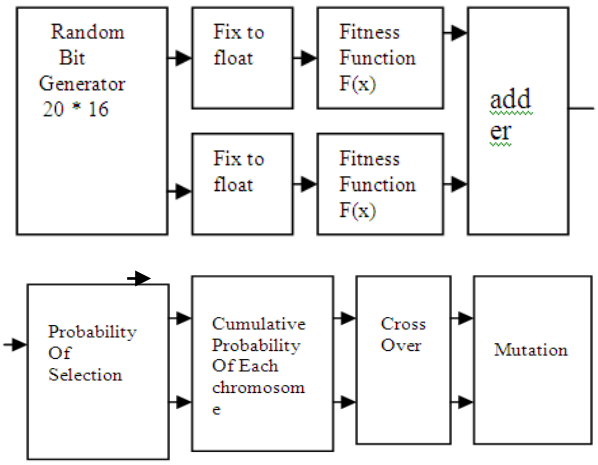


Fig 4 Block diagram of model

Initial Parameters

i. Crossover Rate:100%
ii. Crossover Type: 1 Pt.
iii. Mutation Rate: 2.5%
iv. Population Size: 10
v. Chromosome Length : 40 bit

A. Random Number Generation

In this block all the component of these blocks working to greater on the basic of clock pulse. Whenever we start the

FPGA kit. These all the process depend upon the clock and works as concurrently.
When the clock=0:- Random bit initialized the bit first bit between 0 and one and these bit passes to the solution string block after one clock delay.
When the clock =1:-New random bit generate for the next solution string and solution string passes to the next block and the population

B. Fitness Function

we have used the De Jong's function (1): since It is considered the easiest and simplest test function among De Jong's other functions [13]. It is also called ―TheSphere Model". It is a good example of a continuous, strong convex, unimodel function [15]. The structure of the first is defined as follows:

$$f_1(x) = \sum_{i=1}^{n} x_i^2$$

Fig 5 De Jong's functions

Where fi(x) is Fitness Function , & range for Xi is between 0 to 1.

Implementation

When the clock=0:-1st solution string selected and passed on to block 2nd.When the clock=1:-1st solution string pass to the next block Identifying genes*(X1 to X5) and solution string select a new solution string.
When the clock=2:-Identify genes (X1 to X5)1 pass to next block and convert this value in float form. And 2nd block identify the 2nd solution string and 1st block select 3rd string.
When the clock=3:-3rd block pass to the floating value to the next 4th block this block collect. Identify genes (X1 to X5)2 pass to next block and convert these value in float form. And 2nd block identify the 3nd solution string and 1st block select 4rd string.
When the clock=4:-in this clock 4th block pass to the value to 5th block .this 5th block calculate the some of fitness of each solution string. 3rd block pass to the floating value to the next 4th block this block collect. Identify genes (X1 to X5)3 pass to next block and convert these value in float form. And 2nd block identify the 4nd solution string and 1st block select 4rd string.All the Block works on the basis of clock pulse and each block have one latency time. So all block work after the single clock delay.

C. Selection

All the Block works on the basis of clock pulse and each block have one latency time. So all block work after the single clock delay.
When the clock=0:-fitness of the 1st solution string pass on to block 2nd (probability selection) and at the same clock 1st fitness goes to the fit sum block.
When the clock=1:-this block we have two in coming node 1st value come to the fitness block(2) and other value come to the fit sum (3)block and this block divide the value (fitness/fit sum).
When the clock=2:-this block calculate the cumulative vale of the fitness. it has single incoming node. The value comes from the 2nd block. And this time 2nd block divide the 2nd fitness.
When the clock=3:-4th block passes the cumulative value and this clock also involved to generate the new ten random number in between 0 to 1. And all previous block working in same nature in same clock.
When the clock=4:-in this clock value 6th block take the value from the 5th block and generate the parent

D. Cross Over

The algorithm for crossover operates as follows: The first chromosome is split in two at the crossover point. Both halves are then matched with the second Chromosome to find the longest common subsequence. Matching is done as follows: The first half is matched with the first character from the second chromosome, then with the first two characters and so on. The second half is matched with the last character from the second chromosomes, then with the last two characters and so on. The results of each match are stored in an array. The code in the fig implements the crossover part of GA.

```
begin
ncp<=CONV_INTEGER(cp);
ncp1:=ncp;
if (ncp1>38) then ncp1:=38; end if;
c1<=p1(39 downto ncp1) & p2(ncp1-1 downto 0);
c2<=p2(39 downto ncp1) & p1(ncp1-1 downto 0);
```

E. Mutation

Mutation is implemented by toggling randomly bit. Here mutation rate is 100% but can be varied easily. The code in the fig implements the mutation part of GA.

```
process(clk)
variable nm3:integer;
begin
nm3:=conv_integer(nm2);
if(nm3=40) then nm3:=39; end if;
sout<=sin(39 downto nm3+1)&(not sin(nm3))&sin(nm3-1
downto 0);
end process
```

V. Results

A top-down design methodology was adopted. A High-level VHDL model for the circuits was generated.[2] The logic was partitioned. Each part was re-described in a lower level description (RTL) required for the circuit synthesis, optimization and mapping to the specific technology by assigning current FPGA family and device. The resulting optimized circuit description was verified through extensive simulation. The proposed design was coded in VHDL. It

was functionally verified by writing a test bench and simulating it using ISE simulator and synthesizing it on Spartan 3A using Xilinx ISE 9.2i.[16]

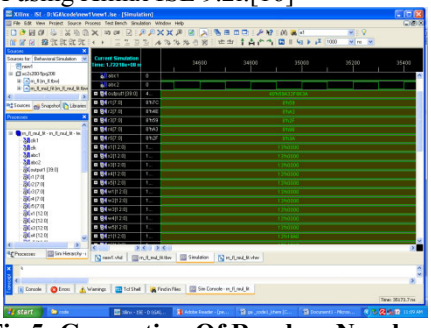


Fig 5: Generation Of Random Numbers

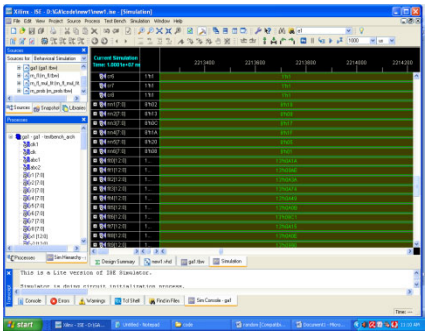


Fig 6: Optimisation Results

HDL Synthesis Report

Macro Statistics

# Adders/ Subtractors	: 1
32-bit adder	: 1
Counters	: 2
32-bit up counter	: 2
#Registers	: 82
13-bit register	: 40
2-bit register	: 1
32-bit register	: 1
40-bit register	: 22
8-bit register	: 10
9-bit register	: 8
#Comparators	: 6
32-bit comparator great equal	: 1
32-bit comparator less	: 1
8-bit comparator greater	: 4
# Multiplexers	: 2
40-bit 10-to-1 multiplexer	: 2
# Xors	: 83
1-bit xor2	: 73
1-bit xor3	: 10

The Xilinx Spartan 3A was adopted in our study as the features such as Suspend power-saving mode, high-speed I/O options, DDR2 SDRAM memory interface, commodity flash configuration support, and FPGA/IP protection using Device DNA Security. [16]. The Spartan-3A FPGA platform is a full feature platform of five devices with system gates ranging from 50K to 1.4M gates, and I/Os ranging from 108 to 502 I/Os, with density migration. The Spartan-3A FPGAs also support up to 576 Kbits of fast-block RAM with byte-write enable, and up to 176 Kbits of distributed RAM. Additionally, there are built-in multipliers for efficient DSP implementation and Digital Clock Managers (DCMs) for system level clock management function.

VI. Conclusion and Future Work

In this paper a We have studied the use of genetic algorithms in the optimization of Function F(x), the initial results are promising.

Other Genetic Algorithm operators could be implemented like, multi-point crossover, Partially Mapped crossover and different selection methods as well. The design can also be enhanced by incorporating a local search engine to create a hybrid memetic GA. The chromosome representation used in this project requires a relatively large amount of external memory to store the population and net list. Alternate chromosome representations can be explored in order to reduce the memory requirements. Furthermore, hardware/software co-design can be implemented and it can be compared with current implementation. Other real-time applications which require rapid and robust optimization can also be tackled with hardware based genetic algorithm.

VII. References

1) Chatchawit and Prabhas, —A Hardware Implementation of compace Genetic Algorithm", in Proceedings of the 2001 IEEE Congress on Evolutionary Computation, pp.624–629, Seoul, Korea, May 2001.
2) Douglas L Perry (2006), —VHDL Programming by Example", McGraw- Hill
3) G. Koonar S. Areibi M. Moussa, —Hardware implementation of Genetic Algorithms for VLSI CAD design", School of Engineering University of Guelph Guelph, Ontario,CANADA N1G 2W1.
4) Goldberg, D. E. (1989). —Genetic Algorithms in Search, Optimization & Machine Learning", Pearson Education, Inc
5) http://www.geatbx.com/docu/fcnindex.html #P74_1604.
6) http://www.obitko.com
7) Javad Frounchi Mohammad Hossein Zarifi Sanaz Asgari Far Hamed Taghipou —Design and Analysis of Random Number Generator for Implementation of Genetic Algorithms using FPGA"
8) . Microelectronic and Microsensor Research Lab, Faculty of Electrical and Computer Engineering, University of Tabriz,Iran
9) J. H. Holland, —Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence" MIT Pre Cambridge, MA, USA, 1992.

10) K. Skahill, VHDL for Programmable Logic, Addison Wesley, Reading, Massuchusetts, 1996.
11) Matti Tommiska and Jarkko Vuori,"Hardware Implementation of GA" Helsinki University of Technology, Otakaari Finland.
12) Mitchell Melanie," An Introduction to Genetic Algorithms", The MIT Press Cambridge.
13) Mustapha Abdulai, Inexpensive Parallel Random Number Generator for Configurable Hardware 2003
14) Nagham Azmi AL-Madi, Ahamad Tajudin Khader, De Jong's Sphere Model Test for A Social-Based Genetic", IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.3, March 2008
15) Russell Pinnington(1999) , —Implementation of Genetic Algorithm"MEng Cybernetics
16) T. Back. —Evolutionary Algorithms in theory and practice Evolution Strategies, Evolutionary Programming, Genetic Algorithms". Accessed 2008.
17) Xilinx Corporation, —ISE Manual", San Jose, CA.
18) [17] Zbigniew Michalewiez (1996)," Genetic algorithm + data structures = Evolution Programs",3rd Edition, Springer press.

HybridSGSA: SexualGA and Simulated Annealing based Hybrid Algorithm for Grid Scheduling

[1]Bhupinder Singh, [2]Seema Bawa

GJCST Computing Classification
F.2.2, H.3.4

***Abstract*-Scheduling jobs on computational grids is a compute intensive problem. Existing methods are unable to perform the required breakthrough in terms of time and cost. A Grid scheduler must use the available resources efficiently, while satisfying competing and mutually conflicting goals. The grid workload may consist of multiple jobs, with varying resource requirements and quality-of-service constraints. In this paper A hybrid algorithm based on SexualGA and simulated annealing is proposed, implemented and tested which tries to minimize makespan and cummulative delay in meeting user specified deadline time. Simulation results show that proposed algorithm performs better than other hybrid genetic simulated annealing algorithms proposed earlier.**

Keywords-Grid Scheduling, Computational Grids, Genetic Algorithms, Simulated Annealing, SexualGA.

I. INTRODUCTION

Computational grids are wide-area (Internet-scale) distributed environments that differ from conventional distributed computing by their focus on large-scale resource sharing, innovative applications, and highperformance orientation [1]. In grid architecture, four levels of management can be distinguished: fabric, connectivity, single resource, and collective multiple resources. The fabric layer typically constitutes computational resources, storage resources, network resources, and code repositories. The connectivity layer deals with easy and secure communication by providing single sign on, delegation, integration with local security solutions, and user-based trust relationships. The single resource layer is concerned with individual resources, and the two primary classes of resource layer protocols are information protocols and management protocols. The collective multiple resources layer provides directory services, co-allocation, scheduling, and brokering services, monitoring and diagnostics services, data replication services, grid-enabled programming systems, workload management systems and collaboration frameworks (problem solving environments) etc.

The responsibilities of a resource management system on the grid includes the ―discovery of available resources for an application, mapping the resources to the application subject to some performance goals and scheduling policies and loading the application to the resource in accordance with the best available schedule.‖ Scheduling, co-allocation and brokering services of resource management system are taken care by a special component of the grid resource management system known as scheduler / broker. A scheduler system provides the interface between a user and the grid resources. Scheduling of jobs on a grid or a cluster is the task of mapping jobs to the available compute-nodes which is NP-complete problem [2] and requires lot of optimization. A scheduler is designed to satisfy one or more of the following common objectives: (a) maximizing system throughput; (b) maximizing resource utilizations; (c) maximizing economic gains; and (d) minimizing the turnaround time for an application [3].

Simulated Annealing (SA) and Genetic Algorithms (GA) had been used in solving many NP-complete problems [4]-[7]. Since grid scheduling is also a compute intensive problem, hence it is a suitable candidate for applying evolutionary techniques. This paper presents a hybrid algorithm based on SexualGA[8] and SA to optimally schedule the jobs on computational grids. The Proposed algorithm tries to satisfy the constraints imposed by resource users and resource providers. Grid scheduler is the important part of Grid Resource Management System (GRMS), which gathers the information about the resources and chooses the best resource as per the job requirements. This is followed by the actual execution of the jobs. The problem of finding the best ―job-resource‖ pair is a compute-intensive problem and need to be formulated mathematically to find the optimal solution.

To formulate the problem, we consider mapping a set of independent user jobs $\{J_1, J_2, J_3, \ldots, J_N\}$ to a set of heterogeneous processors / resources $\{P_1,P_2,P_3,\ldots,P_M\}$ (Though we can have any type of resource that can be shared on a computational grids, but for our simulations we have considered processors as a resource.) This mapping is done with an objective of minimizing the completion time by utilizing the resources effectively, and also minimizing the delay in meeting user specified deadlines. The speed of each resource is expressed in number of cycles per unit time, and the length of each job in number of cycles. Each job *J* has processing requirement C_j cycles and processor P has speed of V_i cycles/second. Any job *J* has to be processed byP, until completion.

This paper is organized as follows. Section 2 provides the details of different components of the proposed HybridSGSA algorithm. In section 3, experimental test bed and setup details are discussed. Section 4 provides the experimental results and finally section 5 provides the conclusion describing the usefulness of the proposed algorithm in scheduling resources on computational grids.

[1]*Bhupinder Singh, Module Lead, Design Tools Department, Ciena India Pvt. Ltd. Gurgaon (Haryana), India. (bsaini@ciena.com)*
[2]*Seema Bawa, Professor, Computer Science and Engineering Department, Thapar University Patiala. (Punjab), India .(seema@thapar.edu)*

II. PROPOSED HYBRIDSGSA ALGORITHM

A. Objective / Fitness Function

The Objective / fitness function is used to differentiate between high and low quality solutions. It is defined in terms of objectives of the problem in hand. For grid scheduling problem, the users have goal of satisfying the deadlines of jobs submitted by them whereas the resource providers like to minimize the makespan. The schedule S is evaluated on the basis of fitness function, which is defined as.

$$\text{Fitness, } F(S) = 1/(\omega * MS + \theta * T_{delay}) \quad (1)$$

Where MS denotes the makespan of the schedule and T_{delay} denotes the cumulative delay in meeting deadlines. ω and θ are the weights to prioritize the components of the fitness function as per the needs of the resource provider and job users. The components MS and T_{delay} of the fitness function are defined as follows:

Let $t_{end}(i)$ and $t_{dline}(i)$ be tunaround time and deadline time for i_{th} job,

then $t_{delay}(i) = t_{end}(i) - t_{dline}(i)$ (2)

{t_{delay} is delay in meeting deadline for i_{th} job, and $t_{delay} = 0$ if $t_{end} < t_{dline}$ }

and $T_{delay} = \Sigma\, t_{delay}(i)$ for $1<i<N$ (3)

Makespan of schedule S is defined as

$MS = Max(t_{processing}(k))$ for $1<k<M$ (4)

{where $t_{processing}(k)$ is the time in which processor P_k will complete processing jobs assigned to it.}

B. Initial Temperature and Annealing Schedule

Initial temperature, T, is set to a value such that the acceptance probability for the chromosomes of the population is very high i.e. close to 1.0. The annealing schedule is generated by multiplying the current temperature by a constant less than 0, which is known as the cooling rate,ϒ. The probability of accepting new solution in the reproduction (crossover and mutation) is implemented as A function of temperature, T. So initially the new off-springs are accepted with a high probability, where as, in the later stage their probability of being accepted is lowered as the temperature decreases. Temperature, T, is decreased at cooling rate,ϒ, in each iteration and is given by Equation 5.

$T_i = \Upsilon * T_{i-1}$ {where i is the iteration number} (5)

C. SexualGA

SexualGA [8] is an enhanced selection scheme that tries to mimic the sexual selection in human beings. In that context sexual selection is described as the concept of male vigor and female choice, meaning that male individuals try to spread their gene material as wide as possible and female individuals are more selective by choosing rather above average fit males to guarantee a high survival probability of their off-springs. So it seems to be preferable not to use identical selection mechanisms for male and female population members also in GAs. Inspired by this view of sexual selection, a new selection mechanism for Gas (SexualGA) is presented by introducing two different selection operators, one for the selection of male and one for the selection of female individuals.

D. Chromosome Representation

Schedule, S, of independent jobs is represented by A chromosome. Chromosome is basically a $N\cdot\_\ \_1$ vector, where the position i $(0 < i < N)$ represents the job / task and the entry at position i $(s[i])$ is the processor / resource to which the job has been assigned. For example, in Fig.1, $S[1] = P_2$, means job J_1 is assigned to Processor P_2.

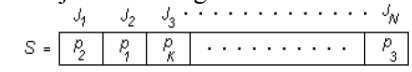


Fig.1. Chromosome representation

E. Crossover

Crossover operation is performed to generate new offsprings for the next generation. Two parents for the crossover are selected using the SexualGA based selection scheme and two-point crossover is performed. Using the vector representation for the chromosomes as shown in Fig.1, crossover operation is similar to exchanging the contents of two vectors for all positions before first crossover point and after the second crossover point. After the generation of new off-springs, change in the energy, δE, _and their probability of acceptance, P, is computed using the Equation 6 and Equation 7 respectively.

$$\delta E = f_{offspring} - f_{Parent}. \quad (6)$$

Where $f_{parent} = Max\ (f_{Parent1}, f_{Parent2})$

$f_{offspring}$ = Fitness of the new offspring

$$P = exp\ (-\delta E / T) \quad (7)$$

and $P = 1$, *if* $\delta E > 0$

After the finding the probability of acceptance, a uniform random number, ⌈, is generated between 0 and 1. New offspring is accepted if, $P >=$ ⌈, otherwise the new offspring is rejected.

F. Mutation

Mutation in the vector representation is considered as moving a job from one resource to another or swapping of two Jobs. For moving a job from one resource to another, two random numbers l and k are generated such that $S[l] = k$, where $0<l<N$ and $0<k<M$. After applying the mutation, the change in energy, δE, is computed using Equation 6 and probability of acceptance, P, is computed using Equation 7. If, $P >$ ⌈, (a uniform random number between 0 and 1), the new offspring is added to the next population, otherwise it is rejected.

G. Exit criteria

An exit criterion specifies when to stop further exploration,

and accept the solution. Proposed algorithm uses two exit criteria.

i. Standard deviation of the fitness value of the individuals in the population and
ii. No change in the elite chromosomes after 50 iterations.

After the fitness function for all the chromosomes in the population is calculated, standard deviation is computed. If the standard deviation reaches a sufficiently low threshold value or there is no change in the elite chromosomes after 50 iterations, solution is said to converge.

H. Building New Population

New population is build from previous population through the crossover, mutation and elitism. This process is executed as follows. X_c% of the chromosomes, for the new population, is constructed using the crossover process as described earlier. X_m% of the chromosomes is constructed by applying the mutation operator. After the application of crossover and mutation, the rest of the chromosomes are taken out of the existing population by making them identical to existing population's best chromosomes, so as to make the size of new population equal to the existing one.

III. Experimental Test Bed and Setup Details

Experimental test bed has been implemented to perform the experiments and to test and evaluate the proposed algorithm. This test bed uses the simulation model as used by Braun et. al. [9] in their study, so that results can be compared.

A. Grid Test Bed

Grid Test bed provides the actual grid like environment with a simple web based user interface for submission of jobs, registering and de-registering of resources and administrative tasks. Fig.2. shows the high level diagram of Grid Test Bed environment

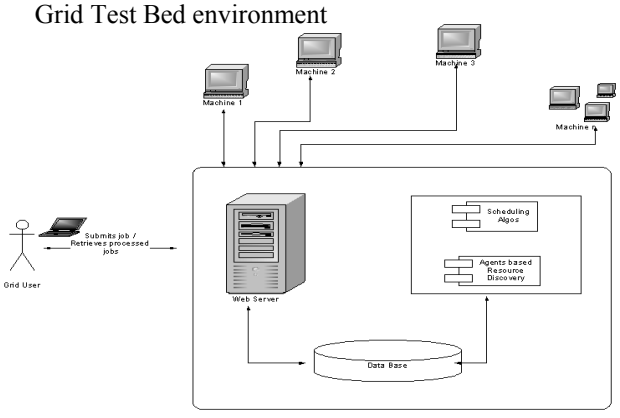


Fig.2. High level view of Grid Test Bed used for experiments

A web server listens to the user requests provided to the system by web interface. The information provided by the user is stored in the database. Users for the system can be grid users or resource owners. Grid users send the job requests for the jobs to be executed. Job characteristics are stored in the database. Grid users also specify the deadline time for the jobs. This is the time before which the job must be executed. Similarly resource owners use web based interface to register and de-register the resources. Whole information about the resources is also maintained in the database. Scheduling algorithm component provides the implementation of proposed HybridSGSA algorithm. Resource discovery is done using the ACO based resource discovery algorithm as discussed in [10]. Proposed algorithm work on the information present in the database about various jobs and resources and update the information as soon as the job's processing is finished.

B. Simulation Model

Grid test bed uses the simulation model as used in [9]. Since the results for the most of existing scheduling heuristics are already available, hence it is possible to compare the performance of proposed algorithm with the existing scheduling heuristics. Details of the simulation model are discussed in [9].

A completion time (CT) matrix is computed based for the job completion times on different resources. The prediction system in the grid test bed uses this completion time matrix to predict the execution time of a job on a resource. CT is an $N \cdot M$ matrix, which contains an entry for every jobresource pair. This entry specifies the completion time for A job on that particular resource.

In the actual grid systems state estimation is generally done based on the task profiling and analytical benchmarking. In this experimental Grid Test Bed, a pre-computed CT matrix is used, instead of actual task profiling.

A row in the CT matrix contains the completion time for A given job on each resource, where as a column consists of the completion time of every job on a given resource. Hence, for a job J_i and a resource R_j, an entry of the CT matrix contains the completion time of job J_i on resource R_j. Completion time matrix can be consistent, partially consistent or inconsistent that corresponds to the homogeneous, semi-heterogeneous, heterogeneous environments respectively. Since grid is highly dynamic and heterogeneous environment, hence only partially consistent and inconsistent matrices are considered for the experiments. Furthermore the heterogeneity can be considered as machine or task heterogeneity. CT matrix can be varied based on low machine or task heterogeneity and high machine or task heterogeneity. Taking into consideration that grid is highly heterogeneous environment; so low and high machine heterogeneity and high task heterogeneity CT matrix are considered in experiments.

IV. Experimental Results

This section discusses the experimental details and results of proposed HybridSGSA algorithm and its comparison with existing algorithms. The objective function of proposed algorithm mainly focuses on minimizing two Quality of Service (QoS) parameters: 1. Makespan 2. Cummulative

time delay in meeting user specified deadlines. Since existing results were available for minimizing the makespan only, so the results are compared with SA and GSA based algorithms from [9], by taking $\theta = 0$ in the fitness function $F(S)$. To support the algorithm proposed in this paper, we have performed several experiments taking different scenarios into consideration. Value of the different parameters used in experiments is given in Table 1.

Parameter Name	Value	Description
T	1	Initial Temperature
γ	0.9	Cooling Rate
X_c	0.75	Cross Over Probability
X_m	0.01	Mutation Probability
ThresholdSD	1.0E-06	Threshold Standard Deviation

Table 1: Parameters and their values used in experiments

As per the simulation model, completion time (CT) matrix can be of three types: consistent, partially- consistent and inconsistent matrices. These can further be varied based on low and high task and machine heterogeneity. As grid is highly dynamic environment where any machines many leave or join at any time, so we have only considered partially consistent and inconsistent CT matrices with low and high machine heterogeneity and high task heterogeneity. Also we have considered $512 \cdot 16$ matrices as given in [9], i.e. $N = 512$ (number of jobs) and $M = 16$ (number of resources). Experiments have been performed for following type of CT matrices: 1) inconsistent matrix with high machine heterogeneity (IC_HM), 2) inconsistent matrix with low machine heterogeneity (IC_LM), 3) partially consistent matrix with high machine heterogeneity (PC_HM) and 4) partially consistent matrix with low machine heterogeneity (PC_LM). Comparison results are presented in graph as shown in Fig. 3. It is established from the experimental results that proposed HybridSGSA algorithm performs better as compared to existing SA and GSA based algorithms.

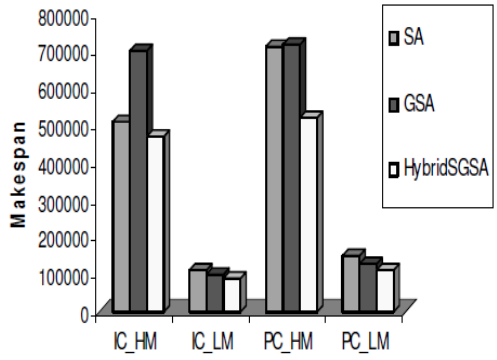


Fig 3: Comparison of SA, GSA and Proposed HybridSGSA

V. Conclusion

In this paper, we have looked at the problem of Grid scheduling and proposed, implemented and tested a new hybrid algorithm (HybridSGSA) based on SexualGA and SA. GA and SA both independently are valid approaches toward problem solving. GA can begin with a population of solutions in parallel, it has poor convergence properties. SA, on the other hand, has better convergence properties but cannot exploit parallelism. Furthermore SexualGA based selection scheme can better represent the concept of male vigor and female choice to more closely mimic the sexual selection in the human beings. Hence, it provides the advantage of not loosing the genetic diversity that result in better control over the selection pressure. The combined effect of all these factors makes proposed algorithm very effective for finding an optimal or near optimal solution quickly when search space is very large and dynamic as in case of computational grids.

VI. References

1) Foster, I. and Kesselman, C. (1998) Computational Grids, The Grid: Blueprint for a New Computing Infrastructure, Morgan- Kaufman, San Fransisco.
2) Abraham, A., Buyya R. and Nath, B.(2000) Nature's Heuristics for Scheduling Jobs on Computational Grids, The 8th IEEE International Conference on Advanced Computing and Communications (ADCOM 2000), India.
3) Wright, D. (2001) Cheap Cycles from the Desktop to the Dedicated Cluster: Combining Opportunistic and Dedicated Scheduling with Condor, Proceeding of HPC Revolu-tion 01, Illinois.
4) Yao, X. (1995) A New Simulated Annealing Algorithm, International Journal of Computer Mathematics, Vol 56, Pp 161-168.
5) Goldberg, D. (1989) Genetic Algorithms in Search, Optimization and machine learning, Addison-Wesley Publishing Company, Inc.
6) Bawa, S. and Sharma, G.K. (2005) Search space pruning for faster test generation based on parallel and adaptive GA. Progress in VLSI Design & Test. Elite publishing, Pp 444-448.
7) Bawa, S. (2004) Parallel and Distributed Algorithms for VLSI Test Generation" Ph.D Thesis. Thapar University, Patiala , India.
8) Affenzeller. M. (2005) SexualGA: Gender-Specific Selection for Genetic Algorithms, Proceedings of the 9th World Multi-Conference on Systemics, Cybernetics and Informatics. In: International Institute of Informatics and Systemics, 4, Pp 76-81.
9) Braun, T. D., Siegel, H. J. and Beck, N. (2001) A Comparison of Eleven Static Heuristics for Mapping A Class of Independent Tasks onto Heterogeneous Distributed Computing Systems, Journal of Parallel and Distributed Computing.
10) Singh, B. and Bawa, S. Grid Resource Discovery using Natural Mobile Agents, International Journal of Computer Science and Network Security. Vol 8. No. 11, Pp 411-415

A Reliable And Energy Efficient Transport Protocol for Wireless Sensor Networks

[1]Kamal Kumar Sharma, [2]Harbhajan Singh and [3]R.B Patel

GJCST Computing Classification
C.2.2 , H.3.4, C.2.1

***Abstract*-In wireless sensor networks (WSN), an ideal transport layer needs to support reliable message delivery and provide congestion control in an efficient manner in order to extend the lifetime of a WSN. The main use of transport protocol in WSN is to overcome the congestion and the reliability with energy efficiency. In this paper, we develop a reliable and energy efficient transport protocol (REETP), which mainly focuses on the reliability and energy efficiency. Our proposed protocol consist of an Efficient Node Selection Algorithm to determine A set of efficient nodes called E-Nodes which form a near optimal coverage set with largest area and highest residual energy level. The key idea of REETP is to transfer encoded packets using LT codes from the source to the sink block by block and each block is forwarded to an E-node. After receiving encoded packets, the E-node tries to reconstruct the original data packets and it encodes the original data packets again and relays them to the next E-node until it reaches the sink. By simulation results, we show that our proposed protocol has more packet delivery ratio with reduced packet loss and energy consumption.**

Keywords-Wireless Sensor Networks, Energy EfficientTransport Protocol, Efficient Node Selection Algorithm,Congestion control, MAC layer.

I. INTRODUCTION

A. *Wireless Sensor Networks*

Wireless Sensor Networks (WSNs) are highly distributed self-organized systems and depends upon a particular number of scattered low cost small devices. These devices include some strong demerits in terms of processing, memory, communications and energy capabilities. Sensor nodes collect measurements of interest over a given space and make them available to external systems and networks at sink nodes. The power saving techniques is commonly implemented to increase the independence of the individual nodes and this technique makes the nodes to sleep most of the time. This can be balanced with low power communications which usually lead to multi hop data transmission from sensor nodes to sink nodes and vice versa [1]. In order to collect the data, WSN uses an event-driven model and depends upon the collective effort of the sensor nodes in the network. Greater accuracy, larger coverage area and extraction of localized features are some of the advantages of the event-driven model over the traditional sensing. It is important that the preferred events are reliably transported to the sink for realizing these potential gains [2]. Habitat monitoring, in-door monitoring, target tracking and security surveillance are some of the applications where WSNs can be used. WSNs have some problems to be overcome such as energy conservation, congestion control, reliability data dissemination, security and management of A WSN itself. These problems often take part in one or more layers from application layer to physical layer and it can be studied separately in each corresponding layer or collaboratively cross each layer. For example, congestion control may involve only in transport layer but the energy conservation may be related to physical layer, data link layer, network layer and higher layers [3].

B. *Transport Protocols In WSN*

The transport protocols in WSN should support

i. Reliable message delivery,
ii. Congestion control, and
iii. Energy efficiency.

The requirement for transport layer protocol in WSN has been discussed. The following are the suggestions given by the researchers [4]:

i. Loss detection and recovery can be handled below the transport layer and mitigated using data aggregation
ii. Congestion is not an issue because sensor nodes spend most of the time sleeping resulting in sparse traffic in the network

Generally the deployment of sensor nodes produces congestion in WSN in the contradiction to the above arguments against the need for a transport layer protocol. In the absence of congestion control, data from sensor nodes to sink may suffer from channel contention which in turn decreases the ability of the sensor nodes to deliver data to the sink. Since the layers under the transport layer do not provide guaranteed end-to-end reliability, it is inadequate to depend upon the loss detection and reliability techniques, in the situation where data‘s are delivered reliably in WSNs [4].

Like other networks, WSNs should have a transport layer in order to posses reliable message delivery and congestion control. An ideal transport layer needs to support reliable message delivery and provide congestion control in an efficient manner in order to extend the lifetime of a WSN [4].

The following are the some of the transport protocols developed in the wireless sensor networks [5]:

i. TCP/IP – Transmission Control Protocol
ii. PCCP - Priority-based Congestion Control Protocol

[1]Kamal Kumar Sharma, [2]Harbhajan Singh and [3]R.B Patel

iii. STCP - Sensor Transmission Control Protocol [6]
iv. MQQT – Message Queuing Telemetry Transport [7]
v. PORT - Price-Oriented Reliable Transport Protocol [8]
vi. PSFQ - Pump Slowly, Fetch Quickly [9]
vii. RMST - Reliable Multi-Segment Transport [10]
viii. ESRT - Event to Sink Reliable Transport [11]

Except STCP, the above mentioned protocols consider either congestion control or reliability guarantees. Some protocols use end-to-end and others hop-by-hop controls and also some guarantees event reliability and others provide packet reliability. The following are the two fundamental demerits of the existing protocols for WSNs [5]:

i. Since sensor nodes in WSNs can be installed with different kinds of sensors and used in different geographical locations, it may have different priorities.
ii. The existing transport protocols for WSNs assume that single path routing us used in the network layer without considering the multipath routing.

We summarize the requirements of a transport layer protocol for sensor networks as follows [6]:

Generic-The transport layer protocol should be independent of the application, Network and MAC layer protocols to be applicable for several deployment scenarios.

Heterogeneous data flow support-Continuous and event-driven flows should be supported in the same network.

Controlled variable reliability-Some applications require complete reliability while others might tolerate the loss of a few packets. The transport layer protocol should leverage this fact and conserve energy at the nodes.

Congestion detection and avoidance-The congestion detection and avoidance mechanism helps in reducing packet retransmissions, thereby conserving energy.

Base station controlled network-Since sensor nodes are energy constrained and limited in computational capabilities, majority of the functionalities and computation intensive tasks should be performed by the base station.

Scalability-Sensor networks may comprise of large number of nodes, hence the protocol should be scalable.

Future enhancements and optimizations-The protocol should be adaptable for future optimizations to improve network performance and support new applications. The main use of transport protocol in the wireless sensor networks is to overcome the congestion and the reliability with energy efficiency. In this paper we develop a reliable and energy efficient transport protocol, which mainly focus on the reliability and energy efficiency. The paper is organized as follows. Section 2 presents the related work done and section 3 presents the Efficient Node Selection Algorithm. The LT coding technique is described in section 4 and the proposed reliable and energy efficient transport protocol is presented in section 5. Section 6 presents the simulation results and the paper is concluded in section 7.

II. Related Work

Sandip Dalvi et al [2] have proposed a transport protocol which provides the desired event reliability to the application, by distributing the load at a sensor among its children based on their residual energies and average MAC layer data rate. The event rate distribution happens in such A way that the application at the sink gets its required event rate and the overall energy consumption of nodes is minimized. They have derived a method for computing average MAC data rate for these two protocols and using simulations they have shown that our transport protocol performs close to optimal.

Nurcan Tezcan et al [12] have addressed the problem of reliable data transferring by first defining event reliability and query reliability to match the unique characteristics of WSNs. They have considered event delivery in conjunction with query delivery. They have proposed an energy-aware sensor classification algorithm to construct a network topology that is composed of sensors in providing desired level of event and query reliability. They have analyzed their approach by taking asymmetric traffic characteristics into account and incorporating a distributed congestion control mechanism. They have evaluated the performance of their proposed approach through an ns-2 based simulation and show that significant savings on communication costs are attainable while achieving event and query reliability.

Yao-Nan Lien et al [13] has proposed the Hop-by-Hop TCP protocol for sensor networks aiming to accelerate reliable packet delivery. Hop-by-Hop TCP makes every intermediate node in the transmission path execute a light-weight local TCP to guarantee the transmission of each packet on each link. It takes less time in average to deliver a packet in an error-prone environment.

Sunil Kumar et al [14] have studied the performance of ESRT in the presence of over-demanding event reliability, using both the analytical and simulation approaches. They have shown that the ESRT protocol does not achieve optimum reliability and begins to fluctuate between two inefficient network states. With insights from update mechanism in ESRT, they have proposed a new algorithm, called enhanced ESRT (E2SRT), to solve the overdemanding event reliability problem and to stabilize the network. Their simulation results show that their E2SRT outperforms ESRT in terms of both reliability and energy consumption in the presence of over-demanding event reliability. It also ensures robust convergence in the presence of dynamic network environments.

Damayanti Datta et al [15] have proposed a new protocol for reliable data transfer in time-critical applications with zero tolerance for data loss in wireless sensor networks which uses less time and fewer messages in comparison to an established protocol PSFQ. The two key features of their proposed protocol are out-of-sequence forwarding of packets with a priority order for sending different types of messages at nodes and delaying the requests for missing packets. They have also presented two methods for computation of the delay in requesting missing packets.

III. Efficient Node Selection Algorithm

Before discussing our proposed reliable and energy efficient

transport protocol (REETP) in detail, we present an Efficient Node Selection Algorithm (Algorithm 1). In the efficient node selection algorithm, we determine a set of efficient nodes called E-Nodes which form a near optimal coverage with set with largest area and highest residual energy level.
Also we assume that sensors are able to monitor their residual energy because many electronic devices are equipped with energy monitoring functions. The energy level (EL) of sensors s_i at the beginning of update interval (UI), denoted by EL is calculated as:

$$EL = \frac{RE(UI)}{IE} \quad (1)$$

where IE is the initial energy corresponding to fully charged battery and RE (UI) is the residual energy of sensors s_i at the beginning of the update interval. In each iteration, Algorithm 1 selects one node from the unselected sensors which covers the largest area with highest residual energy level. For this purpose, a weight value is defined to represent the weight of a sensing region of a sensor based on its residual energy. For a given region, the weight value based on the residual energy level of A sensor is:

$$W(R_i) = EL \times A(R_i) \quad (2)$$

where EL is the energy level given in (1) and $A(R_i)$ is the area of sensing region R_i. Then, we calculate the gain of selecting each sensor using the weight value. To do this, we first find the size of the area that can be covered by sensor s_i and has not been covered yet. Consider the sensor s_i with sensing region Ri. Let R_{CS} be the area that sensors of C covered so far, i.e., $R_{CS} = \bigcup_{s_j \in C} R_j$. Beneficial area of s_i is defined to be the region inside the sensing field which has not been covered, i.e., RB = (R_i_A)/R_{CS}. Hence, gain function for sensor s_i is the total weight of its beneficial area, which is given as:

$$G(Si) = W(RB), s_j \in C \quad (3)$$

Where G(Si) is the gain function and RB is the beneficial area. Algorithm 1 is to find a near-optimal coverage set C. Then each member of the set C is known as an E-node.

Algorithm 1

i. Let C = Φ
ii. Let R_{CS} be the total sensing region of C
iii. Let S-C = {s_1, s_2…..s_n}
iv. G_{max} =0
v. For each $s_i \in$ S - C
vi. Calculate the energy gain of s_i

$$G(s_i) := \sum_{aj \in (R \cap A)/RCS} W_i(aj)$$

vii. If (EL_B _ G_{max})
G_{max} = G
temp: = s_i
End if
viii. End for
ix. C = C∪temp
x. If A⊆R_{CS}, then
Return C
Else Repeat from 3 End if

IV. FEC USING LT CODES

LT codes are rateless because the number of encoding symbols which are generated from the data is unlimited. The required encoding symbols can be generated immediately. From any set of the generated encoding symbols, an exact copy of the data can be recovered by the decoder. Thus the required encoding symbols can be generated without depending on the loss model on the erasure channel. In order to recover the data, the generated symbols are sent over the erasure channel until the adequate number has been arrived at the decoder. The LT codes are near optimal with respect to any erasure channel because the decoder can recover the data from the near optimal number of possible encoding symbols. Moreover, as a function of the data length, the encoding and decoding times are very efficient. When compared with the previous erasure codes, LT codes provide various advantages for different types of data delivery applications. Using LT codes, the minimal number of encoding symbols can be generated and send the packets to the receivers. The minimal number of encoded symbols is required to recover the original data from each receiver [16]. Robust distributed storage, delivery of streaming content, delivery of content to mobile clients in wireless networks, peer-to-peer applications and delivery of content along multiple paths in order to ensure resiliency to network disruptions, are some of the other applications of the LT codes [16].

A. Lt Process

The preferable length L of the encoding symbols can be selected. Due to the overheads with the accounting operations, the overall encoding and decoding is more efficient for larger values of L and this value does not have any influence on the history. Sometimes the length L is selected to be closer to the length of the packet payload in case of transport applications [16].
Encoding-The data of length N is partitioned into K=N / L input symbols such that each transport symbol is of length of L. Each encoding symbols are connected with a key. In order to produce the degree and set of neighbors of the encoding symbol, both the encoder and decoder applies the same function to the key. In order to generate an encoding symbol, the encoding symbol may choose each key randomly and this key is passed to the decoder along with the encoding symbols. Alternatively, each key may be produced by a deterministic process, e.g., each key may be larger than the previous key. The encoder and decoder have the access to the same set of random bits. In order to produce the degree and the neighbors of the encoding symbol, each key is used as the seed to a pseudo-random generator which uses these random bits [16].
Decoding-For a given group of encoding symbols and some illustrations of their associated degrees and sets of neighbors, the decoder recovers the input symbols repeatedly using the following rule as long as it applies [16].
Since the neighbor is a copy of the encoding symbol, it can be recovered immediately if there is at least one encoding

symbol which has exactly one neighbor. The value of the recovered input symbol is XORed into any remaining encoding symbols that also have that input symbol as a neighbor. The recovered input symbol is removed as a neighbor from each of these encoding symbols and the degree of each such encoding symbol is decreased by one to reflect this removal [16].

V. PROPOSED RELIABLE AND ENERGY EFFICIENT TRANSPORT PROTOCOL (REETP)

The key idea of the Reliable and Energy Efficient Transport Protocol (REETP) is to transfer encoded packets using LT codes, block by block. In order to reconstruct the original data packets, the receiver has to receive sufficient encoded packets. The REETP has to guarantee that the receiver can receive enough encoded packets in such a limited time interval. By setting the block size n (i.e., the number of original data packets in each block) appropriately, REETP can control the transmission time and allow the receiver to be able to receive enough packets in order to reconstruct original block even in node motion.

In REETP, a data source first groups data packets into blocks of size n. Then the source encodes these blocks of packets, and sends the encoded blocks into the network. The data packets are forwarded from the source to the sink block by block, and each block is forwarded to an E-node. In each E-node relay, the sender first estimates the number of packets needed to send for the E-node to reconstruct the original packets. We call this number as ―MaxPacket". Within the MaxPacket, the sender pushes the encoded packets to the network fast. When the packet is reached, the sender slows down pack transmission, waiting for a positive feedback from the E-node. After receiving encoded packets, the receiver tries to reconstruct the original data packets. If the reconstruction is successful, it sends back a positive feedback. Upon the reception of a feedback, the sender stops sending packets, while the E-node encodes the original data packets again and relays them to the next E-node until the sink is reached. The operations performed on the sender and the receiver (E-node) is described in the following.

REETP Sender-Estimates the MaxPacket.

i. Encodes a block using LT codes.
ii. Pumps encoded packets fast in a random order within the
iii. MaxPacket.
iv. Sends encoded packets slowly outside the MaxPacket. until receiving a positive feedback from the E-node.

REETP Receiver-

i. Keeps receiving packets until it can reconstruct the original data packets, and sends a positive feedback to the Sender.
ii. Encodes the reconstructed packets again and relay them to the next E-node.

From the above description, we can see that REETP reduces the burden of the sender and the receiver by requiring only one feedback per block. The sender has no additional responsibilities except encoding and injecting packets, and the receiver only needs to send one feedback after reconstructing the original packets.

VI. EXPERIMENTAL RESULTS

A. Simulation Model And Parameters

We use NS2 [17] to simulate our proposed protocol. In our simulation, the channel capacity of mobile hosts is set to the same value: 2 Mbps. We use the distributed coordination function (DCF) of IEEE 802.11 as the MAC layer protocol. In our simulation, 100 sensor nodes are deployed in a 1000 mr x 1000 m region for 50 seconds simulation time. All nodes have the same transmission range of 250 meters. The simulated traffic is Constant Bit Rate (CBR). The simulation settings are summarized in the following table.(Table 1).

No. of Nodes	100
Area Size	1000 X 1000
Mac	802.11
Simulation Time	50 sec
Traffic Source	CBR
Packet Size	512
Transmit Power	0.360 w
Receiving Power	0.395 w
Idle Power	0.335 w
Initial Energy	3.1 J
No. of sources	2,4,6,8
Transmission Rate	250,500,750 and 1000 kb.

B. Performance Metrics

We compare the performance of our proposed REETP protocol with A MAC-aware Energy Efficient Reliable Transport Protocol (MAEERTP) [2] for WSN. We evaluate mainly the performance according to the following metrics:

Average Energy Consumption-The average energy consumed by the nodes in receiving and sending the packets are measured.

Packet Delivery Ratio-It is the ratio of the fraction of packets received successfully and the total no. of packets sent.

Average Packet Loss-It is average number of packets lost at each receiver and the sink. The performance results are presented graphically in the next section.

C. Results

A. Varying No. Of Sources

In the first experiment, in order to study the impact of increased the number of sources, we vary the no. of sources as 2,4,6 and 8 and measure the performance of the protocols.

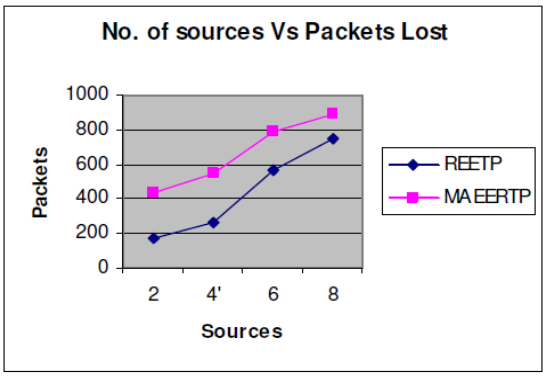


Fig: 1 No. of Sources Vs Packets Lost

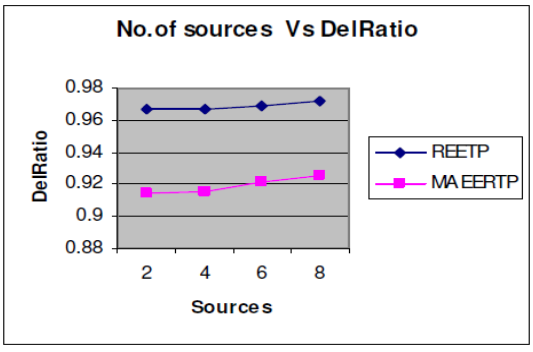


Fig: 2 No. of Sources Vs DelRatio

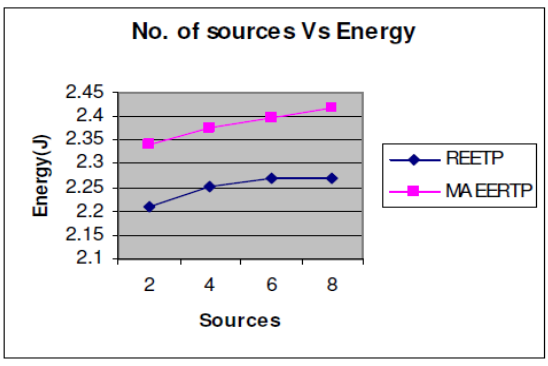


Fig: 3 No. of Sources Vs Energy

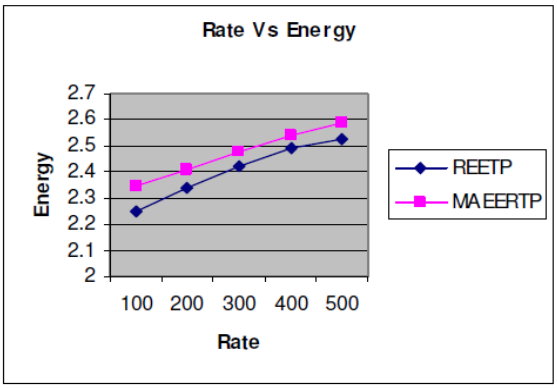


Fig: 4 Rate Vs Energy

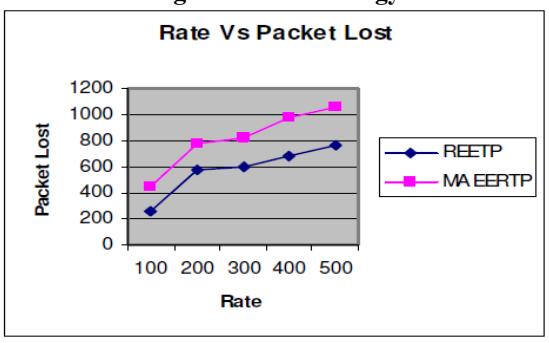


Fig: 5 Rate Vs Packet Lost

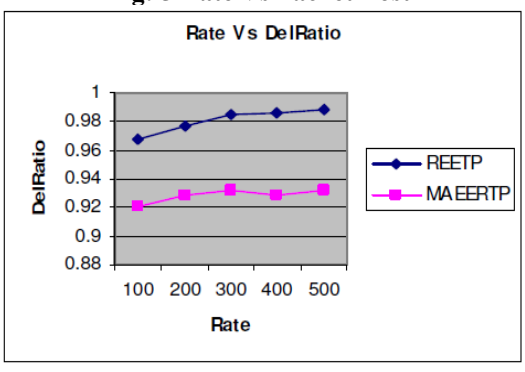


Fig: 6 Rate Vs DelRatio

Fig.1-shows the packet lost obtained with our REETP protocol compared with MAEERTP protocol. It shows that the packet lost is significantly less than the MAEERTP, as sources increases. From

Fig 2- we can see that the packet delivery Ratio (PDR) for REETP increases, when compared to MAEERTP protocol.

Fig.3- shows that the average energy consumed by the nodes in receiving and sending the data. Since REETP make use of energy efficient scheduling, the values are considerably less in REETP when compared with MAEERTP protocol.

B. Varying The Transmission Rate

In the second experiment, in order to study the performance of increased traffic sending rate, we vary the transmission rate as 100,200,300,400 and 500Kb to measure the performance of the protocols.

Fig. 4-shows that the average energy consumed by the nodes in receiving and sending the data. Since REETP make use of energy efficient scheduling, the values are considerably less in REETP when compared with MAEERTP protocol.

Fig.5-shows the packet lost obtained with our REETP protocol compared with MAEERTP protocol. It shows that the packet lost is significantly less than the MAEERTP, as rate increases.From

Fig. 6- we can see that the packet delivery Ratio (PDR) for REETP increases, when compared to MAEERTP protocol.

VII. CONCLUSION

In this paper, we have developed a reliable and energy efficient transport protocol (REETP), which mainly focuses on the reliability and energy efficiency. Our proposed protocol consist of an Efficient Node Selection Algorithm to

determine a set of efficient nodes called E-Nodes which form a near optimal coverage set with largest area and highest residual energy level. The objective of REETP is to transfer encoded packets using LT erasure codes from the source to the sink block by block and each block is forwarded to an E-node. The sender first estimates MaxPacket which is the number of packets needed to send for the E-node to reconstruct the original packets. When the packet is reached, the sender slows down pack transmission, waiting for a positive feedback from the E-node. After receiving encoded packets, the receiver tries to reconstruct the original data packets. If the reconstruction is successful, it sends back a positive feedback. Upon the reception of a feedback, the sender stops sending packets, while the Enode encodes the original data packets again and relays them to the next E-node. By our simulation results, we have shown that our proposed protocol has more packet delivery ratio with reduced packet loss and energy consumption.

VIII. References

1) Paulo Rogerio Pereira, Antonio Grilo, Francisco Rocha, Mario Serafim Nunes, Augusto Casaca, Claude Chaudet, Peter Almström and Mikael Johansson, "End-To-End Reliability in Wireless Sensor Networks: Survey and Research Challenges", Euro FGI Workshop on IP QoS and Traffic Control, 2007.
2) Sandip Dalvi, Anirudha Sahoo and Ashutosh Deo, "A MAC-aware Energy Efficient Reliable Transport Protocol for Wireless Sensor Networks", In the Proceedings of the IEEE conference on Wireless Communications & Networking Conference, 2009.
3) Chonggang Wang, Kazem Sohraby, Bo Li and Weiwen Tang, "Issues of Transport Control Protocols for Wireless Sensor Networks", In the Proceedings of International Conference on Communications, Circuits and Systems (ICCCAS), 2005.
4) Justin Jones and Mohammed Atiquzzaman, "Transport Protocols for Wireless Sensor Networks: State-of-the-Art and Future Directions", International Journal of Distributed Sensor Networks, 2007.
5) Chonggang Wang, Kazem Sohraby, Bo Li, Mahmoud Daneshmand and Yueming Hu, "A Survey of Transport Protocols for Wireless Sensor Networks", IEEE Networks, 2006.
6) Yogesh G. Iyer, Shashidhar Gandham and S. Venkatesan, "STCP: A Generic Transport Layer Protocol for Wireless Sensor Networks", In the Proc. of IEEE Intl. Conf. on Computer Communications and Networks (ICCCN), 2005.
7) Urs Hunkeler, Hong Linh Truong and Andy Stanford- Clark, "MQTT-S – A Publish/Subscribe Protocol For Wireless Sensor Networks", In the Proc. of 2nd Workshop on Information Assurance for Middleware Communications, (IAMCOM), 2008.
8) Yangfan Zhou, Michael R. Lyu, Jiangchuan Liu and Hui Wang, "PORT: A Price-Oriented Reliable Transport Protocol for Wireless Sensor Networks", Proceedings of the 16th IEEE International Symposium on Software Reliability Engineering, 2005.
9) Chieh-Yih Wan, Andrew T. Campbell and Lakshman Krishnamurthy, "PSFQ: A Reliable Transport Protocol for Wireless Sensor Networks", In the Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications, 2002.
10) Fred Stann and John Heidemann, "RMST: Reliable Data Transport in Sensor Networks", In the 1st IEEE International Workshop on Sensor Net Protocols and Applications (SNPA), 2003.
11) Yogesh Sankarasubramaniam, Ozgur B. Akan and Ian F. Akyildiz, "ESRT: EventtoSink Reliable Transport in Wireless Sensor Networks", Proceedings of 4th ACM International symposium on Mobile ad hoc networking & computing, 2003.
12) Nurcan Tezcan and Wenye Wang, "ART: An Asymmetric and Reliable Transport Mechanism for Wireless Sensor Networks", International Journal of Sensor Networks, 2007.
13) Yao-Nan Lien, "Hop-by-Hop TCP for Sensor Networks", International Journal of Computer Networks & Communications (IJCNC), 2009
14) Sunil Kumar, Zhenhua Feng, Fei Hu and Yang Xiao, "ESRT: enhanced event-to-sink reliable transport for wireless sensor networks", Wireless Communications and Mobile Computing, 2008.
15) Damayanti Datta and Sukhamay Kundu, "An Application-Specific Reliable Data Transfer Protocol in Wireless Sensor Networks", Journal of Networks, 2008.
16) Michael Luby, "LTCodes", Proceedings of the 43rd Symposium on Foundations of Computer Science, 2002.
17) Network Simulator: www.isi.edu/nsnam/ns

A Proficient Content-Based Video Retrieval System Using Extensive Features And LSI

[1]Priya Rajendran and [2]T.N.Shanmugam

GJCST Computing Classification
H.5.1 , I.2.10

Abstract- **Rapid growth in multimedia technologies facilitates the acquisition and storage of videos in a cost effective manner; leads to the processing of ginormous videos. However, for effective processing, suitable search methodologies are essential pre-requisite in any video processing system. In this paper, we propose a proficient content-based video retrieval system with the aid of extensive features and LSI. The system is comprised of three stages, shot segmentation, feature extraction and retrieval of similar video clips. Initially, the system performs the task of shot segmentation and feature extraction for every database video. In the feature extraction, we extract the most dominating features that include motion object, color histogram, texture and contour features from every shots. All the extracted features that correspond to a video are stored in a feature database. In the retrieval of similar video clips, the system retrieves a required number of video clips that are similar to a given query clip. The retrieval can be accomplished by performing the shot segmentation and feature extraction for the query clip. Then, the LSI-based similarity is measured between the feature set of the query clips and the feature database. Hence, the database videos that are relevant to the query clip can be retrieved more proficiently using the proposed system. Eventually, the system is evaluated by determining the precision and recall for different query clips.**

Keywords- Content-based Video Retrieval (CBVR), Shot segmentation, Motion Estimation, Gaussian Mixture Model, Color Histogram, Query Clip, Latent Semantic Indexing (LSI)I.

I. INTRODUCTION

Nowadays, it is very easy to acquire, store, upload and deliver the video contents because of extreme variation in the development of digital devices, Internet infrastructures, and Web technologies. The search for video content over the Web is yet seems to be extremely difficult even after the success of the web search engines. Generally, a number of the web search engines index merely the Meta data of videos and search them by texts. In video retrieval, conventional search engines may be restricted, in case if they are devoid of the capability to comprehend media contents. By means of the usage of the rich media contents for the purpose of video retrieval there must be ample room in improving conventional search engines. In order to construct future video search engines [1] this has done content-based video retrieval (CBVR) a hopeful direction. The stability in the numerous video libraries that are becoming publicly available at present has resulted in the demand for methodologies which are efficient of manipulating the video data based on the content [3]. However, conventional database management systems which work on relational or object oriented data model yet do not offer sufficient facilities in order to maintain and retrieve video contents. As we have portrayed in [2], three chief reasons can be noted for this: (i) lack of amenities for the administration of spatiotemporal relations, (ii) lack of knowledge-based methods for deducing raw data into semantic contents, and (iii) lack of query representations.

Owing to the increasing abundance of digital video contents effective techniques for analysis, indexing, and retrieval of videos according to their contents have turn out to be ever more important. Video content is dealt at various levels in the existing works: raw data, low-level visual content and semantic content [4]. Raw video data consists of fundamental video units together with common video formats like color, shapes, textures and the like. On the contrary, semantic content includes high-level concepts such as objects and events. By means of a number of visual presentations it is probable to represent semantic content. For the extraction of these contents the major difference between the two kinds of content is the different necessities. The process of extracting the semantic content is comparatively difficult due to the fact that domain knowledge or user interaction is necessary for this. On the other hand the extraction of visual features is naturally domain independent [4].

Based on their visual content such as color distribution, texture and shape [5] a number of researchers have paid attention on the retrieval of video and image data and the approaches works on the basis of similarity measurements. VisualSEEk [6], Photobook [7], Blobworld [8], as well as Virage video engine [9], CueVideo [10] and VideoQ [11] are a few of the prominent examples for image and video retrieval systems. Based on the visual image contents the image retrieval systems make possible the users to devise queries– properties such as color percentages, color layout and textures present in the images generally with the assistance of instances of prior matches (query by example). Some of these systems make use of spatial information and allow the user to build queries either by drawing the layout of color regions, or by giving the URL of a seed image. In the former approaches of video retrieval, functionality for segmentation and key frame extraction were combined to the existing image retrieval systems. Subsequently after the extraction of the key frame similarity measurement based on the low level characteristics was applied. The above mentioned procedure is not appropriate due to the fact that video is temporal data and therefore sequencing of individual frames, which produced new semantics which may not exist in any of the individual shots, was performed.

In recent times, the requirement for intelligent processing as well as the analysis of multimedia information has been on a stable way up. There are a number of methodologies for the video management. For instance the shot transition detection, key frame extraction, video summarization and video retrieval and further more were developed by the researchers. Amongst various methodologies content based retrieval is well thought-out to be the most challenging and important issue of practical value. Based on the video contents via the user interactions [12] it is capable of assisting the users in the retrieval of the chosen video segments from a huge video database effectively.

Usually it is possible to segment the video retrieval system into two chief constituents. They are as follows: a module for the extraction of representative characteristics from video segments and one defining a fitting similarity model in order to locate identical video clips from video database. A number of approaches have made use different features in order to indicate a video sequence of which color histogram [13], shape information [14], motion activity [15], and text analysis [16] are some noteworthy features. In order to improve the retrieval performance [17] a few approaches concatenated the above mentioned features.

A novel and efficient CBVR system is offered in this paper. From a gathering of videos, alike video clips are retrieved for a query video clip is performed in the proposed system. The pre-annotation of video shots is not necessitated by the proposed system. Traditionally, in a number of content-based video analysis techniques available, the first step is to separate a video into elementary shots, each one of them constituting a sequence of consecutive frames recording a video event or scene continuous in time and space. The elementary shots in this are composed in order to form a video sequence during the video sorting or editing with either cut transitions or gradual transitions of visual effects for instance fades, dissolves, and wipes.

As a next step four different kinds of video features, comprising motion, texture, color and contour for every video shot is extracted in our system. In order to minimize the dimensionality of the data, the features to be extracted are preferably compact and discriminative. In order to symbolize the temporal information of videos, motion is considered to be the key feature. For the CBVR, efficient motion feature extraction is a significant step. A Gaussian mixture model of texture is stored as another feature. One of the most broadly used methods is color histogram because it is more robust to changes owing to scaling, orientation, perspective, and occlusion of images, which are recognized after segmentation process using anisotropic diffusion. As a final feature, contour of all the objects in the shots are extracted by using sparse field method. After the extraction of all the features they are stored in the feature library. The related videos are retrieved when a query clip is provided in the proposed system. For that reason, the abovementioned four diverse features are extracted for a query video clip and are evaluated against the features in the feature library. With the assistance of the similarity measure the comparison is performed between the query features and the features in the feature library. For the purpose of similarity measure calculation the proposed system utilizes the LSI method. On the basis of the calculated LSI related videos are retrieved from the collection of videos later.

The rest of the paper is organized as follows. A concise review of a few of the existing works in CBVR is presented in the Section 2. In the section 3 the proposed efficient CBVR system is detailed with necessary pictorial representation as well as mathematical formulations. The outcomes and discussions are described in the Section 4. In Section 5 the conclusions are summed up.

II. Related Works

Our work has been moved by a variety of former works available in the literature for CBVR. Few of them are described as follows

Arnab Ghoshal et al. [42] have taken into account the consequence of novel sources of video data on the performance of content-based retrieval systems. The typically utilized visual features contained one source of mismatch is demonstrated by them. It is required to carry out more work in order to recognize robust visual features because the color, edge or texture features are not anticipated to be robust to changes in photometric conditions. Yet by this moment one can take into account learning a set of transforms for matching features from one source to another by means of the non-robust features, which is a solution that can show the way to improvements over baseline conditions. Besides, it is illustrated by them that concepts whose location remains comparatively fixed are robustly identified in comparison to the concepts whose spatial location changes in novel sources.

A. Dyana et al. [43] have proposed a Gabor filter based representation of motion trajectory, for the motion-based video retrieval. They proposed a spectro-temporal representation of the trajectory, which concerned about the process of identifying a set of salient points from the peaks (locally) of the Gabor filter responses. The feature set (formed by the frequency, temporal location and turning direction at each salient point) provided a semantic representation of the trajectory. Their approach was a global trajectory representation where matching was carried out on the basis of the edit distance and was revealed to carry out well even for partial trajectory matching. They experimented on real world videos. Experimental results illustrated that better performance than existing systems on the basis of Fourier descriptors, polynomial representation and two state-of-the-art methods of symbolic representations based on PCA and characteristics of movement.

Liang-Hua Chen et al. [44] have proposed a video retrieval algorithm on the basis of the integration of a number of visual cues. On the contrary to key-frame based representation of shot, their approach examined all frames within a shot in order to build a compact representation of video shot. In the video matching step, by combining the color and motion features, a similarity measure was defined to spot out the occurrence of related video clips in the database. For that reason, their approach was able to fully make use of the spatio-temporal information contained in

video. Experimental results specified that their approach was efficient and outperformed a few of the existing techniques.

Dyana A. et al. [45] have proposed a system for CBVR based on shape and motion features of the video object. They utilized Curvature scale space for shape representation and Polynomial curve fitting for trajectory representation and retrieval. The shape representation was invariant to translation, rotation and scaling and robust with respect to noise. Trajectory matching incorporated visual distance, velocity dissimilarity and size dissimilarity for retrieval. The cost of matching two video objects was on the basis of shape and motion features, in order to retrieve alike video shots. They tested their system on standard synthetic databases. They also tested their system on real world databases. Experimental results illustrated good performance.

Xiao-Ming Chen et al. [46] have proposed a Vector Quantisation (VQ)-based video retrieval algorithm by means of global motion features. Their contribution comprised of two points: first, they designed a VQ-based algorithm for getting rid of the singular points in the motion vector space; second, they make use of the global motion vector index histogram of all frames in the query video clip to match those of all video clips in the database. The experimental results illustrated that their algorithm could efficiently extract the statistical characteristics of the global motion features.

The properties of spatio-temporal volumes in videos and proposed a novel and robust video matching framework is examined by Arslan Basharat et al. [47]. The basis of the constructed volumes is the clustering of the interest point trajectories. In order to model the appearance of the volumes multiple features comprising color, texture, motion, and interest point descriptors are extracted. In order to calculate the similarity between the two videos the maximum matching problem of the graph formed by the volumes is solved. The employment of their proposed video matching framework in video matching for retrieval attained very challenging as well as competitive performance.

Alexander Haubold, and Apostol (Paul) Natsev [48] have offered an approach for finding out information content from two web-based sources. They also illustrated its application to concept-based video retrieval. Characteristically, the measure employed is the information content. On the other hand, rigorous flaws due to outdated information from a too small text corpus are demonstrated by its underlying database of term frequencies. By means of two approaches of retrieving term frequency from web-based resources it is demonstrated by them that how to produce an enhanced information content database. They carried out widespread validation of both the approaches and determined that in comparison to the commonly used IC on the basis of the Brown corpus. The semantic similarity on the basis of the enhanced information content has in general a considerably better performance.

III. Proposed Content-Based Video Retrieval System by Extensive Features

In this section, we offers an efficient CBVR system which retrieves similar videos based on an extensive feature set. The video database, which is a collection of video sequences, is completely processed offline in our system. The proposed system depicted in the figure 1 illustrates the fundamental operations performed in the system for the retrieval of video clips based on the contents.

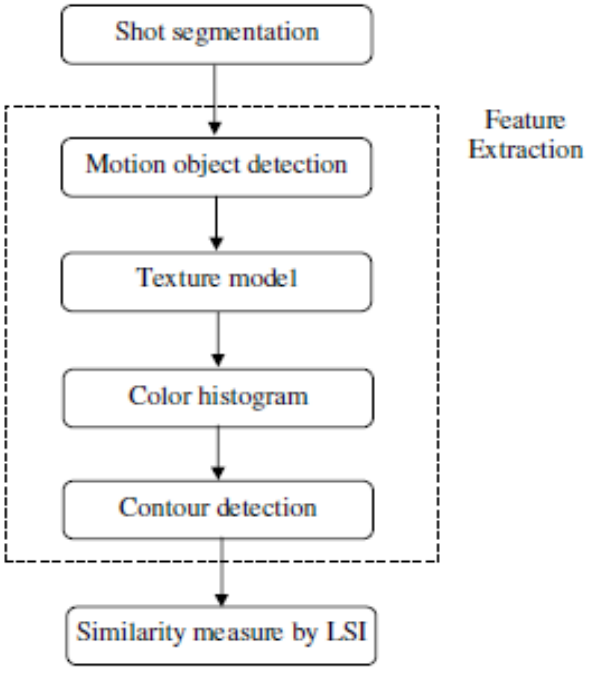


Fig.1: Steps involved in the proposed extensive CBVR system

On the basis of the shots each individual video is segmented and then tracking of the video objects is carried out throughout the frames for every shot. The primary step in our system is to partition a long video sequence into a number of video shots, i.e. shot segmentation, where each shot is the fundamental unit for video retrieval. As a next step, we are extracting four different types of features, namely, motion, texture, color and contour for all the video shots. After that the extracted features are stored in the feature library. Then the same features (above mentioned) are extracted for a query clip (single clip) and are compared with the features in the feature library. By means of the assistance of LSI similarity measure the comparison is carried out. As a final point based on the LSI the videos are retrieved from the videos collection.

A. Shot Segmentation

Prior to carry out any video object analysis it is essential to split the video into —chunks" or video shots. Scene change detection is utilized to carry out the video shot separation, no matter whether it is abrupt scene changes or transitional (e.g. dissolve, fade in/out, wipe). Meng et al. [18] have presented an effective scene change detection algorithm that

functions on compressed MPEG streams. Consequently, the heuristic models of abrupt or transitional scene changes are verified by means of these measurements [19]. Nowadays, the majority of the video retrieval or scene change detection systems make use of shots as the basic element in the video database [23].

A shot is defined as a sequence of frames taken by a single camera without any major variation in the color content of consecutive images. Due to this purpose a number of researchers employed robust techniques on the basis of the color histogram comparison. Hereby, we have utilized an approach to extract the frames belongs to similar shots based on DCT and Euclidean distance. As we have already known, a video is a collection of frames. Consider the given a video clip has n_f number of frames and we have to segment these frames based on the shots in that video. In our segmentation approach, we have divided each frame into several numbers of blocks and DCT will be applied to every block of the frame. The DCT calculation is performed as

After this calculation, we have DCT values for all the blocks. Hence the vector has been formed as $B^{(j)}_{(u,v)_i}$ which is constituted by the DCT values of $n_b \times n_f$ blocks where $i = 1,2,3,\ldots,n_b$ and $j = 1,2,3,\ldots,n_f$. The DCT calculation can be given as

$$B^{(j)}_{(u,v)_i} = \left(\frac{2}{m}\right)^{1/2}\left(\frac{2}{n}\right)^{1/2}\alpha_u\alpha_v \sum_{x=0}^{m-1}\sum_{y=0}^{n-1} b^{(j)}_{(x,y)_i}\cos\left[\frac{\pi u}{2m}(2x+1)\right]\cos\left[\frac{\pi v}{2n}(2y+1)\right] \quad (1)$$

where

$$\alpha_u = \begin{cases}\left(\frac{1}{2}\right)^{1/2} & if \quad u = 0\\ 1 & if \quad 1 \le u \le m-1\end{cases}$$

$$\alpha_v = \begin{cases}\left(\frac{1}{2}\right)^{1/2} & if \quad v = 0\\ 1 & if \quad 1 \le v \le n-1\end{cases}$$

Then as the next step of shot segmentation, Euclidean distance is calculated between two blocks of consecutive frames. This can be performed as follows

$$\left[E_d^{(j)}(i)\right] << \sqrt{\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}\left(B_i^{(j)}(u,v) - B_i^{(j+1)}(u,v)\right)^2} \quad (2)$$

By equation (2), a Euclidean distance for each block will be calculated. Then for a single frame, average of Euclidean distance of n_b blocks will be calculated as

$$\left[M\left(E_d^{(j)}\right)\right] = \frac{1}{n_b}\sum_{i=1}^{n_b} E_d(i) \quad (3)$$

Thus we have obtained a single Euclidean distance value for each frame which is the average of the Euclidean distances of all the blocks of the corresponding frame. This Euclidean distance will be subjected for a criterion which will segment the whole clip into segment of shots as follows

$$M\left(E_d^{(j)}\right) \sim M\left(E_d^{(j+1)}\right) \le 2 \; ; f_j \; and \; f_{j+1} \; are \; same \; shots \quad (4)$$

$$M\left(E_d^{(j)}\right) \sim M\left(E_d^{(j+1)}\right) > 2 \; ; f_j \; and \; f_{j+1} \; are \; different \; shots \quad (5)$$

Based on the criterion described in the equation (4) and (5), the segments are obtained on the basis of shot change. Each frame is tested with the mentioned criteria sequentially and then on the basis of the criteria results, segmentation of the video clips based on the shots change is obtained. Now, we have the n_s number of segmented video clips based on the DCT combined with Euclidean distance. These segmented frames will be subjected for the next process of Motion object detection, a process of extracting the moving object from a frame comprised by motion and still object.

B. Feature Extraction

In this subsection the extraction of features from the segmented shots is explained in detail. For instance, A shot of a person (the person is the "object" here) walking can be segmented into a collection of adjoining regions which are differing in criteria such as shape, color, Edge and texture, on the other hand all the regions may be consistent in their motion attribute. A feature can be considered to be good if and only identical objects are close to each other in the feature space, and different objects are far apart [33]. The features which are extracted from our system are as follows: Motion estimation, Texture model, Color Histogram and Object contour. For all the video shots the abovementioned features are extracted and stored in the feature library.

i. Motion Estimation

Apart from the typical image features for instance color, texture and shape, motion is considered to be the most significant feature in video which signifies two dimensional temporal change of video content [24]. By means of the shot segmentation, a single video clip will be segmented into several video clips based on the change of shot. Each such video clip is a collection of frames which belongs to the same shot. These frames are now subjected for detecting the objects which exhibits motion. In terms of motion, videos and images can be considerably differentiated. Several applications such as motion based segmentation, and structure from motion make use of the motion information. The major focus of this sub-section is the detection of moving object and the estimation of its moving direction. A number of significant applications in the areas of computer vision and video processing make use of the estimation of motion. One of the applications that involve a direct appliance of motion estimation technique [25] is motion compensation based video coding. Several efforts have been

made on motion estimation in the past two decades which is even now one of the most active research areas in video analysis. On one hand, motion can help in finding interesting objects in the video [26]. In our system an innovative approach for motion estimation is proposed. Initially, an operation of subtraction is performed between a selected numbers of frames as

$$\left[d_{f_k}^{(l)}\right] << f_k^{(l)} - f_{k-2}^{(l)} \qquad (6)$$

In equation (6), $k = 2(j+1)-1$. This means odd number of frames is selected for the process of detecting the motion object. Selection like this exposes the motion object clearly because when the operation mentioned in the equation (6) is performed, an abrupt change of position of an object from one state to another obtains clearly. The number of frames thus obtained by the subtraction operation is given by

$$\left|d_{f_k}^{(l)}\right| = \begin{cases} \dfrac{n_f^{(l)}}{2} & if \;\; n_f \%2 = 0 \\ \dfrac{n_f^{(l)}}{2} + 1 & else \end{cases} \qquad (7)$$

Then the resultant frames are converted into binary frames. Over these binary frames, a morphological operator is applied. On these binary frames the operator undergoes morphological closing and returns the closed frames. ‗Disk' is the structuring element used here to perform the suggested morphological operation. After the operation, the objects which exhibit motion in each frame are extracted. Then the motion direction is determined by calculating gradient and then orientation as follows

$$\nabla f_k^{(l)} = \left[\frac{\partial}{\partial x} f_k^{(l)}(x,y), \frac{\partial}{\partial y} f_k^{(l)}(x,y)\right] \qquad (8)$$

In equation (8), gradient of the k^{th} block of the l^{th} shot is determined. Then the further vectors are determined as

$$g_{xy_k}^{(l)} = \left[\frac{\partial}{\partial x} f_k^{(l)}(x,y)\right] . \left[\frac{\partial}{\partial y} f_k^{(l)}(x,y)\right] \qquad (9)$$

$$g_{xx_k}^{(l)} = \left[\frac{\partial}{\partial x} f_k^{(l)}(x,y)\right]^2 + r \qquad (10)$$

$$g_{yy_k}^{(l)} = \left[\frac{\partial}{\partial y} f_k^{(l)}(x,y)\right]^2 + r \qquad (11)$$

Then with the gradient calculation, orientation vector is determined for each frame as

$$G_{xy_k}^{(l)} = \sum_{a=0}^{m_g}\sum_{b=0}^{n_g} g_{xy_k}^{(l)}(a,b) \qquad (12)$$

$$G_{xx_k}^{(l)} = \sum_{a=0}^{m_g}\sum_{b=0}^{n_g} g_{xx_k}^{(l)}(a,b) \qquad (13)$$

$$G_{yy_k}^{(l)} = \sum_{a=0}^{m_g}\sum_{b=0}^{n_g} g_{yy_k}^{(l)}(a,b) \qquad (14)$$

Then the angle of direction of motion is determined as

$$\theta_k^{(l)} = \begin{cases} \dfrac{1}{2}\tan^{-1}\left(\dfrac{2G_{xy_k}^{(l)}}{G_{xx_k}^{(l)} - G_{yy_k}^{(l)}}\right) & ; if \; E_{k_1}^{(l)} \geq 0 \\ \dfrac{1}{2}\left(\tan^{-1}\left(\dfrac{2G_{xy_k}^{(l)}}{G_{xx_k}^{(l)} - G_{yy_k}^{(l)}}\right) + \pi\right) & ; \; if \; E_{k_1}^{(l)} < 0 \; and \; E_{k_2}^{(l)} > 0 \\ \dfrac{1}{2}\left(\tan^{-1}\left(\dfrac{2G_{xy_k}^{(l)}}{G_{xx_k}^{(l)} - G_{yy_k}^{(l)}}\right) - \pi\right) & ; \; else \end{cases} \qquad (15)$$

In equation (15), two conditional variables are used which can be given as $E_{k_1}^{(l)} = G_{xx_k}^{(l)} - G_{yy_k}^{(l)}$ and $E_{k_2}^{(l)} = 2G_{xy_k}^{(l)}$. From the equation (15), the angle of direction $\theta_k^{(l)}$ of the object motion that is distributed over all the quadrants is determined. Now the quadrant of $\theta_k^{(l)}$ changed as adaptive for further process as follows

$$\theta_{\left(-\frac{\pi}{2},\frac{\pi}{2}\right)} = \begin{cases} \theta + \pi/2 \; ; \; if \;\; \theta \leq 0 \\ \theta - \pi/2 \; ; \; else \end{cases} \qquad (16)$$

Thus obtained direction of the motion object is stored in the feature library for further process of the proposed CBVR.

ii. Extraction Of Texture Features

For the proposed CBVR, the next feature to be extracted is the texture pattern of the frames. Hereby, we are using Gaussian mixture model to extract the texture features of frames. Before employing the Gaussian Mixture Model, the frames in RGB color space are converted to YCbCr color space. Then they are subjected for Gaussian Mixture Modeling. Mixture Models are a type of density model which consist of a number of component functions, generally Gaussian. In order to offer a multimodal density, the component functions are joined. Mixture models are a semi-parametric substitute to non-parametric histograms [20] (which can also be used as densities) and offer better flexibility and precision in modeling the underlying statistics of sample. They are capable of smoothing over gaps resulting from sparse sample data and offer tighter constraints in assigning object membership to color-space regions. A precision like this is essential in order to acquire the best results possible from color-based pixel classification for qualitative segmentation requirements.

Once a model is produced, conditional probabilities can be computed for color pixels [21]. Gaussian mixture models can also be seen as a form of generalized radial basis

function network in which every Gaussian component is a basis function or `hidden' unit. The component priors can be viewed as weights in an output layer. Finite mixture models have also been discussed at length elsewhere [20], [22], [27] and [28] even though the majority of this work has concentrated on the general studies of the properties of mixture models rather than developing vision models for use with real data from dynamic scenes.

Given a texture vector which is indicated as where n is the dimension of the feature vector. We model the distribution of all samples by the following formula [29]:

$$p(x \mid \lambda) = \sum_{i=1}^{M} \omega_i p_i(x) \qquad (17)$$

where $p_i(x)$ is a normal PDF which is a component of the GMM. It is parameterized by a mean vector μ_i , and a covariance matrix R_i :

$$p_i(x) = \frac{1}{2\pi \mid R_i \mid^{1/2}} \exp\left[\frac{-1}{2}(x-\mu_i)^T R_i^{-1}(x-\mu_i)\right] \qquad (18)$$

ω_i is the weight of the component $p_i(x)$, $0 < \omega_i < 1$ for all components, and $\sum \omega_i = 1$. Mixture model specified in the equation (17) is called the Gaussian Mixture Model (GMM) [29].

In our work, GMM is applied for texture extraction as stated earlier. Prior to model, each frame is partitioned into several blocks and then DCT calculation is applied to each block followed by zigzag scanning of the same. The Gaussian mixture model for texture for image retrieval is detailed clearly in [29], [31] and [32]. Based on the GMM, clustering is performed, which is among the majority of statistically mature methods for clustering. GMM dedicatedly performing for clustering operation is given in [34] provides n_c number of clusters. The covariance matrix R is calculated for the n_c^{th} cluster. The covariance matrix of a frame R_j is determined by the mean value of the covariance matrix of the blocks corresponds to the frame.

iii. Color Histogram

The color histogram is widely used as an important color feature indicating the content of the image [35], due to its robustness to scaling, orientation, perspective, and occlusion of images [30]. Initially, a smoothening operation is performed over each frame of the shot segmented clips. Anisotropic diffusion is utilized in our proposed CBVR approach for the smoothening the frames, prior to color histogram

Anisotropic diffusion is an excellent method for image denoising, which was first introduced by Perona and Malik. It prefers intra-region smoothing to inter-region smoothing, and is thus capable to smooth the noise when keeping the boundaries from being smoothed [36]. For $f(\cdot,t)$, a sequence of frames, the anisotropic diffusion is given by

$$\frac{\partial f}{\partial t} = div\left(c(a,b,t)\nabla f\right) \qquad (19)$$

where, $div(\cdot)$ is the divergence operator, $c(a,b,t)$ is the diffusion co-efficient and ∇f denotes the gradient. From the Anisotropic diffusion, smoothed frames are obtained in the RGB color space. For these frames color histogram is obtained, by means of a three dimensional vector representing RGB respectively followed by a normalization function which converts the three dimensional vector into a single dimensional vector. Hence the color histogram, another important feature for the proposed CBVR scheme is extracted.

iv. Contour Detection

Contour (or) Shape features generally necessitate that images have segmentation information available which makes them useless for numerous tasks. On the other hand, different edge histograms that can be taken into account as statistical shape features do not want segmentation which has made them popular in Content Based Retrieval applications [37]. As initial process, the RGB image is converted into grayscale and then binary image. For both the combination of the grayscale and the binary image, sparse field method, an active contour method is applied. The sparse field level set method which was introduced by Whitaker in 1998 eliminates the time complexity [38]. It makes use of a set of linked lists to track the active voxels around the interface. This permits incremental extension of the active region as needed without incurring any important overhead. Though it provides less computational complexity, it requires more storage space.

The sparse-field algorithm utilizes an approximation to the distance transform that supports to re-compute the neighborhood of the level-set model in every time step. Therefore, it obtains the maximum narrow-band strategy; determines one point wide updates on a band of grid points. By means of the up-wind scheme and the mean-curvature flow the values of the points in the active set can be updated. While computing updates on so few points maintaining a neighborhood around those points is must therefore accurate derivatives can be determined that are controlling the process. The strategy is to make the embedding wide from the active points outward in layers in order to generate a neighborhood around those points that is precisely the width required to calculate the derivatives for the next time step.

This approach has a number of advantages. The algorithm does exactly the number of calculations required to compute the next position of the level curve. It does not require explicitly recalculating the positions of level sets and their distance transforms. In the sparse-field algorithm the number of points being computed is so little, it is possible to use a linked-list to keep track of them. In addition, the sparse-field approach recognizes a single level set with a particular set of points whose values control the position of

that level set. This permits one to compute external forces to an accuracy that is superior to the grid spacing of the model [38] [39]. So, the contour of the objects, one of the significant features from all the shot segmented frames is obtained.

C. Retrieving Video Clips Based On The Query Clip

For all the video shots in the database the motion orientation, texture model, color histogram and shape features are extracted by means of the approaches stated in the preceding subsections and stored in the feature library. Based on the input query clip, in our CBVR system, the related videos are retrieved from the video database in our content-based video retieval system. For the query clip the abovementioned features are calculated and evaluated against the features in the feature library. With the help of the similarity measure the comparison of the features is performed. In our system we have employed the LSI as similarity measure and so the similar videos are retrieved as a final point.

i. LSI Based Similarity Measure

For a specific retrieval system when the feature set is fixed, the researchers can possibly develop is the similarity measure. Obviously, the similarity measure plays an important role as the original feature space in deciding —close" or —far". Among the broadly used similarity measures euclidian distance and other Minkowski-type distances are a noteworthy few. For similarity measure computation we have utilized LSI method for in our system.

For information retrieval, Matrix computation is used as a basis in the retrieval strategy called LSI [40]. The premise is that more conventional retrieval strategies all have problems because they match directly on keywords. Since by means of many different keywords the same concept can be described, this type of matching is prone to failure. LSI attempts to search for something that is closer to represent the underlying semantics of a document. By means of matrix computation the searching is performed in particular Singular Value Decomposition (SVD). The noise found in a document is out by these filters, such that two documents that have same semantics will be located close to one another in a multi-dimensional space [41].

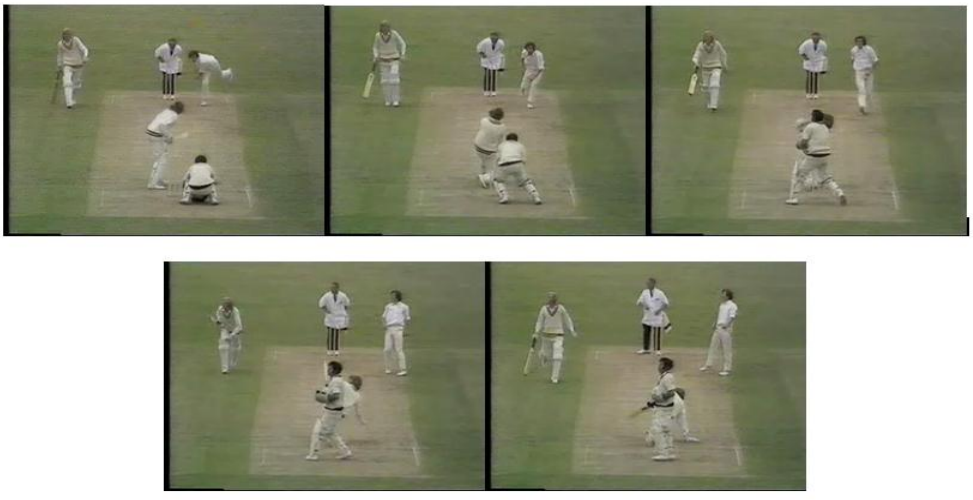
For the query clip, the required features are extracted as discussed in the sections. The query clip feature and each element of feature library are applied for LSI. The LSI determines the similarity of the query clip feature with the feature library element. The LSI procedure involves Singular value decomposition (SVD) of the term- document matrix A which can be given as

$$A = USV^T \qquad (20)$$

In the equation (20), $S \in R_{m\times n}$ is a diagonal matrix with nonnegative diagonal elements called the singular values, $U \in R_{m\times n}$ and $V \in R_{m\times n}$ are orthogonal matrices. . The columns of matrices U and V are called the left singular vectors and the right singular vectors respectively. The decomposition can be computed in such a way that the singular values are sorted by decreasing order. Only a few largest singular values of A and the corresponding left and right singular vectors are computed and stored in memory rather then finding out all. The number of singular values and vectors which are computed and kept in memory can be selected as a compromise between the speed and precision ratio of the LSI procedure. Based on the mean similarity of all the features, the video clips in database which are alike to the query clip are obtained. Thus obtained video clips are similar to the query clip that is retrieved on the basis of contents of the clips.

IV. Results and Discussion

The proposed CBVR system has been validated by querying some video clips and we have checked the resultant clips retrieved. For the purpose of validation, we have implemented the proposed system in the working platform of MATLAB (version 7.4). We have affixed some of the intermediate results obtained at the end of each step included in our retrieval system.



(a)

(b)

Fig. 2: Shots obtained after shot segmentation from a video clip (a) frames belongs to one shot (b) frames belongs to another shot

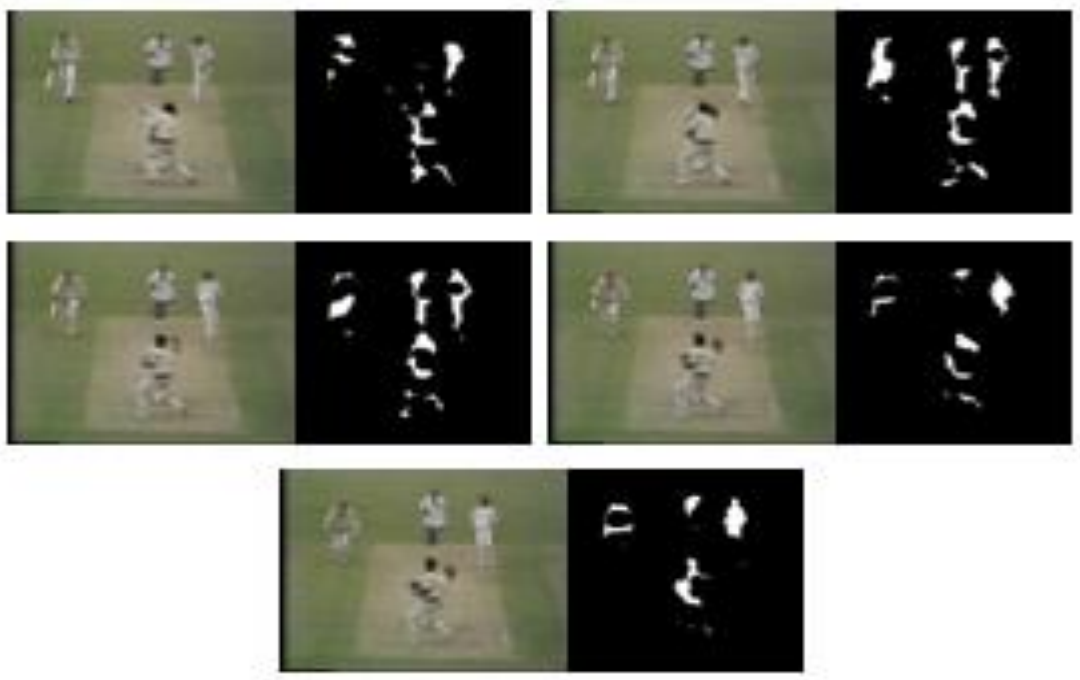


Fig. 3: Motion object detection from segmented video clip

Fig. 4: YCbCr color space converted clips from RGB color space for Gaussian mixture model

Fig. 5: Sample frames and their corresponding color segmented frames

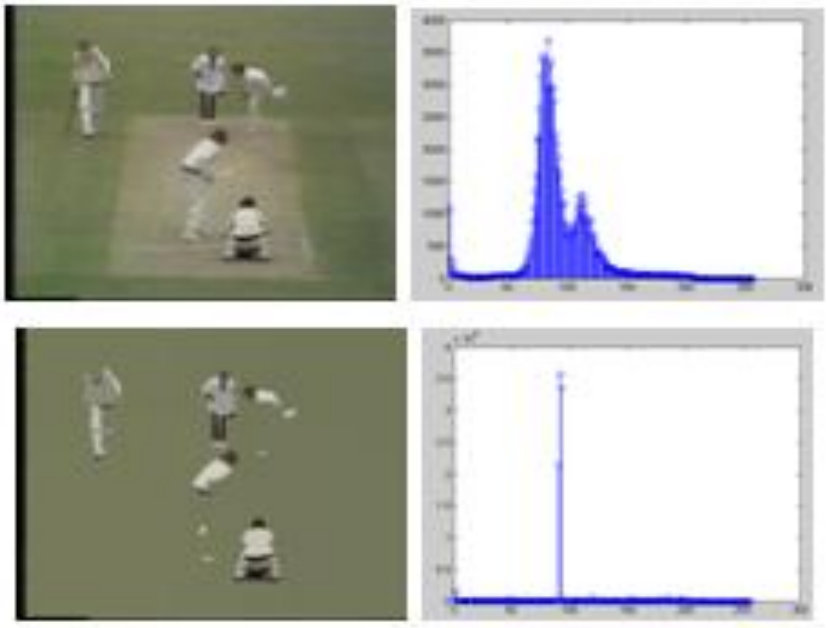


Fig.6: Color histogram for normal frames and the color segmented frames

Fig. 7: Contour of the object present in the corresponding frames extracted from the shots

(i)

(ii)

Fig. 8 (a): Results of the proposed CBVR system: (i) input query clip and the (ii) retrieved video clips

(i)

(ii)

Fig. 8 (b): Results of the proposed CBVR system: (i) input query clip and the (ii) retrieved video clips

Figure 2 shows some sample frames of the segmented shots which are obtained from a single video clip. Figure 3,4,5 and 6 illustrates extracted objects which exhibits motion in the shots obtained after the application of morphological operator, RGB to YCbCr color space conversion output for Gaussian mixture model, some sample frames and the corresponding color segmented frames based on Anisotropic diffusion and color histogram for the shots respectively. Figure 7 depicts the object contour of some sample frames of the shots. Figure 8 details some input query clips and the corresponding video clips retrieved by the proposed content based-video retrieval approach. For the given query clips, on the basis of videos retrieved, we have determined the precision and recall for the proposed system. The precision and recall measurements are the more frequently using measurements to analyze the performance of an image retrieval system which can be defined as

$$precision = \frac{Number\ of\ relevant\ images\ selected}{Total\ number\ of\ retrieved\ images}$$

$$recall = \frac{Number\ of\ relevant\ images\ selected}{Total\ number\ of\ similar\ images\ in\ the\ dataset}$$

Some of the query clips utilized for the precision and recall calculation is given in the figure 9.

fig. 9: The input query clips, Q1, Q2, Q3, Q4 and Q5 given to the proposed system for precision and recall calculation

For the entire given query clips Q1, Q2, Q3, Q4 and Q5 as illustrated sequentially in the figure 9, precision and recall have been determined and the values are tabulated in the table 1.

Table 1: Precision and Recall for a given set of query clips

Sl. No.	Query clip	Precision	Recall
1	Q1	0.13	0.93
2	Q2	0.29	0.87
3	Q3	0.38	0.85
4	Q4	0.76	0.39
5	Q5	0.89	0.28

In the table 1, we have given a sample of five query clips and the corresponding precision and recall values and figure 10(a) and 10(b) gives the precision-recall plot for the query clips

.

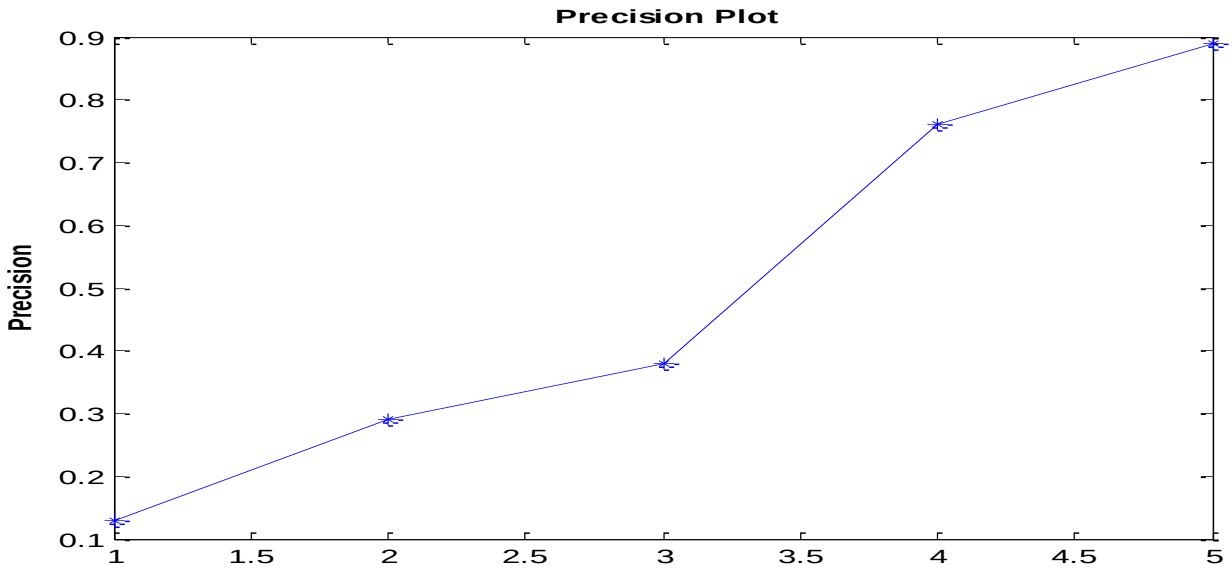


fig. 10(a): precision plot for the input query clips

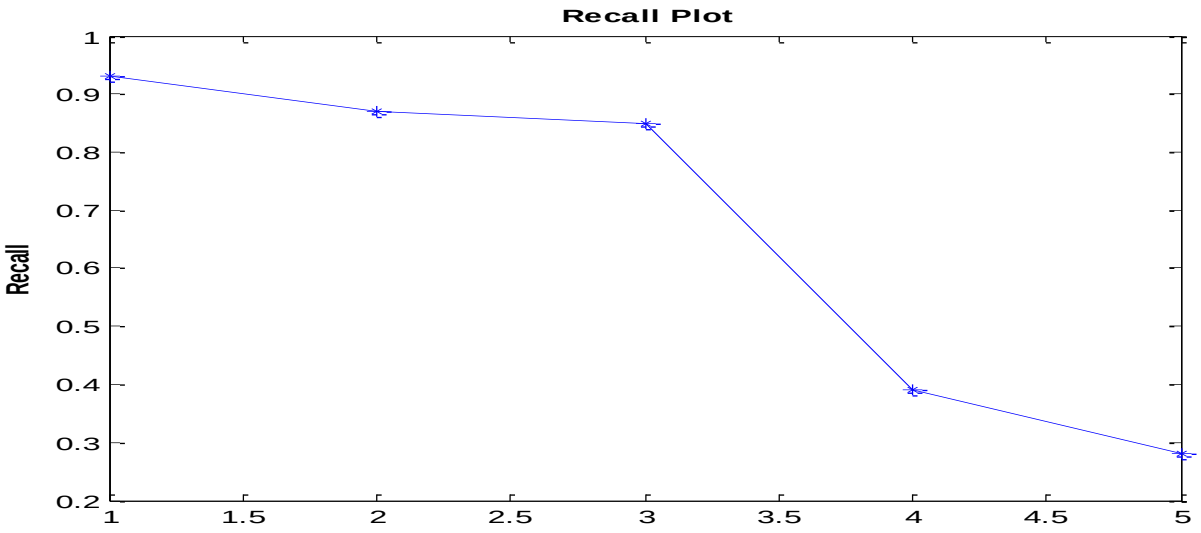


fig 10(b): Recall plot for the input query clips

For the given five input query clips Q1, Q2, Q3, Q4 and Q5, we have plotted the precision – recall graph. From the graph, it can be decided that the proposed system exhibits a good acceptable precision and recall which proves the system as more effective.

V. Conclusion

The content-based retrieval of visual information is a growing area of research that has gained the attention of the research community in the recent past. We have offered an extensive CBVR system in this paper. With the help of the proposed approach the elementary shots in the long video sequence are separated effectively. Then, in the feature library, the features like motion vector, texture model, color histogram and objects' contour of the video sequence are extracted and stored. Similarly the features of the query clip are extracted and evaluated against features in the feature library. By LSI the comparison is carried out efficiently. Related videos are retrieved from the video database more effectively as the retrieval is based on the four different features. Use of the LSI for similarity measure added more proficiency to the proposed CBVR system.

VI. References

1) Steven C.H. Hoi and Michael R. Lyu, "A Multimodal And Multilevel Ranking Framework

For Content-Based Video Retrieval", 2007 International conference on Acoustics, speech, and Signal processing, Hawaii, USA, 15-20 April 2007.
2) Yoshitaka, T. Ichikawa, —A Survey on Content-Based Retrieval for Multimedia Databases", IEEE Transactions on Knowledge and Data Engineering, vol. 11, no. 1, pp. 81-93, 1999.
3) M. Petkovic, W. Jonker, "Content-Based Video Retrieval by Integrating Spatio-Temporal and Stochastic Recognition of Events", Proceedings of IEEE Workshop on Detection and Recognition of Events in Video, pp. 75-82, 2001.
4) M. Petkovic, "Content-based Video Retrieval", VII Conference on Extending Database Technology (EDBT), Ph.D. Workshop, Konstanz, Germany, March 2000.
5) P. Aigrain, H. Zhang, D. Petkovic, Content based Representation and Retrieval of Visual Media: A State-of-the-Art Review, Multimedia Tools and Applications, Kluwer Academic Publishers, vol. 3(3), 1996, 179-202.
6) R. Smith, S.F. Chang, —Visual SEEk: A Fully Automated Content-Based Image Query System", ACM Multimedia Conference, Boston, MA, November 1996.
7) Pentland, R. W. Picard, S. Sclaroff, —Photobook: Content-Based Manipulation of Image Databases", Int. J. Computer Vision, vol. 18, no. 3, pp. 233-254.
8) Carson, M. Thomas, S. Belongie, J. M. Hellerstein, J. Malik, —Blobworld: A System for Region-Based Image Indexing and Retrieval", Third Int. Conf. On Visual Information and Information Systems, Amsterdam, 1999, pp. 509-516.
9) Hampapur, A. Gupta, B. Horowitz, C.F. Shu, C. Fuller, J. Bach, M. Gorkani, R. Jain, —Virage Video Engine", SPIE Vol. 3022, 1997.
10) Ponceleon, S. Srinivasan, A. Amir, D. Petkovic, D. Diklic, —Key to Effective Video Retrieval: Effective Cataloging and Browsing", ACM Multimedia, '98, pp. 99-107.
11) S-F. Chang, W. Chen, H. Meng, H. Sundaram, D. Zhong, —A Fully Automated Content Based Video Search Engine Supporting Spatio-Temporal Queries", IEEE Transaction on Circuits and Systems for Video Technology, Vol. 8, No. 5, Sept., 1998.
12) Chi-Jiunn Wu Hui-Chi Zeng Szu-Hao Huang Shang-Hong Lai Wen-Hao Wang, "Learning-Based Interactive Video Retrieval System", Proceedings of IEEE International Conference on Multimedia and Expo., pp: 1785-1788, 9-12 July 2006,
13) M. Ferman, A. M. Tekalp, and R. Mehrotra, —Robust color histogram descriptors for video segment retrieval and identification," IEEE Transactions on Image Processing, Vol. 11, No. 5, pp 497-508, 2002.
14) Erol, and F. Kossentini, —Shape-based retrieval of video objects," IEEE, Trans. on Multimedia, Vol. 7, No. 1, pp 179-182, 2005.
15) C.W. Ngo, T.C. Pong, H.J. Zhang, —Motion-based video representation for scene change detection," Int. Journal Computer Vision, pp 127-142, 2002.
16) L. Chen and T.S. Chua, —A match and tiling approach to content-based video retrieval," Proc. ICME, pp. 301-304, 2001.
17) M.Y. Chen and A. Hauptmann, —Searching for a specified person in broadcast news video," Proc. ICASSP, Vol. 3, pp 1036-1039, 2004.
18) Meng, Y. Juan, S.F. Chang, Scene Change Detection in a MPEG Compressed Video Sequence, SPIE Symposium on Electronic Imaging: Science and Technology - Digital Video Compression: Algorithms and Technologies, SPIE Vol. 2419, San Jose, Feb. 1995.
19) Shih-Fu Chang, William Chen, Horace J. Meng, Hari Sundaram, Di Zhong, "VideoQ: An Automated Content Based Video Search System Using Visual Cues", Proceedings of the fifth ACM international conference on Multimedia, Seattle, Washington, United States, pp. 313-324, 1997.
20) Bishop, —Neural Networks for Pattern Recognition", Oxford University Press, ISBN 0-19-853864-2 1995.
21) S. McKenna, S. Gong, and Y. Raja, —Face recognition in dynamic scenes", Proceedings of the British Machine Vision Conference 1997, BMVC 1997, University of Essex, UK.
22) S. Everitt and D. J. Hand, Finite Mixture Distributions, Chapman and Hall, ISBN 0412224208, New York, 1981.
23) Yang Liu, Weiqiang Wang, Wen Gao, Wei Zeng, "A novel compressed domain shot segmentation algorithm on H.264/AVC", in Proc. of International Conference on Image Processing, ICIP 2004, 24-27 October 2004.
24) Y. F. Ma and H. J. Zhang, —Detecting Motion Object by Spatio-Temporal Entropy", IEEE International Conference on Multimedia and Expo, Tokyo, Japan, August 22-25, 2001.
25) Otsuka, T. Horikoshi, S. Suzuki and M. Fujii, —Feature Extraction of Temporal Texture Based on Spatio-Temporal Motion Trajectory", Proceedings of the 14th International Conference on Pattern Recognition, ICPR'98, pp.1047-1051, Aug. 1998.
26) P. Bouthemy, R. Fablet, —Motion Characterization from Temporal Co-occurrences of Local Motion-Based Measures for Video Indexing", Int. Conf on Pattern Recognition, ICPR'98, pp.905-908, Vol. 1, Australia, Aug. 1998.
27) McLachlan and K. E. Basford, —Mixture Models: Inference and Applications to Clustering", Textbooks and Monographs, Marcel Dekker Inc., New York, 1988.

28) M. Titterington, A. F. M. Smith, and U. E. Makov, Statistical Analysis of Finite Mixture Distributions, John Wiley, New York, 1985.
29) Gu, J., Chen, J. and Zhang, H., ―Gaussian mixture model of texture for extracting residential area from high-resolution remotely sensed imagery", ISPRS Workshop on Updating Geo-spatial Databases with Imagery, 157-162, 2007-08
30) W.Y. Ma and H. Zhang, ―Content-Based Image Indexing and Retrieval," Handbook of Multimedia Computing, CRC Press, 1999.
31) Permuter, H. Francos, J. Jermyn, I.H., ―Gaussian mixture models of texture and colour for image database retrieval" In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003, (ICASSP '03), April 2003, Vol. 3, p.p. 569-72 ISSN: 1520-6149, DOI. 10.1109/ICASSP.2003.1199538
32) Wu, Y., Chan, K.L., and Huang, Y., ―Image texture classification based on finite Gaussian mixture models", In 3rd International Workshop on Texture Analysis and Synthesis, ICCV'03, France, M. Chantler (Ed.), pp. 107-112, 2003.
33) Tsuhan Chen, "From Low-Level Features to High-Level Semantics: Are We Bridging the Gap?", Proceedings of the Seventh IEEE International Symposium on Multimedia, pp. 179, 2005.
34) Hichem Sahbi, ―A particular Gaussian mixture model for clustering and its application to image retrieval", Source Soft Computing - A Fusion of Foundations, Methodologies and Applications archive, Vol. 12 , No. 7, p.p. 667-676, 2008, ISSN:1432-7643, DOI. 10.1007/s00500-007-0247-y
35) Wang, S.L. Liew, A.W.C., "Information-Based Color Feature Representation for Image Classification", IEEE International Conference on Image Processing, ICIP 2007, Volume: 6, p.p. 353- 356, ISSN: 1522-4880, DOI: 10.1109/ICIP.2007.4379594.
36) Yongsheng J, Pan, Douglas Birdwell, Seddik M. Djouadi, "Image Segmentation Using Curve Evolution and Anisotropic Diffusion: An Integrated Approach", Proceedings of the Seventh IEEE International Symposium on Multimedia, IEEE Computer Society Washington, DC, USA, p.p 387 – 394, 2005, ISBN:0-7695-2489-3, DOI: 10.1109/ISM.2005.68.
37) Rami Rautkorpi and Jukka Iivarinen, "A Novel Shape Feature for Image Classification and Retrieval", Lecture Notes in Computer Science, Image Analysis and Recognition, Springer Berlin / Heidelberg, Volume. 3211/2004, p.p, 753-760, 2004, DOI 10.1007/b100437
38) Ross T. Whitaker, ―ALevel-Set Approach to 3D Reconstruction from Range Data", International Journal of Computer Vision, Vol. 29 (3), September 1998, P.P: 203 – 231, 1998, ISSN: 0920-5691, DOI. 10.1023/A: 1008036829907.
39) Museth, D.E. Breen, R.T. Whitaker, S. Mauch and D. Johnson, ―Algorithms for Interactive Editing of Level Set Models," Computer Graphics Forum, Vol. 24, No. 4, pp. 821-841, December 2005.
40) S. Deerwester, S. Dumais, G. Furnas, T. Landauer, R. Harshman, ―Indexing by latent sematic analysis", Journal of the American Society for Information Science, 41(6):391-407, 1990.
41) P. Praks, J. Dvorský, and V. Snášel. Latent semantic indexing for image retrieval systems. In SIAM Conference on Applied Linear Algebra. The College of William and Mary, Williamsburg, USA, 2003, http://www.siam.org/meetings/1a03/proceedings/Dvorsky.pdf
42) Arnab Ghoshal, Sanjeev Khudanpur, Dietrich Klakow, "Impact of Novel Sources on Content-Based Image and Video Retrieval", IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2009, Taipei, Taiwan, April 19-24, 2009
43) Dyana and Sukhendu Das, "Trajectory representation using Gabor features for motion-based video retrieval," Pattern Recognition Letters, Vol. 30, No. 10, pp. 877-892, 2009.
44) Liang-Hua Chen, Kuo-Hao Chin, Hong-Yuan Mark Liao, "Integration Of Color And Motion Features For Video Retrieval," International Journal of Pattern Recognition and Artificial Intelligence, Vol. 23, No. 2, pp. 313-329 , 2009.
45) Dyana A., Subramanian M.P., Das S., "Combining Features for Shape and Motion Trajectory of Video Objects for Efficient Content Based Video Retrieval," Proceedings of Seventh International Conference on Advances in Pattern Recognition, pp. 113-116, February 4-6, Kolkata, 2009.
46) Xiao-Ming Chen, Zhe-Ming Lu, Zhen L., "Video retrieval using VQ-based global motion features," Proceedings of 5th International Conference on Visual Information Engineering, pp. 577-581, July 29-August 1, Xian China, 2008.
47) Arslan Basharat, Yun Zhai, Mubarak Shah, "Content based video matching using spatiotemporal volumes", Computer Vision and Image Understanding, vol.110, pp. 360–377, 2008.
48) Alexander Haubold, Apostol (Paul) Natsev, "Web-based Information Content and its Application to Concept-Based Video Retrieval", Proceedings of the 2008 international conference on Content-based image and video retrieval, pp. 437-446, 2008.

Service Productivity Optimization Using Genetic Algorithm

GJCST Computing Classification
H.3.4

[1]S. Tackenberg, [2]S. Duckwitz

Abstract- **The efficient development and prospective evaluation of valid service scenarios is a critical success factor for companies. With an increasing complexity of service provision the planning process gains importance. But the process of planning a service and optimizing the underlying service plan still lacks adequate knowledge and support. Existing approaches are not suitable for an optimization of services with regard to productivity prior to their implementation. To achieve efficient services, the concepts for service productivity are reviewed and novel aspects of a developed Genetic Algorithm (GA) for solving the resource constraint service scheduling problem are presented. The GA is verified and validated based on a case study in the German chemical industry. Due to the extensive documentation of this company, detailed information about task structure and workload could be gathered and used for model development. The results of the validation study confirm the structural and external validity of the developed project model. Therefore, the proposed GA-based optimization method has demonstrated its validity and utility in this particular application domain.**

Keywords- optimization, Genetic Algorithm, Service Productivity, Service Management.

I. Introduction

Service science focuses on studying the characteristics of service processes intending to simultaneously reduce service lead-times, reduce cost, and improve quality. Several formal theories of service management have been published (Koumpis, 2010). While the relevant authors fail to agree on a single theory of best service processes as a benchmark for a service enterprise, four facts in terms of planning service operations are generally regarded as important: (1) Uncertainties regarding the duration of activities are a typical characteristic of service projects (Blackburn, 1991), (2) predecessor constraints of tasks within a (service) project are not strictly specified and are therefore often refined with respect to the situation and the person (Eppinger et al., 1994), (3) the assignment of actors and resources (tools, facilities) to tasks is a complex resource constraint scheduling problem whose solution determines the service lead-time (Hildum, 1994), (4) iterations are a characteristic of complex services as well as a major source of unexpected rework and budget overruns (Reijers, 2003). Due to these facts, both the prospective evaluation of uncertainty and the consideration of complexity have become central issues in developing a service organization. Research in this field has shown that modeling provides an avenue to improved understanding and optimization of service operations (Barnard, 2006). Optimization of a service organization and the underlying service processes can be achieved in several ways:

Improving the sequence of activities to streamline the workflows, which directly determine the service provision (Eppinger et al., 1994); developing simulation models to describe, evaluate and minimize uncertainty of a service process due to adding actors and resources to bottle neck tasks (Yassine & Braha, 2003). Instability can be reduced by reordering tasks and considering alternative options so that the effects of delayed activity processing and feedback are minimized. Thereby, an ideal work organization is defined as one with no iterations (Yassine et al., 2007) and which fully achieves the planning objectives. However, in real complex service projects, an optimal sequence of activities as well as assignment of actors and resources is unlikely to be uncovered. Thus, a heuristic for the development of an optimal service provision is still missing. Therefore this paper focuses on a genetic optimization heuristic to improve the productivity of service projects. The approach has to be suitable for decision support in the context of developing a service plan.

II. Service Productivity

A theoretical concept or a procedure model to evaluate the effectiveness of services is missing. To close the lack of knowledge the German Federal Ministry of Education and Research set up the research program „productivity of services" in 2009 to identify the success parameters of efficient services. The incentive for such a research program was the missing of a general theory for service productivity. Existing definitions of service productivity are based on elements of definitions for production processes. In the context of production, the term productivity is understood as the alignment between the entire output and the input used (Ojasalo, 1999). A measurement of the input parameters, e.g. labor, capital, energy, material and information (Sink, 1985) is a central challenge for service science. But also the definition of a valuation baseline is compared to the evaluation of a production processes difficult to achieve due to weakly structured service operations and a high impact of actors' behavior (Ojassalo, 1999). In particular for service providers, the dependence on superordinate workflows leads

About 1st author: Sven Tackenberg completed an apprenticeship as Electronician in the German electronics industry in 2000. He studied industrial engineering at the Technical University of Kaiserslautern, Oita University and received his degree from Paderborn University in 2006. Since June 2007 he is research assistant at the Institute of Industrial Engineering and Ergonomics at RWTH Aachen University.
(telephone: +49 241 80 99462 email: s.tackenberg@iaw.rwth-aachen.de
About 2nd author: Sönke Duckwitz received the Dipl.-Ing. and the Dipl.-Wirt.Ing. degrees from RWTH Aachen University, Germany, in 2003 and 2006, respectively. He is currently a research assistant at the Institute of Industrial Engineering and Ergonomics at RWTH Aachen University.
(telephone: +49 241 80 99461 email: s.duckwitz@iaw.rwth-aachen.de)

to a close cooperation between different actors (Cerasale, 2004). Similar to the lack of methods for a precise measurement of the input parameters, an approach for the evaluation of output parameters is missing. Apart from the quantitative parameters (quantity of service per time period, costs, utilization of the workforce) especially the ―soft" parameters for the evaluation of service productivity are of high relevance, e.g. customer satisfaction, workload of employees (Karapidis, 2005). Although, the comprehension of the term ―service productivity" is related to the definition of production productivity (Johnston & Jones, 2003) a general approach to determine a service specific basis of valuation is missing. On the basis of experience with simulation models for the development of production processes a simulation based evaluation of service productivity seems promising. Therefore, the following aspects have to be considered:

i. The measurement of productivity is distorted due to physical distinction during the analysis of inputs and outputs, which are scaled in most cases in different units (Johnston & Jones, 2003).
ii. The high number of combinations between different characteristic of outputs and inputs as well as missing restriction of productivity ratios does not provide valid indicators of performance (Sink, 1985; Johnston & Jones, 2003).
iii. The significance of the productivity ratio is limited due to the insulated view from other performance factors (e.g. the demanded productivity or internal bench marks) (Sink, 1985).
iv. The characteristics of elements (activities, actors etc.) and relations (iterations, overlapping of activities etc.) of a service organization are fraught with uncertainty. Groenross (2000) describes a service provision as an open process by which changes in the input variables (such as employees, systems, physical resources and technologies, and customer participation) influence the efficiency of the process.
v. The behavior and the decision making of actors have a high impact on service productivity (Cerasale, 2004). This must take into account that services operations are not exclusively designed and implemented by the service provider due to the fact that there are strong interrelations to the customers as well as the suppliers (Edvardsson & Olsson, 1996).

Ojasalo (1999) assumes that existing economic approaches cannot adequately describe or measure service productivity. In this context (Groenross & Ojasalo, 2004) asserted that the comprehensive analysis of service productivity is still at the beginning. Due to the fact that only a few papers about service productivity were published during the last five years, a research gap can still be identified. Therefore, only a few detailed models can be found in literature that explain the complexity and multidimensionality of service productivity, e.g. Corsten & Stuhlmann (1998), Groenross & Ojasalo (2004), and Johnston & Jones (2003). However, all these models are very generic and limited with regard to their applicability by service managers. Thus, an efficient design of service operations and the application of the methods to real service processes are missing.

The approach presented in this paper focuses on a simulation based model to measure service productivity. It is suitable for the decision support in the context of service design and considers the cooperation of enterprises, suppliers and subagents. Thereby, retrospective methods of analysis are not accessible for the prospective evaluation of different service scenarios (Winkelmann & Luczak, 2006) and are not considered for the development of an optimization heursitic.

III. Genetic Algorithm for the Optimization of Service Productivity

Genetic Algorithms (GA) are robust stochastic search algorithms inspired by the process of biological evolution and are often used for scheduling problems (Koza et al., 2003). This class of optimization heuristics considers a set of population of solutions as opposed to only one solution throughout local search. Following an initial population with several individuals through a random generation, new individuals are produced by mating two existing ones (crossover operation) and/or by altering an existing one (mutation operation). Thereby, an individual represents a solution to the optimization problem. Selection scheme determines which individuals survive via evaluating fitness or objective function. The fittest individuals take over next generation while the others are deleted. This optimization process finishes with a convergence to an optimal or near-optimal solution. The structure and the operations of Standard Genetic Algorithms (SGA) for scheduling problems are widely mentioned in literature (Koza et al., 2003) and are not further covered in this paper. The novelty of the presented GA is the characteristic of a chromosome, the transformation of the modified chromosomes into a detailed, feasible service organization model and the evaluation of the service plan quality. These three aspects allow the initial description of uncertainty and stochastic events within a service organization.

I. Genetic Representation Of A Service Organization

The chromosome consists of a number of genes corresponding to the number of activities in a service project or a multi service portfolio of a company. More precisely, a gene represents an activity and has therefore a unique identification number (ID). By the use of an ID, the loci of genes do not necessarily have to be fixed. Each gene consists of several attributes (Fig. 1):

i. Status Of Activity- The boolean value (true, false) indicates if the activity is an integral element of the service plan. Due to this construct the alternative execution of activities or a sequence of activities can be described.
ii. Starting time ti- The value defines the starting time of the activity i and describes the work on predecessor activities j completed by that time.

iii. Within a predetermined range [0, 2.5] the value zi is randomly selected during the optimization process:
zi < 1; Overlapping of activities i and j,
zi = 1; Starting time of activity i and end time of activity j are identical.
zi > 1 Pause between the end time of activity j and the starting time of activity i.

iv. Actor: The value references to the characteristic of at least one actor who has to execute the activity. Thereby, the assigned actor fulfills the formal qualifications required by the task.

v. Duration of Activity di: The attribute describes the duration of the activity i. The stochastic parameters (variance of estimation errors, random events etc.) and deterministic parameters (effect of complexity of task etc.) are integrated in this value.

vi. Events: The attribute ―events" describes the occurrence and the structure of an iteration. The information about the characteristic is stored in the gene which triggers the event.

vii. Additional Information: The information on the probability of the occurrence of this specific activity characteristic is stored in the gene. This information is used during the optimization process to calculate the probability for the occurrence of the generated service process.

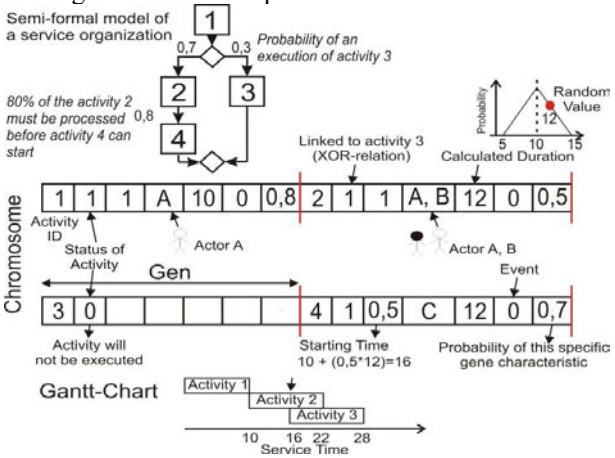


Fig.1. Gene structure

The values of starting time t_i and duration d_i are used to clearly assign the activity into a plan for service provision. Except the gene's ID, which is internally stored by the data structure, the attributes of the gene are subject to crossover and mutation operations and will be changed during evolution. Based on the description of the most important and widely used types of genetic representation by Rothlauf (2006), the developed encoding can be characterized as a real-valued messy representation. The position of genes and the corresponding real value are coded together in the string. Thereby, the encoding is independent of the position of the alleles in the chromosome.

II. Gp Mapping

A chromosome (genotype) represents an individual and is a blueprint for a service organization. Based on the information of varied, unvaried chromosomes and data base entries the genotype-phenotype mapping fg (GP-mapping) is used to generate a feasible, detailed service process for an individual. Information about the characteristics of actors and activities is stored in a central data base and is not varied during optimization. The developed algorithm therefore concerns only for the attributes of the genes of an individual. A mapping results in a description of a detailed service organization which includes information about the sequence of activities, the assignment of actors to tasks as well as further characteristics of the service scenario, i.e., the starting time and the durations of activities, the probability of the occurrence of the service scenario, etc. Uncertainties within a service organization are modeled by stochastic variables and are interpreted during the optimization process by the GA. Within a population the stochastic elements of the service model lead to a variety of heterogeneous individuals. Although uncertainties of service projects have been considered in the approach, due to the novel genotype structure and fg, the mapping of a specific characteristic of a genotype into a phenotype always generates an identical service organization. The steps for the GP mapping process are as follows:

Step 1-A gen (activity i) is randomly chosen from the set of genes of an individual. At the same time the algorithm verifies whether the activity i is already a part of the developed service organization and its predecessors activities have been sufficiently processed. If not, step 1 is continued until an activity i is found.

Step 2-Assignment of actor(s) to the activity i according to the ID of an actor stored in the gene.

Step 3-The exact duration of activity i is listed in the gene. The calculation of the value is done before the GP-mapping by the mutation operator and is based on:

i. Basic effort de(i) of activity i estimated by experts
ii. Qualification q of assigned actor
iii. Random value h, to consider the estimation errors of the requested durations – probability distribution, e.g., Gaussian, right- or left-skewed β-distribution.
The duration d of activity i is calculated: $di = de(i) \bullet q \bullet h$.

Step 4- Calculation of the absolute starting time tj of activity i, based on:

i. Time period related to the end time of predecessor activities tj (variable factor zi, stored in gene)
ii. Starting time and duration of the predecessor(s).
iii. The absolute starting time t_i is calculated as follows:
$t_i = t_j + (z_i \bullet d_j)$. For activities with more than one predecessor: $t_i = max\ \{ t_j + (z_i \bullet d_j) \mid j = 1, \ldots, n\}$

Step 5- In this step it is verified if the assigned actor executes an activity j, j≠i at any particular time of the period pi:

i. Status of the actor (employed, unemployed) during: $p_i = t_i + d_i$

If the actor is unemployed during pi the assignment leads to a change of the status (employed). If an actor is employed, the starting time of the activity i is modified under consideration of the predecessor constraints. For the identification of a valid placement the starting time is iteratively modified: ti = max { tj + 1 + (zi • dj) | j = 1, …, n} until a valid solution is found.

Step 6- Activity Freeze – the parameters of the activity are defined (global starting time, duration etc.) and the activity is placed in the plan.

Step 7- The steps 1 to 6 are repeated until all tasks of a chromosome have been placed for the development of a comprehensive service organization.

III. Evaluation Of Plan Quality

The development of efficient service organizations has to be characterized as a multi-criteria optimization problem. Thereby, the planning objectives are often complementary or conflicting. This can result in a plurality of optimal solutions, but only according to a specific objective or a target combination. Therefore, one optimal plan can only be identified if a prioritizing of objectives is given. Otherwise a pareto front with different optimal solutions must be approximated. In subject to the personal preferences of the planner a solution can be situation-related selected.

The planning problem considered here has a two-dimensional targeting vector (minimization of service budget and service duration). Two-dimensional targeting vectors are partially ordered, i.e. two vectors are incomparable if each of them contains a more advantageous component. A vector dominates another one if at least one component shows a better value and is not concurrently worse in any other component. For example, the Individual No. 1 is dominated with regards to time- and cost-targets by five individuals (Fig. 2). To quantify the performance of the generated individuals, the S-metric by Zitzler et al. (2006) was implemented for the GA. This metric has been deemed one of the fairest and most useful existing methods for the evaluation of multi-criteria solutions (Zitzler et al., 2006). The metric measures the dominated hyper volume. The volume describes the area of a solution space and those vectors which are worse than at least one individual of the set. The S-metric value is to be maximized, which can graphically be regarded as the conquest of the solution space. It is further necessary to distinguish between two sets of individuals non-dominated and dominated vectors. If a population contains dominated individuals, the one with the highest dominance value is removed. The value describes the number of individuals which are dominated by a solution. If a population contains only non-dominated individuals, the one with the lowest s-metric value (hyper volume) is deleted. Thereby, the hyper volume is the sub-space which is only dominated by a single individual (Fig. 2). To minimize the loss of quality the GA selects in a case of exclusive non-dominated individuals the one with the smallest s-metric value. For the dominated individuals a further VS-metric value is calculated. This value describes the smallest distance of a dominated solution to a non-dominated pareto-optimal solution. Thus apart from the non-dominated individuals, the dominated individuals are sequenced in an ascending order.

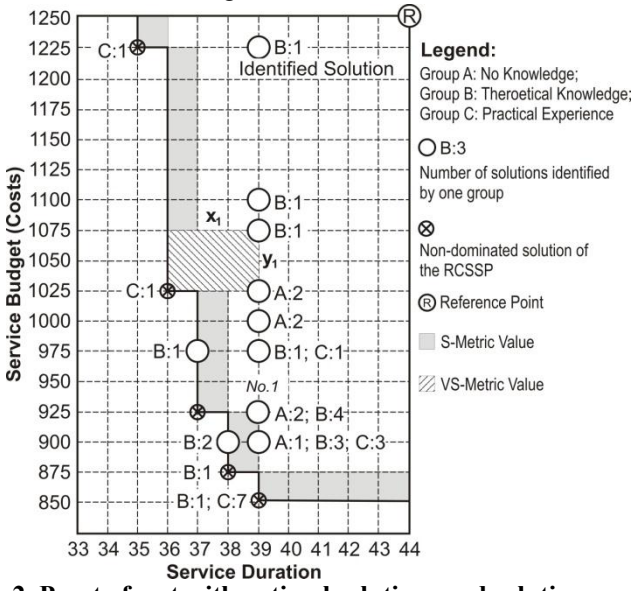


Fig. 2. Pareto front with optimal solutions and solutions identified by participants

IV. Validation Study

An experimental study and a case study in a medium-sized enterprise in the German chemical industry were conducted to verify and validate the developed GA.

The goal of the experimental study was to compare service plans generated by the GA with solutions developed by humans. Therefore, an existing measure of problem solving- and planning competencies (Arling, 2006) was refined. The study was conducted as a web-based planning game and could be executed by multiple participants simultaneously. The advantage of a decentralized execution – due to independence regarding time and place – was that experienced service managers could be recruited for the two-hour participation. On the other hand, the decentralization of the test procedures limited the significance of the test due to a missing control of the framework conditions. A total of 34 participants, 9 female and 25 male between 18 and 56 years of age participated in the study (average 28.03 years, SD=6.85). For the experimental task a typical computer-aided development of a service plan, such as the definition of predecessors and successors for activities or the parameterization of activities (starting time, duration, costs etc) had to be completed by the participants. The task the subjects had to deal with was the development of a project plan for the service —organization of an event" consisting of 34 activities. Each subject received a graphical model indicating the sequence of activities and further information about their characteristics. The web-based experimental environment is similar to MS Project. In parallel to these investigations, the corresponding service model for the GA was developed and

the multi-criteria optimization was carried out. The goal was the simultaneous minimization of service costs and time and therefore the approximation of the pareto front for this specific RCSSP.

The analysis of the planning results of the participants shows a significant deviation of the solutions from the optima, identified by the GA (Fig. 2). For the multiple criteria planning game (cost versus service duration) 5 non-dominated solutions exists. The matrix shows all solutions of the participants in relation to the pareto front. With regard to the information about the classification of the subjects regarding their project planning knowledge/experience the graphical visualization allows a comprehensive evaluation to what extent the subjects were able to identify optimal solutions.

The case study in a medium-sized enterprise involves the service of a facility for the production of polymer. Due to complex interrelations between service projects within the company, only one project and acquired accurate time and cost data was considered. The essential information to build up a service organization model was enquired from the service manager of the enterprise. Finally, the most efficient work organization relating to the fulfillment of predefined planning objectives was compared to the historical data of the enterprise. Identified similarities as well as differences regarding the key project performance indicators (duration, costs, utilization of workers etc.) and the progression of services were discussed with the project manager and project team members. The validation procedure chosen in this study ensures by the separation of data ascertainment (interview and experience-based organizational modeling), the development of the project simulation model and the analysis of the existing project documentation an objective evaluation of the planning quality achieved.

The service project consisted of 62 development tasks, from the identification and analysis of customer specifications of the chemical process, to the commissioning of the installation. Thereby, the service organization was comprehensively modeled through the characterization of internal organizational units of the company, as well as external actors (engineering service providers, maintenance enterprises etc.). The characterization of organizational units included the modeling of qualifications and experiences of the employees, as well as resource descriptions. The tasks have been processed in several projects, so that deviations from the arithmetic mean could be identified by the project manager. Because the task predecessor constraints were deterministic and stochastic, and there were many degrees of freedom in assigning actors to tasks, the considered service process in this particular company was a good candidate for evaluating the performance of the GA. The identification of the global optimum implies an efficient parameterization of the population size and the mutation rate. Based on previous work of Zhuang and Yassine (2004), the correlation between the arithmetic mean of the fitness value and the population size were investigated. The fitness function was a weighted accumulation function with a strong emphasis on the minimization of project lead-time and cost. The weighting of each objective was based on project manager‘s individual preferences concerning the investigated process engineering project. Fig. 3 shows the progress of the arithmetic mean of the fitness influenced by population size and mutation probability pmc. An individual that passed into the next generation is chosen randomly (probability pmc) and it genes are than accidentally selected for mutation. As the population size (number of individual) increases, the fitness increases in value. The results revealed a clear influence of the mutation probability pmc and the fitness. For low values of pmc fitness values become stable when the population size reaches the number of genes of an individual (number of tasks: 62). For higher values of pmc the continuous modification of a high proportion of partial efficient individuals lead to an oscillation of the fitness. These results are consistent with the findings of Zhuang and Yassine (2004). Therefore, it can be assumed that with an adequate population size, good schemes for the project planning problem can be located within the search space.

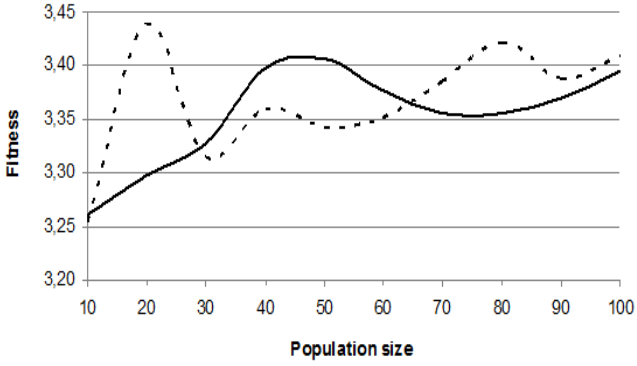


Fig. 3. Arithmetic mean of the fitness value over the chromosomes of all 50 generations for different population size and mutation probability pmc

The validity of the results was further confirmed by analyzing the service lead-time and cost for several generations. This requires the transformation of the genotypes of each generated individual into a detailed service plan. Two scenarios with different actor competences were investigated in detail: (1) The actors involved in the project have a medium competence (value: 1.0). (2) The actors have different values for their competences (0.8 to 1.2). The competence of an actor is negatively correlated with the makespan of a specific task and positively correlated with the cost per time unit. Both correlations are based on estimates of the service managers. The curve progression (Fig. 4) concisely describes the correlation of assigning —competent" actors to tasks to reach a minimization of the service lead-time. In particular, the steep negative slope for the lead-time of scenario 2 until generation 30 indicates the high degrees of freedom regarding the improvement of the service organization. With an increasing number of generations the lead-time converges due to precedence constraints of tasks. The stopping criterion of optimization was set to 100 generation. A further reduction of the makespan of single tasks through the assignment of actors to tasks with a higher competence has no influence and only the service budget increases.

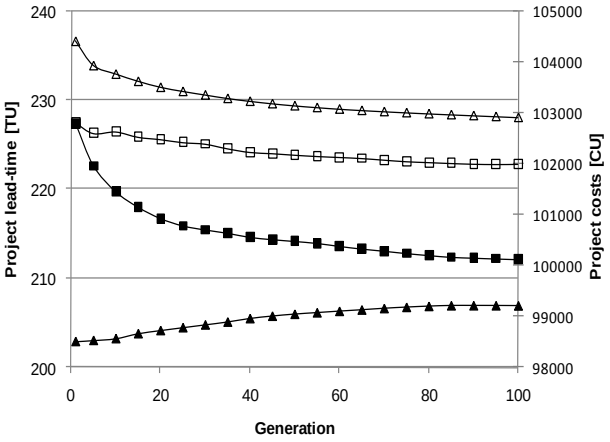


Fig. 4. Arithmetic mean of the service lead-time and costs

V. Conclusion

In the previous sections, a GA was introduced to describe and optimize service productivity. Determining the sequence of tasks with given and indefinite predecessor constraints as well as uncertainty regarding the time-on-task is a NP-hard scheduling problem. The objective was to minimize the service lead-time and costs through the improvement of activity sequences, assignment of actors, and resources under certain constraints (availability; qualification etc.). The GA structure and operations allow the identification and evaluation of optimal service organizations under uncertainty. Furthermore, the transformation of a modified gene resulted in time-grained and valid project plans despite several stochastic influences.

In future papers additional optimization studies of more complex project organizations with a larger number of coupled activities and more organizational units will be presented. A scientific focus is the integration of individual human behavior into the project model. In conclusion, the GA enables manager to develop more efficient service organizations. In the long term, an effect on the service productivity due to a comprehensive identification and evaluation of valid service organizations within a short time will take place.

VI. References

1) Arling, V.S. (2006). Entwicklung und Validierung eines Verfahrens zur Erfassung von Planungskompetenz in der beruflichen Rehabilitation: „Der Tour-Planer―. Ph.D. dissertation, psychology program, RWTH Aachen University, Aachen, Germany.
2) Barnard, J.H. (2006). A Multi-View Framework for Defining the Services Supply Chain Using Object Oriented Methodology. PhD Thesis, University of Central Florida.
3) Blackburn, J.T. (1991). New Product Development: The New Time Wars. In J.T. Blackburn (Ed.), Time-Based Competition: The Next Battleground in Manufacturing. Homewood, IL: Business One Irwin Publishers.
4) Cerasale, M. (2004). Business Solutions on Demand. London: Kogan Pages.
5) Corsten, H., & Stuhlmann, S. (1998). Capacity management in service organizations. Technovation, 18(3), 163-178.
6) Edvardsson, B., & Olsson, J. (1996). Key Concept for New Service Development. The Service Industries Journal, 16(2), 140-164.
7) Eppinger, D., Whitney, D. E., Smith, R.P., & Gebala, D.A. (1994). A model-based method for organizing tasks in product development. Research in Engineering Design, 6(1), 1-13.
8) Groenroos, C. (2000). I Did It My Way. In R.P. Fisk, S.J. Grove, & J. John (Eds.). Service Marketing Self-Portraits: Introspections, Reflections, and Glimpses from Experts. Chicago: American Marketing Association, 71-108.
9) Groenroos, C., & Ojasalo, K. (2004) Service Productivity. Towards a Conceptualization of the Transformation of Inputs into Economic Results in Services. Journal of Business Research, 54(4), 414-423.
10) Hildum, D.W. (1994). Flexibility in a knowledge-based system for solving dynamic resource-constrained scheduling problems. UMass CMPSCI Technical Report 94-77, Department of Computer Science, University of Massachusetts.
11) Johnston, R., & Jones, P. (2003). Service Productivity – Towards Understanding the Relationship between Operational and Customer Productivity. presented in the Workshop Service Productivity, IAO, Stuttgart.
12) Karapidis, A. (2005). Service Management in Production Companies. In G. Zülch, H.S. Jagdev, & P. Stock (Eds.). Integrating Human Aspects in Production Management. Proceedings of the International Conference on Human Aspects in Production Management. New York: Springer: 375-385.
13) Koumpis, A. (2010). Service Science for Socio-Economical and Information Systems Advancement. Hershey: Information Science Publishing.
14) Koza, J.R., Keane, M.A., Streeter, M.J., Mydlowec, W., Yu, J., & Lanza, G. (2003). Genetic Programming IV: routine human-competitive machine intelligence. Newell, MA: Kluwer Academic Publishing.
15) Ojasalo, K. (1999). Conceptualizing productivity in services. Thesis (doctoral). Svenska handelshögskolan, Helsingfors: Swedish School of Economics and Business Administration.
16) Reijers, H.A. (2003). Design and Control of Workflow Processes – Business Process Management for the Service Industry. Berlin: Springer.

17) Rothlauf, F. (2006). Representations for Genetic and Evolutionary Algorithms, 2nd ed. Berlin: Springer.
18) Sink, S. (1985). Productivity Management: Planning, Measurement and Evaluation, Control and Improvement. New York: John Wiley & Sons Inc.
19) Winkelmann, K., & Luczak, H. (2006). Modelling, simulation and prospective analysis of cooperative provision of industrial services using coloured petri nets. International Journal of Simulation, 7(7), 10-26.
20) Yassine, A. A., & Braha, D. (2003). Four Complex Problems in Concurrent Engineering and the Design Structure Matrix Method. Concurrent Engineering Research & Applications, 11(3), 165-176.
21) Yassine, A. A., Meier, C., & Browning, T.R. (2007). Design Process Sequencing With Competent Genetic Algorithms. Transaction of the ASME, 129, 566-585.
22) Zhuang, M., & Yassine, A. A. (2004). Task Scheduling of Parallel Development Projects using Genetic Algorithm. Proceedings of ASME 2004 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Salt Lake City, Utah USA, September 28-October 2. 1-11.
23) Zitzler, E., Deb, K., & Thiele, L. (2000). Comparison of multiobjective evolutionary Algorithms: Empirical Results. Evolutionary Computation, 8(2), 173-195.

Global Journals Guidelines Handbook 2010

www.GlobalJournals.org

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

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

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

- AICCT title will be awarded to the person/institution after approval of Editor-in-Chef and Editorial Board. The title '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.

Scope

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

2. ETHICAL GUIDELINES

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

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

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

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

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

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

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

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

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

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

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

Please mention proper reference and appropriate acknowledgements wherever expected.

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

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

3. SUBMISSION OF MANUSCRIPTS

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

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

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

4. MANUSCRIPT'S CATEGORY

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

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

Research articles: These are handled with small investigation and applications

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

5. STRUCTURE AND FORMAT OF MANUSCRIPT

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

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

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

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

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

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

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

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

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

(h) Brief Acknowledgements.

(i) References in the proper form.

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

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

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

Format

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

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

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

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

Structure

All manuscripts submitted to Global Journals, ought to include:
Title: The title page must carry an instructive title that reflects the content, a running title (less than 45 characters together with spaces), names of the authors and co-authors, and the place(s) wherever the work was carried out. The full postal address in addition with the e-mail address of related author must be given. Up to eleven keywords or very brief phrases have to be given to help data retrieval, mining and indexing.
Abstract, used in Original Papers and Reviews:
Optimizing Abstract for Search Engines
Many researchers searching for information online will use search engines such as Google, Yahoo or similar. By optimizing your paper for search engines, you will amplify the chance of someone finding it. This in turn will make it more likely to be viewed and/or cited in a further work. Global Journals have compiled these guidelines to facilitate you to maximize the web-friendliness of the most public part of your paper.

Key Words
A major linchpin in research work for the writing research paper is the keyword search, which one will employ to find both library and Internet resources.
One must be persistent and creative in using keywords. An effective keyword search requires a strategy and planning a list of possible keywords and phrases to try.

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

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

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

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

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

Acknowledgements: Please make these as concise as possible.

References

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

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

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

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

Tables, Figures and Figure Legends

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

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

Preparation of Electronic Figures for Publication

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

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

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

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

6. AFTER ACCEPTANCE

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

6.1 Proof Corrections

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

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

(Free of charge) from the following website:

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

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

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

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

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

6.3 Author Services

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

6.4 Author Material Archive Policy

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

6.5 Offprint and Extra Copies

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

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

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

Techniques for writing a good quality Computer Science Research Paper:

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

choose that topic. In most of the cases, you may have to conduct the surveys and have to visit several places because this field is related to Computer Science and Information Technology. Also, you may have to do a lot of work to find all rise and falls regarding the various data of that subject. Sometimes, detailed information plays a vital role, instead of short information.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

31. Adding unnecessary information: Do not add unnecessary information, like, I have used MS Excel to draw graph. Do not add irrelevant and inappropriate material. These all will create superfluous. Foreign terminology and phrases are not apropos. One should NEVER take a broad view. Analogy in script is like feathers on a snake. Not at all use a large word when a very small one would be sufficient. Use words properly, regardless of how others use them. Remove quotations. Puns are for kids, not grunt readers. Amplification is a billion times of inferior quality than sarcasm.

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

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

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

Informal Guidelines of Research Paper Writing

Key points to remember:

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

Final Points:

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

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

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

General style:

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

To make a paper clear

· Adhere to recommended page limits

Mistakes to evade

- Insertion a title at the foot of a page with the subsequent text on the next page
- Separating a table/chart or figure - impound each figure/table to a single page

- Submitting a manuscript with pages out of sequence

In every sections of your document

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

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

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

· Align the primary line of each section

· Present your points in sound order

· Use present tense to report well accepted

· Use past tense to describe specific results

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

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

Title Page:

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

Abstract:

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

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

Write your summary when your paper is completed because how can you write the summary of anything which is not yet written? Wealth of terminology is very essential in abstract. Yet, use comprehensive sentences and do not let go readability for briefness. You can maintain it succinct by phrasing sentences so that they provide more than lone rationale. The author can at this moment go straight to shortening the outcome. Sum up the study, with the subsequent elements in any summary. Try to maintain the initial two items to no more than one ruling each.

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

Approach:

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

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

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

Approach:

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

Procedures (Methods and Materials):

This part is supposed to be the easiest to carve if you have good skills. A sound written Procedures segment allows a capable scientist to replacement your results. Present precise information about your supplies. The suppliers and clarity of reagents can be helpful bits of information. Present methods in sequential order but linked methodologies can be grouped as a segment. Be concise when relating the protocols. Attempt for the least amount of information that would permit another capable scientist to spare your outcome but be cautious that vital information is integrated. The use of subheadings is suggested and ought to be synchronized with the results section. When a technique is used that has been well described in another object, mention the specific item describing a way but draw the basic principle while stating the situation. The purpose is to text all particular resources and broad procedures, so that another person may use some or all of the methods in one more study or referee the scientific value of your work. It is not to be a step by step report of the whole thing you did, nor is a methods section a set of orders.
Materials:

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

Methods:

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

Approach:

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

What to keep away from

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

Results:

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

The page length of this segment is set by the sum and types of data to be reported. Carry on to be to the point, by means of statistics and tables, if suitable, to present consequences most efficiently.

You must obviously differentiate material that would usually be incorporated in a study editorial from any unprocessed data or additional appendix matter that would not be available. In fact, such matter should not be submitted at all except requested by the instructor.
Content

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

What to stay away from

- Do not discuss or infer your outcome, report surroundings information, or try to explain anything.
- Not at all, take in raw data or intermediate calculations in a research manuscript.
- Do not present the similar data more than once.
- Manuscript should complement any figures or tables, not duplicate the identical information.
- Never confuse figures with tables - there is a difference.

Approach

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

Figures and tables

- If you put figures and tables at the end of the details, make certain that they are visibly distinguished from any attach appendix materials, such as raw facts
- Despite of position, each figure must be numbered one after the other and complete with subtitle

- In spite of position, each table must be titled, numbered one after the other and complete with heading
- All figure and table must be adequately complete that it could situate on its own, divide from text

Discussion:

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

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

Approach:

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

Administration Rules Listed Before Submitting Your Research Paper to Global Journals

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

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

- The **major constraint** is that you must independently make all content, tables, graphs, and facts that are offered in the paper. You must write each part of the paper wholly on your own. The Peer-reviewers need to identify your own perceptive of the concepts in your own terms. NEVER extract straight from any foundation, and never rephrase someone else's analysis.
- Do not give permission to anyone else to "PROOFREAD" your manuscript.

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

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

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

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

Index

F

G

H

I

L

M

N

O

P

Q

R

S

T

U

V

W

Global Journal of Computer Science and Technology

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

ISSN 9754350