IJACSA
WHERE WISDOM SHARES
International Journal of Advanced Computer Science and Applications
Volume 3 Issue 2
February 2012
ISSN 2156-5570(Online)
ISSN 2158-107X(Print)
SAI
www.ijacsa.thesai.org

IJACSA
WHERE WISDOM SHARES

INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS

THE SCIENCE AND INFORMATION ORGANIZATION

www.thesai.org | info@thesai.org

IJACSA

WHERE WISDOM SHARES

INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS

THE SCIENCE AND INFORMATION ORGANIZATION

www.thesai.org | info@thesai.org

Editorial Preface

From the Desk of Managing Editor…

IJACSA seems to have a cult following and was a humungous success during 2011. We at The Science and Information Organization are pleased to present the February 2012 Issue of IJACSA.

While it took the radio 38 years and the television a short 13 years, it took the World Wide Web only 4 years to reach 50 million users. This shows the richness of the pace at which the computer science moves. As 2012 progresses, we seem to be set for the rapid and intricate ramifications of new technology advancements.

With this issue we wish to reach out to a much larger number with an expectation that more and more researchers get interested in our mission of sharing wisdom. The Organization is committed to introduce to the research audience exactly what they are looking for and that is unique and novel. Guided by this mission, we continuously look for ways to collaborate with other educational institutions worldwide.

Well, as Steve Jobs once said, Innovation has nothing to do with how many R&D dollars you have, it's about the people you have. At IJACSA we believe in spreading the subject knowledge with effectiveness in all classes of audience. Nevertheless, the promise of increased engagement requires that we consider how this might be accomplished, delivering up-to-date and authoritative coverage of advanced computer science and applications.

Throughout our archives, new ideas and technologies have been welcomed, carefully critiqued, and discarded or accepted by qualified reviewers and associate editors. Our efforts to improve the quality of the articles published and expand their reach to the interested audience will continue, and these efforts will require critical minds and careful consideration to assess the quality, relevance, and readability of individual articles.

To summarise, the journal has offered its readership thought provoking theoretical, philosophical, and empirical ideas from some of the finest minds worldwide. We thank all our readers for their continued support and goodwill for IJACSA. We will keep you posted on updates about the new programmes launched in collaboration.

We would like to remind you that the success of our journal depends directly on the number of quality articles submitted for review. Accordingly, we would like to request your participation by submitting quality manuscripts for review and encouraging your colleagues to submit quality manuscripts for review. One of the great benefits we can provide to our prospective authors is the mentoring nature of our review process. IJACSA provides authors with high quality, helpful reviews that are shaped to assist authors in improving their manuscripts.

We regularly conduct surveys and receive extensive feedback which we take very seriously. We beseech valuable suggestions of all our readers for improving our publication.

Thank you for Sharing Wisdom!

Managing Editor
IJACSA
Volume 3 Issue 2, February 2012
ISSN 2156-5570 (Online)
ISSN 2158-107X (Print)

Editorial Board

Reviewer Board Members

CONTENTS

A Comprehensive Analysis of E-government services adoption in Saudi Arabia: Obstacles and Challenges

Mohammed Alshehri, Steve Drew, Osama Alfarraj
School of ICT, Griffith University
Brisbane, Australia

***Abstract*: Often referred as Government to Citizen (G2C) e-government services, many governments around the world are developing and utilizing ICT technologies to provide information and services to their citizens. In Saudi Arabia (KSA) e-government projects have been identified as one of the top government priority areas. However, the adoption of e-government is facing many challenges and barriers including technological, cultural, organizational which must be considered and treated carefully. This paper explores the key factors of user adoption of e-government services through empirical evidence gathered by survey of 460 Saudi citizens including IT department employees from different public sectors. Based on the analysis of data collected the researchers were able to identify some of the important barriers and challenges from these different perspectives. As a result, this study has generated a list of possible recommendations for the public sector and policy-makers to move towards successful adoption of e-government services in Saudi Arabia.**

Keywords— Challenges; E-government services; adoption; Saudi Arabia; Citizens perspective; IT employees.

I. INTRODUCTION

E-government represents a fundamental change in the whole public sector structure, values, culture and the ways of conducting business by utilizing the potential of ICT as a tool in the government agency. The Organization for Economic Co-operation and Development (OECD) [1] defines e-government as "the use of information and communication technologies, and particularly the Internet, as a tool to achieve better government". E-government offers services to those within its authority to transact electronically with the government. These services differ according to users' needs, and this diversity has given rise to the development of different types of e-government. According to Carter & Belanger [2] the relationship of government with recipients of its electronic services can be characterized as: Government to Citizen (G2C), Government to Business (G2B); Government to Employees (G2E); or, Government to Government (G2G).

However, many governments are still in the early stages of implementation and adoption of e-government services. Kingdom of Saudi Arabia (KSA), the biggest country in the Middle East, is in process for a transition to e-government. Today, most of the Saudi government ministries, currently 22, have their own web sites. According to Al Nuaim's study [3] which evaluates the Saudi ministries web sites, it was found that 8 (41%) of 21 ministries did not implement the main features of an e-government web site. In addition, 10 ministries (45.4%) were completely or partially in the first stage (web presence); 3 ministries (13.6%) were in the second stage (one-way interaction); and 6 ministries had no online service at all. This paper reports on research that seeks to answer the question: "What are the challenges and barriers that affect the adoption of e-government services in Saudi Arabia from citizen and government perspectives?" The findings of this study verified some systemic barriers [10] and the extent that are likely to influence the adoption of e-government services. Systemic barriers include: IT infrastructural weakness in government sector, lack of public knowledge about e-government, lack of systems to provide security and privacy of information, and lack of qualified IT and government service expert personnel. Finally, it presents a number of critical strategic priorities for the attention of any e-government project implementation in KSA.

II. RESEARCH METHODOLOGY

In this study a quantitative research method using questionnaires was used to conduct an interpretive study with two sample populations from government IT department employees and members of the general public. Questionnaires are a widely used data collection instrument for recording participant responses to research related questions presented in a predetermined order [4]. Rigorous questionnaire design was undertaken to provide the research with reliable measures that have been validated for this application [5], [6], [7]; also ensuring participants can understand the questions and answer accurately with the most appropriate responses. The questionnaire was used to determine the strength of general citizens' perceptions of obstacles and challenges facing the adoption and diffusion of the e-government services in Saudi Arabia. Also, the same challenges were investigated from IT services providers' perspectives. IT employees in this study represent an element of the government sector and it is very important to explore their expert views and opinions about this issue.

At the beginning of the questionnaire the researcher explained the purpose of the survey and directions for filling out the questionnaire. The questionnaires were distributed to a range of Saudi citizens in public locations such as: shopping centers, internet café and other such locations. For the second sample made up of IT employees such as programmers, software engineers and web designers, participants were approached for participation at their work locations.

The first section of the questionnaire was designed to capture demographic information such as age, occupation, work experience, and educational background. The second section was designed to obtain information on their capabilities using computers and Internet services. The last section contained eleven previously determined barriers [10] to be identified by respondents as either not a barrier (0) or important barrier (1) or very important barrier (2) as shown in Table 1.

It was included to gain better understanding of challenges and obstacles that prevent or influence e-government services acceptance and use in KSA. Sample populations for this study comprises of two group Saudi citizens and IT employees. 400 respondents were ordinary Saudi citizens while 60 were employees from ten (10) governmental public sectors in KSA. Data were then analyzed using SPSS software where selected variables were subjected to exploratory, descriptive and inferential statistical analysis.

III. Data Analysis And Discussion

The following sections highlight the main findings and provide indications as to how the research question might be answered based on the survey results. The first section presents an overview of the results of the online survey questionnaire then the following two sections illustrate the implications for the research question in more detail.

A. *Demographic information*

Table1 and Table 2, following, provide a general overview of the Saudi citizens group and IT employees group in terms of the demographic information, such as gender, age, education level, computer knowledge and internet knowledge. The general population sample might be characterized as being between 21 and 40 years old, mostly degree educated, self identified as having moderate computer knowledge and moderate to good Internet knowledge. By comparison, the major differences in the IT employee group were self identification as having mostly very good computer and Internet knowledge.

B. *Interpretation of research Question: Barriers and challenges to E-Government services adoption*

There are many organizational, technical, social and financial barriers that are facing e-government services adoption and diffusion in KSA. Berge, Muilenburg & Haneghan [8] emphasized that the diffusion of technology into society and citizens is not without obstacles and barriers. However, the government sectors face challenges from Saudi citizens, who expect higher levels of service than from the private sector [9]. The researchers identified eleven barriers to e-government services adoption based empirical research [10] and verified by literature review.

Consequently, participants were asked to evaluate their perceptions of the levels of importance of each barrier by selecting one of the following (0: not a barrier, 1: important barrier, 2: very important barrier).

The barriers that might provide challenges to e-government service adoption are listed in Table 3 and explained in the following sections based on the survey questionnaire groups (Saudi citizens and IT employees).

TABLE I. Demographic information of Saudi citizens

Variable		Frequency	Percent
Gender	Male	295	62.7%
	Female	105	37.3%
Age	Less than 20	85	0.6%
	21-30	125	48.3%
	31-40	162	46.5%
	41-50	19	3.5%
	More than 50	9	1.1%
Education	H.Shool	14	0.5%
	Diploma	122	22.0%
	Bachelor	260	69.3%
	Higher education	6	8.2%
Computer knowledge	Poor	13	3.5%
	Moderate	154	55.7%
	Good	225	39.7%
	Very good	8	1.0%
Internet knowledge	Poor	29	3.5%
	Moderate	108	49.9%
	Good	251	45.1%
	Very good	12	1.5%

TABLE II. Demographic information of IT Staff

Variable		Frequency	Percent
Gender	Male	60	100%
Age	21-30	25	41.7%
	31-40	35	58.3%
Education	Diploma	20	33.3%
	Bachelor	40	66.7%
Computer knowledge	Good	13	21.7%
	Very good	47	78.3%
Internet knowledge	Good	9	15.0%
	Very good	51	85.0%

TABLE IV. BARRIERS OF E-GOVERNMENT SERVICES ADOPTION

No	Barriers	Not a barrier	Important barrier	Very important barrier
1	IT Infrastructural weakness of government public sectors	0	1	2
2	Lack of knowledge and ability to use computers and technology efficiently	0	1	2
3	Lack of knowledge about the e government services	0	1	2
4	Lack of security and privacy of information in government's websites	0	1	2
5	Lack of users' trust and confidence to use e-government services	0	1	2
6	Lack of policy and regulation for e-usage in KSA	0	1	2
7	Lack of partnership and collaboration between the governmental sectors	0	1	2
8	Lack of technical support from government's websites support team	0	1	2
9	Governmental employees resistance to change to e-ways	0	1	2
10	The shortage of financial resources of government sectors	0	1	2
11	The availability and reliability of internet connection	0	1	2

1) Perception of citizens regarding barriers to e-government services

As shown in Table 4 all eleven barriers were selected as either an important or very important barrier and no one of them was selected as not a barrier. In this way the barriers identified both through empirical qualitative investigation [10] and literature review are validated for this sample population.

a) Barriers perceived as being "important"

Inspecting the top three barriers that citizens perceived as being "important" it can be seen that the IT Infrastructural weakness of government public sectors was popular at (53.5%). Moon [11] confirmed that the lack of technical, personnel, and financial capacities are seen as significant obstacles to the development of e-government services in many countries. Moreover, several researchers have mentioned the importance of ICT infrastructure as one of the main barriers in Saudi Arabia [12-15]. Lack of security and privacy of information in government's websites come as the second barrier also with a popularity of (53.5%). West [16] emphasized that e-government will not grow without a sense of privacy and security among citizens regarding their online services and information and services providers need to take care of these issues more seriously. The third barrier is about lack of knowledge and ability to use computers and technology efficiently with a popularity of (51.2%). This is confirmed as an important barrier that is relevant to Saudi Arabia [17].

b) Barriers perceived as being "very important"

From the "very important" angle it is clear that lack of technical support from government's websites support team got the highest percentage at (67.5%) followed by the availability and reliability of Internet connection with (67.0%). Al-shehry [19] highlighted the importance of ensuring positive user experiences in building trust in e-government systems. High quality Internet services will provide system performance to create excellent user experiences. Importantly poor user experiences present risks of citizen rejection of e-government services which may prove difficult to recover [19]. The next most popular barrier in the "very important" category was lack of knowledge about the e-government services at (66.5%). This indicates that a program of promotion is likely to be a significant factor for successful of e- government systems. For any new technology there are many steps to convince and encourage people to accept it and then use it. Research into technology adoption indicates that potential users must perceive that it is useful [28], that it is easy to use [29], and that it provides some relative advantage over the current way of doing things [30]. For citizens to develop these perceptions before extensive experience is gained, programs of promotion and advertising can be key tools to accomplish this task.

TABLE V. ANALYSIS OF E-GOVERNMENT SERVICES OBSTACLES FROM CITIZEN'S PERSPECTIVE

No	Barriers	Important barrier		Very important barrier	
		Frequency	Percent	Frequency	Percent
1	IT Infrastructural weakness of government public sectors	214	53.5%	186	46.5%
2	Lack of knowledge and ability to use computers and technology efficiently	205	51.2%	195	48.8%
3	Lack of knowledge about the e-government services	274	33.5%	544	66.5%
4	Lack of security and privacy of information in government's websites	214	53.5%	186	46.5%
5	Lack of users' trust and confidence to use e-government services	197	49.3%	203	50.7%
6	Lack of policy and regulation for e-usage in KSA	198	49.5%	202	50.5%
7	Lack of partnership and collaboration between the governmental sectors	166	41.5%	234	58.5%
8	Lack of technical support from government's websites support team	130	32.5%	270	67.5%
9	Governmental employees resistance to change to e-ways	170	42.5%	230	57.5%
10	The shortage of financial resources of government sectors	198	49.5%	202	50.5%
11	The availability and reliability of internet connection	132	33.0%	268	67.0 %

2) Perception of IT employees toward obstacles of e-government services

Table 5 summarizes the barriers from the analysis of IT employees' perspectives. Again, the three most popularly identified barriers from the two perception levels will be illustrated in the following subsections.

a) Barriers perceived as being "important"

Of the barriers that IT employees perceived as "important" it is clear that lack of knowledge and ability to use computers and technology efficiently ranked most highly at (68.3%). The ability of citizens to effectively use computers and the Internet is a critical success factor in e-government projects, and the lack of such skills may lead to marginalization or even social exclusion [20]. Lack of security and privacy of information in government's websites presented as the second most popular barrier at this level of importance at (65.0%). Ndou [21] considered privacy and confidentiality as critical obstacles toward the realization of e-government in developing countries. It was revealed that citizens studied were deeply concerned with the privacy of their information and confidentiality of the personal data they are providing as part of obtaining government services. Thus, it was pointed out that privacy and confidentiality are high priorities when establishing and maintaining web sites in order to ensure the secure collection of data. On the other hand governments should provide a secure authenticated access to their online services in order to maintain citizen trust in use of e-government services. Practically, media campaigns and promotion through awareness seminars and brochures about safe Internet use and security principles is an important supporting strategy in citizen acceptance of any e-government system. E-government systems are revolutionary in many developing countries around the world and support its effective use it requires appropriate policies and regulatory framework. Such laws and regulation should be "e-aware" to cover all e-applications such as e-payments, e-mail usage, copyright rules, e-crimes, e-business, e-commerce and others [22]. In the KSA case, the Saudi government has issued many government regulations and laws such as e-transaction law, information criminal law, shift to electronic methods decision and many other laws. These laws and regulations are playing an important function in promoting effective communication between citizens, business and government to accelerate the adoption of e-government service on all levels. However, the existence of these laws and regulations is but one step in e-government adoption process and needs information about them to be published in the community domain to facilitate and provide confidence in their use.

b) Barriers perceived as being "very important"

From the barriers that IT employees viewed as "very important" Table 5 reveals that the lack of technical support from government's websites support teams got the highest percentage at (93.3%). Thus a fast and accurate technical support service is an essential part of an effective and efficient e-government system. Citizens may understandably be easily deterred by technical failures, so it is very important to have a professional team to detect and respond to technical issues and to help users as soon as possible. Citizens require high-quality technical support, in order to learn how to use the e-services and become familiar with them. Hoffman [23] defined technical support as ''knowledge people assisting the users of computer hardware and software products'', which can include help desks, information centre support, online support, telephone response systems, e-mail response systems and other facilities. Technical support is one of the significant factors in the acceptance and use of technology [24, 25], and accordingly in the adoption of e-applications such as e-government services. The second most popularly perceived "very important" barrier was lack of knowledge about e-government services at (81.7%). As raised earlier, effective promotion is likely to be one of the most significant factors influencing successful citizen adoption of e-government systems [26]. For any new technology there are many steps to convince and encourage people to adopt and use it so promotion and advertising are tools central to accomplishing this task. The survey results indicate that the lack of programs to promote the e-government services benefits and advantages may be a significant barrier to the adoption of e-government in Saudi society. The third most popular "very important" barrier was IT infrastructural weakness in government public sectors at (80.0%). The ICT infrastructure is an essential part of successful e-government implementation and diffusion [17]. It enables government agencies to cooperate, interact and share work in an effective and professional fashion. Development of ICT infrastructure, both in government and private domains, needs to be sensitively handled in the Saudi context and accompanied by an effective, staged roll-out strategy [27].

TABLE VI. ANALYSIS OF E-GOVERNMENT SERVICES OBSTACLES FROM IT EMPLOYEE'S PERSPECTIVE

No	Barriers	Important barrier		Very important barrier	
		Frequency	Percent	Frequency	Percent
1	IT Infrastructural weakness of government public sectors	12	20.0%	48	80.0%
2	Lack of knowledge and ability to use computers and technology efficiently	41	68.3%	19	31.7%
3	Lack of knowledge about the e-government services	11	18.3%	49	81.7%
4	Lack of security and privacy of information in government's websites	39	65.0%	21	35.0%
5	Lack of users' trust and confidence to use e-government services	26	43.3%	34	56.7%
6	Lack of policy and regulation for e-usage in KSA	38	63.3%	22	36.7%
7	Lack of partnership and collaboration between the governmental sectors	36	60.0%	24	40.0%
8	Lack of technical support from government's websites support team	4	6.7%	56	93.3%
9	Governmental employees resistance to change to e-ways	33	55.0%	27	45.0%
10	The shortage of financial resources of government sectors	18	30.0%	42	70.0%
11	The availability and reliability of internet connection	15	25.0%	45	75.0%

3) Comparison of perceptions of barriers

The aim of this section is to compare between the viewpoints of Saudi citizens and IT employees about barriers to adoption of e-government services. It is clear from the previous sections that there are many perceived barriers that are common to both. Firstly, both sample populations nominated lack of technical support for government websites as a “very important” barrier and ranked that as the most important barrier in the list. This agreement between both sample populations indicated that it is a critical barrier to be resolved with a high level of priority. Next, both groups agreed that lack of knowledge about the e-government services was considered the second or third most important barrier in the “very important barriers” list. Finally, there was a distinction in the next most popular “very important” barrier to e-government adoption and this reflects the individual perspectives of the sample populations. For the IT employees within the government IT infrastructural weakness was seen as a significant barrier to them being able to provide reliable and effective services. From the perspective of the ordinary citizenry having access to reliable and effective Internet services has impact on their ability to access and make effective use of the available services. Table 6 presents the common barriers and distinct barriers with their relative popularity.

TABLE VII. COMMON AND DISTINCT BARRIERS BETWEEN THE TWO GROUPS

Barrier	Rank	Percent	
		Citizens	***IT employees***
Lack of technical support from government’s websites support	1	67.7%	93.3%
Lack of knowledge about the e-government services	2	66.5%	81.7%
The availability and reliability of internet connection	3	67.2%	75.0%
IT Infrastructural weakness of government public sectors	3	46.5%	80.0%

IV. IMPLICATIONS FROM RESEARCH

Based on the research outcomes, the highest priority strategies to be implemented to help successful adoption of e-government services in Saudi Arabia are:

- Instantiation of reliable and responsive technical support systems that cover all Saudi government organizations and agencies. Results indicate that technical support is the foundation stone of a successful adoption program and should be integral to e-services rollout. Any weakness in technical support systems may present a barrier to all e-government implementation stages.

- Instantiation of comprehensive information and training programs that raise citizen awareness and knowledge of e-government services as they become accessible in each region. An associated advertising campaign focusing on each emerging e-government system and service with its benefits and advantages.

- Acceleration of the rollout of high performance network and Internet infrastructure to government agencies and service providers. Acceleration of the rollout of high performance Internet infrastructure starting with the areas of highest population density.

These can then be supported by:

- Development and adoption of a set of standards and processes for the design, development, and maintenance of all government websites. Instantiation of government Web developer training programs to ensure appropriately skilled professionals needed to build and maintain websites that provide a high level of transaction security and ease of use for all e-services.

- Instantiation of interagency agreements that require and promote high levels of collaboration, cooperation, and service consistency between all government agencies and with the ongoing Yesser national e-government development project [13].

- Instantiation of standards and protocols to ensure and enforce highest levels of information security and privacy across all e-government agencies and services. Instantiation of standardized trusted payment systems and gateways which are highly secure, highly accessible, and simple to use for all online transactions.

V. CONCLUSION

Citizen’s adoption of e-government services is an important goal for many governmental service providers, however the success of this adoption process is not easy and requires a thorough understanding of the needs of citizens and system requirements.This paper focuses on the barriers of e-government services adoption in Saudi Arabia.

The result of this work based on the response of Saudi citizens and IT employees in public sectors to the research question. Based on the data collected through a questionnaire survey the researchers identified and uncovered many important factors which affect directly the adoption process. It is clear that there are many important factors which common between the two groups and that need to be addressed in quick and professional manner. As result of this study, a brief set of recommendations has been made in order to help the public sectors and government originations to improve the e-government services outcome and to achieve the aspirations of Saudi citizens and their satisfaction with electronic services.

REFERENCES

[1] OECD, OECD E-Government Flagship Report “The E-Government Imperative,” Public Management Committee, Paris: OECD, 2003

[2] Carter, L., & Belanger, F. (2005). The utilization of e-government services: citizen trust, innovation and acceptance factors. Information Systems Journal, 15(1), 5–25.

[3] Al-Nuaim, H. (2011) An Evaluation Framework for Saudi E-Government, Journal of e-Government Studies and Best Practices, Vol. 2011 (2011)

[4] Neuman, W. (2006) Social research methods: qualitative and quantitative approaches, 6th edn, Allyn and Bacon, Boston, MA, USA.

[5] Al- Shareef, K. (2003). The challenges that face e-government in Saudi Arabia. Work report, King Saud University, Riyadh, Kingdom of Saudi Arabia.

[6] Al-Smmary, S., (2005). E-readiness in Saudi Arabia. Work report, King Fahd University of Petroleum and Minerals, Damamm, Kingdom of Saudi Arabia.

[7] Al-Solbi, H. & Al-Harbi,S. (2008). An exploratory study of factors determining e-government success in Saudi Arabia. Communications of the IBIMA, Volume 4, 2008

[8] Berge, Z. L., Muilenburg, L. Y., & Haneghan, J. V. (2002). Barriers to distance education and training. The Quarterly Review of Distance Education, 3(4), 409-418.

[9] Chavez, Regina. (2003). the utilization of the mazmanian and sabatier model as a tool for implementation of e-government for Fresno county, California. Doctoral Dissertation. California State University.

[10] Alshehri, M. & Drew, S. (2010). Challenges of e-Government Services Adoption from an e-Ready Citizen Perspective. World Academy of Science, Engineering and Technology, 66, 1053–1059.

[11] Moon, M. (2002).The evolution of e-government among municipalities: rhetoric or reality. Public Administration Review 62 (4): 424-33.

[12] Al-Ghaith, W.A., Sanzogni, L. and Sandhu, K. (2010), "Factors influencing the adoption and usage of online services in Saudi Arabia", Electronic Journal on Information Systems in Developing Countries, Vol. 40 No. 1, pp. 1-32,

[13] Al-shehry, A., Rogerson, S., Fairweather, N. & Prior, M,. (2006). Egovernment adoption: an empirical study. The Information Communication & Society (ICS) 10th Anniversary Symposium, York, 20–22 September 2006.

[14] Al-Sobhi, F., Weerakkody, V. & Kamal, M. M. 2010. „An exploratory study on the role of intermediaries in delivering public services in Madinah City: Case of Saudi Arabia‟. Transforming Government: People, Process and Policy, 4(1):14-36.

[15] A. Al-Solbi and S. Al-Harbi, (2008) An exploratory study of factors determining e-Government success in Saudi Arabia, Computer Centre, Air Academy, Riyadh, Saudi Arabia.

[16] West, D.M. (2001). State and Federal E-government in United States, Available at:http://www.insidepolitics.org/egovt01us.html. Accessed on 25/08/2011

[17] Al-Zumaia, Abdulrahman. S. (2001). Attitudes and Perceptions Regarding Internet-Based Electronic Data Interchange in Public Organization in Saudi Arabia. Doctoral Dissertation. University of Northern Iowa

[18] Feng. L. (2003). Implementing E-government Strategy is Scotland: Current Situation and Emerging Issues. Journal of Electronic Commerce in Organizations 1(2), 44-65

[19] Al-shehry, A., (2008). Transformation towards E-government in The Kingdom of Saudi Arabia: Technological and Organisational Perspectives . Doctoral Dissertation De Montfort University, UK.

[20] United Nations Division for Public Economics and Public Administration, (2001). Benchmarking Egovernment: A Global Perspective - Assessing the Progress of the UN Member States., from http://pti.nw.dc.us/links/docs/ASPA_UN_egov_survey1.pdf Accessed on 19/12/2011.

[21] Ndou , V. (2004). E-government for developing countries: opportunities and challenges. The Electronic Journal on Information Systems in Developing Countries 18, 1, 1-24

[22] Ralph, W. (1991). Help! The art of computer technical support. California: Peachpit Press.

[23] Hofmann, D. W. (2002). Internet-based distance learning in higher education. Tech Directions, 62(1), 28–32.

[24] Williams, P. (2002). The learning Web: the development, implementation and evaluation of Internet-based undergraduate materials for the teaching of key skills. Active Learning in Higher Education, 3(1), 40–53.

[25] Geetika, P. (2007). Strategic Marketing of E-Government for Technology Adoption Facilitation Available online at: www.csisigegov.org/critical_pdf/6_51-60.pdf . Accessed on 22/05/2010.

[26] Jaegar, P. and Thompson, K. (2003), E-Government around the world: Lessons, Challenges and Future Directions, Government Information Quarterly 20, 329-394

[27] Dada, D. (2006) The Failure of E-Government in Developing Countries, retrieved 2006, available at http://www.lse.ac.uk/collections/informationSystems/iSChannel/Dada_2006b.pdf. Accessed on 20/11/2011

[28] Davis, F. (1989) Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology, MIS Quarterly, 13(3), 319-340.

[29] Venkatesh, V. (2000) Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model, Information Systems Research, 11(4), 342-365.

[30] Moore, G, C. and Benbasat, I. (1991) Development of an instrument to measure the perceptions of adopting an information technology innovation, Information Systems Research, 2, 192-222.

AUTHORS PROFILE

Dr Steve Drew is a Senior Lecturer in the School of ICT at Griffith University. He heads a group of researchers looking at different aspects of information systems e-services acceptance and adoption.

Mohammed Alshehri is a Ph.D candidate at ICT School at Griffith University at Australia. He received his MSC in Computer and Communication Engineering from QUT in Brisbane in 2007.

Osama Alfarraj is a Ph.D candidate at ICT School at Griffith University at Australia. He received his MSC in IT from Griffith University in Brisbane in 2008.

A Flexible Approach to Modelling Adaptive Course Sequencing based on Graphs implemented using XLink

Rachid ELOUAHBI
Department of mathematics and computer Science, Moulay Ismail University, Faculty of Sciences Meknes, Morocco

Driss BOUZIDI
Department of mathematics and computer Science, Hassan II Ain Chock University,Faculty of Sciences, Casablanca, Morocco

Noreddine ABGHOUR
Department of mathematics and computer Science, Hassan II Ain Chock University,Faculty of Sciences, Casablanca, Morocco

Mohammed Adil NASSIR
Department of mathematics and computer Science, Moulay Ismail University, Faculty of Sciences Meknes,Morocco

***Abstract*—A major challenge in developing systems of distance learning is the ability to adapt learning to individual users. This adaptation requires a flexible scheme for sequencing the material to teach diverse learners. This is where we intend to contribute to model the personalized learning paths to be followed by the learner to achieve his/her determined educational objective. Our modelling approach of sequencing is based on the pedagogical graph which is called SMARTGraph. This graph allows expressing the totality of the pedagogic constraints under which the learner is submitted in order to achieve his/her pedagogic objective. SMARTGraph is a graph in which the nodes are the learning units and the arcs are the pedagogic constraints between learning units.**

We shall see how it is possible to organize the learning units and the learning paths to answer the expectations within the framework of individual courses according to the learner profile or within the framework of group courses. To implement our approach we exploit the strength of XLink (XML Linking Language) to define the sequencing graph.

Keywords-* *e-learning; pedagogical graph; adaptability; profile; learning units; XML; XLINK; XSL.

I. INTRODUCTION AND STATE OF THE ART

In the e-Learning context, the sequencing means the scheduling, planning or also the organization of the elements constituting the content to be taught. In order to carry out the pedagogic sequencing, it is necessary to know the learners profile and the objectives of the training. From a given detailed representation of the contents, the sequencing authorizes a dynamic planning of the course elements that will be the subject of a pedagogic activity. This representation describes in a structured and hierarchical way the scheduling of the objectives to be achieved in order to meet the terminal goals as fixed by the course.

It is very important to design the course sequencing, since it leaves the traditional relation between teacher and learners [6]. It is necessary for us to exploit various contexts so that the trainings are built through various interactions. In the context of distance learning, there is no more exposure of knowledge, and declaratory knowledge assimilation by learners, who will thereafter be implemented in evaluation situations, and translated into competency. It is rather necessary to offer the learner an interactive framework which will not only guide his/her course in the sequences of teaching, but also will justify its the spirit of the initiative.

The sequencing is present in several areas of research. In the context of Adaptive Educational Hypermedia systems (AEH) applied to distance learning, sequencing of content should be carefully designed so that the learner does not get lost in hyperspace [16] [17]. The content to be presented to learners must be selected and adapted in terms of presentation and navigation [13] [15]. The sequencing of content has also a vital part in the Intelligent Tutoring Systems (ITS) [12] developed under the hypothesis that computer systems can model human learning in selecting the best scheduling learning strategy for each learner [14]. The objective of the majority of ITS is to adapt their training offer in a dynamic way, according to pedagogical rules that depend also on the reactions of the learner [8] [5].

In parallel with this research on sequencing, considerable effort has been devoted to developing standards to enable systems-based learning on the web to find, share, reuse, and export content in a standardized way such as the IMS [10], the ADL [2], the AICC [4] , the ISO/IEC JTC1 SC36 [11] , the IEEE/LTSC [7] and the W3C [25]. The function of the sequence was included in the specifications of the standard IMS Simple Sequencing [9] which is taken up by SCORM Sequencing and Navigation [3]. This feature provides the learner with the sequencing of content in order to guide him/her in the learning space.

The three approaches STI, AEH and IMS SS are superficially similar. They are intended to provide not only an appropriate content to achieve the learning objective efficiently, but also have fundamental important differences. Indeed, the ITS and AEH use a model student and a model of concepts in learning to decide what content and what navigation structure to display, and how to present this content. The learner model can be initiated by previous knowledge and dynamically updated according to the behaviour of the learner. The model concepts described how information is structured and linked together using the concept of relationship. However, IMS SS specification describes three data models: Sequencing Definition Model that stores the rules of progression in the activities, the Tracking Model describes the results of the learner's interactions related to an objective, and Activity State Model which records the status of current activity. The learning model is implicit in the tracking model. It contains only the educational progress and does not support the features and preferences of the learner.

The aim of IMS SS is to ensure that the learner can perform all the activities that an author considers important, while avoiding unnecessary ones. The objective is to focus on the strategies of the author. On the other hand, the objective of ITS and AEH is to help the learner to be familiar with the complex space of relations between domain concepts to achieve the goals he has chosen. Thus, these systems are learner-centered.

The main objective of this paper is to describe our approach for modeling course sequencing based on graphs which is called a SMARTGraph pedagogic graph. This will help the author to well structure his/her course and pedagogic relations between the various pedagogic units which compose it and help the learner to be oriented during the browsing in the course.

Proposed approach for modeling the course sequencing

In our approach the teaching course/path was designed to allow each learner to express his/her capacities as well as possible, and to build assets in agreement with his/her pedagogical objectives. The course is composed of Learning Units (LU) which are interrelated by pedagogic relations [19]. It is up to the pedagogic sequencing to define these relations. It will have to determine which LU will be presented to the learner and when it will take place.

It must allow not only the conditional branch from one learning resource to other learning resources, according to whether the learner has carried out a certain stages or obtained a sufficient note, but also the fact that this last is permitted to subscribe in a LU. The pedagogic sequencing is the result of the application of composition techniques which can be, for example, the browsing of the tree structure in a linear way from a LU to another (Fig. 1). A more complex pedagogic sequencing can be based on the achievement of certain LUs, as prerequisite LUs, on the learner preferences or on the evaluation results.

SMARTGraph is a pedagogic graph which is included in this perspective of the pedagogic sequencing modelling. In SMARTGraph the nodes represent the learning units and the arcs (links) are the pedagogical constraints between pedagogical sequences. In order to express the pedagogic constraints, we use the prerequisites of formalism. A prerequisite of a LU is the set of the knowledge necessary to follow this LU in order to achieve a defined objective.

The prerequisite of LUs can be components of the same LU (it is the case of the chapters of the same course) or external LUs belonging to other LUs (it is the case of a chapter of another course or even of another cursus). Using this formalism, the notation used to express that the LUi is a prerequisite of the LUj is as follows: Pr(LUj)=LUi. The logical operators that allow us to express pedagogical constraints between different units of the course are:

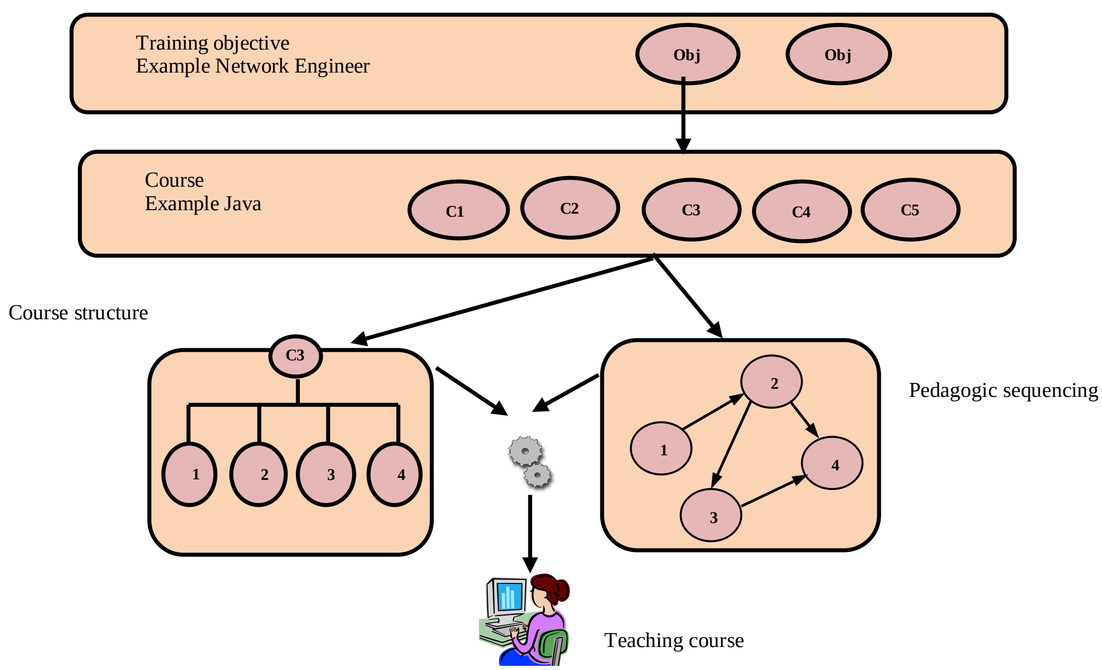


Figure 1. Proposed approach for course sequencing

A. No prerequisite

$Pr(LU_1)$ = NULL. There is no LU required to follow LU_1

Figure 2. No prerequiste

B. Simple prerequisite

- Without condition: Pr(LU2) = LU1. The learner can follow LU2 if he/she has followed LU1. It is equivalent to an unconditioned sequencing between two LUs of a graph.

Figure 3. Simple prerequiste

- With condition: Pr(LU2) = LU1 : c. The learner can follow LU2 if he/she has followed LU1 and satisfied the condition c. It is equivalent to a conditioned sequencing between two LUs. A condition c is boolean expression where the operators are =, >, >=, <, <= and the operands are strings such as test score or number of attempts. To combine between multiple expressions we use the logical operators that are conjunction, disjunction and negation: &, | and !.

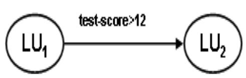


Figure 4. Prerequiste with condition

C. Conjunction prerequisite

- Without order: Pr(LU3)= LU1& LU2. The learner can follow LU3 if he/she has successfully followed LU1 and LU2.

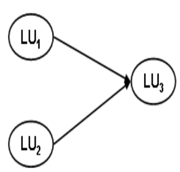


Figure 5. Conjonction prerequiste without order

- With order : Pr(LU3)= LU1&& LU2. The learner can follow LU3 if he/she has successfully followed LU1 and LU2 in the order.

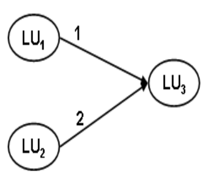


Figure 6. Conjonction prerequiste witht order

D. Disjunction prerequisite

- Simple: Pr(LU3)= LU1| LU2. Choice between many LUs. The learner can follow either LU1 or LU2 to consider that the group LU1, LU2 has been done.

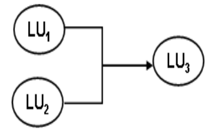


Figure 7. Disjonction prerequiste

- Choice of n LU among m without any constraint on the followed order as prerequisite: Pr(LU4)={LU1,LU2,LU3}2. 2 LUs of the set LU1, LU2, LU3, are sufficient to allow access to LU4.

- Choice of n LU among m with constraint on the followed order as prerequisite:

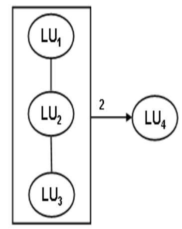


Figure 8: Choice 2 LUs among 3 without order

- Pr(LU5)=[LU1,LU2,LU3,LU4]2. Two LUs of the group are sufficient to consider that the group has been done, but the order must be respected.

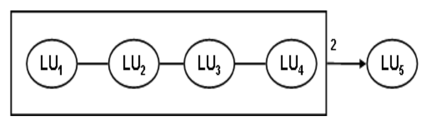


Figure 9. Choice 2 LUs among 4 witht order

E. Exclusive prerequisite

- Pr(LU3)= LU1^ LU2. The learner can follow either LU1 xor LU2 to consider that the group LU1, LU2 has been done.

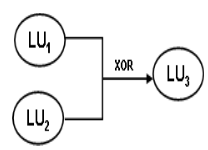


Figure 10. Exclusive prerequiste

In a complex expression of prerequisites we use parentheses to determine the precise order of evaluation of the expression. For example, on the expression Pr(LU4)= LU1& LU2| LU3, the learner can follow LU4 if he/she simply has followed LU3 or (LU1 and LU2). But if we add () like this Pr(LU4)= LU1 & (LU2|LU3), the learner can follow LU4 if he/she has successfully completed LU2 or LU3 and the LU1.

Then, once the course is generated, the learner starts the first part, and at the end of this part, the system must take into account the learner's interactions during this learning unit. The learner will then have access to a specific learning unit if necessary with a modified profile. The transition from an educational unit to another is then made according to an educational approach.

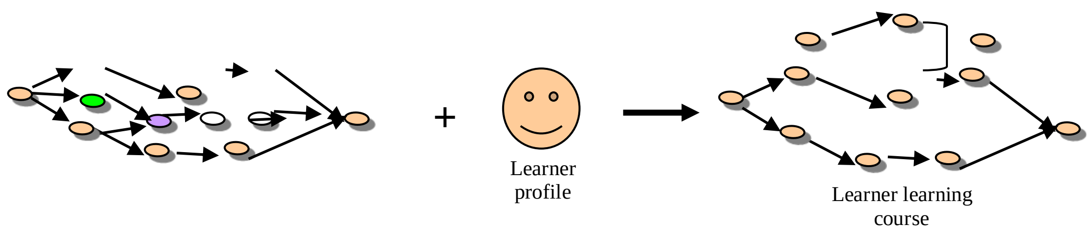


Figure 12. Personalization

The total graph is deduced starting from the prerequisites expressions because each definition of prerequisites makes it possible to deduce a portion from the pedagogical graph. This made, the fusion of the whole of under graph makes it possible to reconstitute the general graph.

The following scheme shows an example of SMARTGraph:

1) $Pr(LU_1)$= NULL
2) $Pr(LU_2)$= LU_1 | LU_2 : c_1
3) $Pr(LU_3)$= LU_2 & LU_6
4) $Pr(LU_4)$= LU_3
5) $Pr(LU_5)$= LU_3
6) $Pr(LU_6)$= LU_1
7) $Pr(LU_7)$= LU_6 : c_2
8) $Pr(LU_8)$= LU_7 & (LU_4 | LU_5)

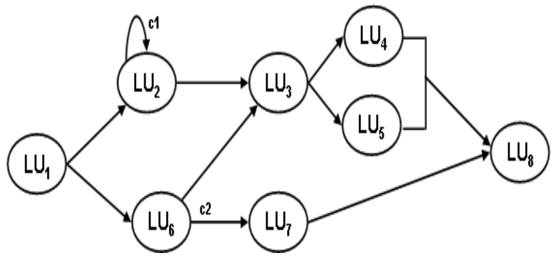


Figure 11. Example of SMARTGraph

While being based on the formalisms defined above, two approaches are offered to represent the pedagogical Graph: The first consists of representing the graph from the final or atomic elements of the hierarchical structure of the course. The second consists of breaking up the total graph into a hierarchy of under graph gathered by stage or level [18]. In this second case of figure, the course will be organized so that the passage of node to another is conditioned by the validation of under graph corresponding to this node. Contrary to the first representation, this one offers much more visibility and comprehension, because it makes it possible to preserve semantics on the regrouping of the course item.

II. Graph Adaptability According To Learner Profile

Thus the generation of the specific course sequencing consists of a simple extraction of a sub-graph of the general graph of the course (Fig. 12). The elements of the learners' profile will act here as a filter which lets pass only the pedagogical sequence with which they are in conformity [1].

The learner's profile is a key element in the implementation of adaptability in our approach. The profile is represented by an amount of information which characterizes the learner in the learning process. It consists in particular of the capacities, language, objectives of training and psychological factors of learning. Thus for the same learning objectives, each learner will have an adapted course, corresponding to his language, his learning speed or his capacities [20].

This requirement can be reached only by one good knowledge (modelling) of the learner. The purpose of the modelling of the learner is to establish the characteristics of learning depending in particular on his program, on his answers, on his preferences and his behavior to adapt the training to each individual; i.e. to define a profile for learning. This profile of learning can be subdivided into three components:

- The Intellectual profile which reflects the image of the learner on the educational level. The objective of this sub-model is to appreciate at every moment and with its right value, the state of knowledge of the learner.
- The criteria of localization which aim to facilitate the communication between the learner and the actors of the e-learning system and to allow a better comprehension of the course. We can quote such criterion of localization; the language, cultural origins and other criteria of localization.

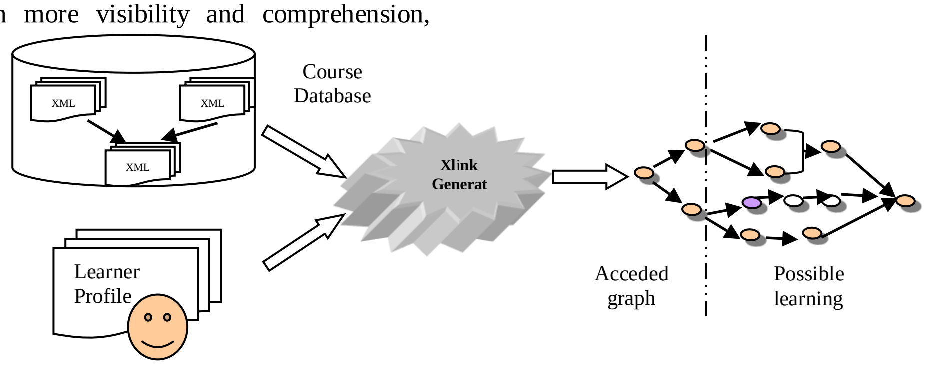


Figure 13: Implementation of SMARTGraph

- A preference of the learner who represents these personal requirements so that the learner can evolve in an environment is familiar and convivial for him. For example, each learner will be able to visualize the course in the aesthetics of personalized colors.

III. Implementation Of Sequencing

Our objective is to provide an adaptive sequencing across pedagogical sequence a need for a good definition and structuring of the course documents is obvious for a comprehension of this structure and this definition by the system in order to extract some information targeted to provide course sequencing adapted to each learner's profile (Fig. 13).

In the search for a solution to our needs, our choice was made on XML [23] to describe the structure of the course, XLink (XML Linking Language) [24] to define the sequencing graph.

The adaptation of the course sequencing according to the user's profile is made possible using XSLT [22]. It will first of all be necessary to adopt a structure defining the relationship between the course and the profile.

The generation process consists of generating the already accomplished course by the learner in the first part of the graph and all the other possible courses in the second part. The continuation of the learner's course will be done according to his/her personal choices, and especially to his profile evolution.

A. Technical choice: use of Xlink

To satisfy our objectives for the implementation of the course sequencing, namely:

- The ability to know the paths followed by a learner during his progress in the course;
- Support for multi-directional links corresponding to an expression of combined pre-requisites (expression formulated from other expressions of prerequisites);
- To be able to manage the documents involved in the graph as LUs may be sources, target connection, or even both (i.e, a link with a series of locations and connections between them);
- And in the end to reach our main objective to have a sequencing document between learning units entirely separate from its contents.

Use of HTML hyperlinks has proven insufficient in our context. However, the XML linking technologies meets our needs perfectly. Indeed, XLink provides mechanisms rendering even more flexible hypermedia documents by:

- Extended Links: although richer than HTML links because they are multidirectional between multiple documents. An extended link is a directed graph in which the locations correspond to vertices and links between nodes correspond to arcs.
- Out-of-Line Links: such a link does not appear in the document or documents for which it constitutes a link. This allows storing a series of definitions of links in a separate document which we called "basic links". Thus, we can define a document that can contain out-of-line links that are totally separated from documents for which the links are active. This involves an intelligent separation between the document content and its sequencing. So we can change the order of the sequence of parts of a document without changing its contents; it suffices to modify its sequencing in the "basic links" document.

The notion of out-of-line extended links perfectly meets our afore-mentioned needs. We model the operators used in expressions of the prerequisites, linking the different learning sequences described in the pedagogical progression graph SMARTGraph; in arcs oriented more specifically in out-of-line extended links while taking account of the conditions of passage defined by the author.

In order to structure the elaboration of the implementation of SMARTGraph we proceed by decomposition in two stages: generation and adaptation.

B. Generation process

As shown in Fig. 14, starting from the list of the expressions of the prerequisites stored in the database, we generate generic XLINK corresponding to the total pedagogical graph of the Fig. 11 via an XSLT transformation.

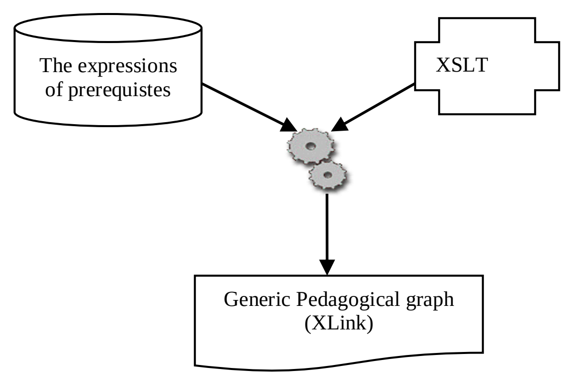


Figure 14. Pedagogical graph generation process

C. Adaptation process

As shown in Fig. 16, knowing that each learner has a profile which is brought to change constantly during all the learning process, it is unimaginable to envisage all possible XSLT transformations being able to be applied to the pedagogical graph.

This led us to choose the solution of a generic XSLT file containing a certain number of parameters which have a direct relationship with the profile (Fig. 17).

Thus, the parameter setting of this generic XSLT by the profile item will dynamically give a specific XSLT according to this profile (Fig. 18). The transformation of the generic XSLT into specific XSLT will be done by using XML parser such as DOM (Document Object Model) [21].

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<?xml version="1.0" encoding="utf-8"?>
<graph  xmlns:xlink="http://www.w3.org/2000/xlink"
        xlink:type="extended">
<nodes>
<node ID="LU1" xlink:type="locator"/>
<node ID="LU2" xlink:type="locator"/>
<node ID="LU3" xlink:type="locator"/>
<node ID="LU4" xlink:type="locator"/>
<node ID="LU5" xlink:type="locator"/>
<node ID="LU6" xlink:type="locator"/>
<node ID="LU7" xlink:type="locator"/>
<node ID="LU8" xlink:type="locator"/>
</nodes>
<links>
<operator type="or">
<arc xlink:type="arc" xlink:from="LU1" xlink:to="LU2" condition="null"/>
<arc xlink:type="arc" xlink:from="LU2" xlink:to="LU2" condition="c1"/>
</operator>
<operator type="simple">
<arc xlink:type="arc" xlink:from="LU1" xlink:to="LU6" condition="null"/>
</operator>
<operator type="simple">
<arc xlink:type="arc" xlink:from="LU2" xlink:to="LU3" condition="null"/>
</operator>
<operator type="and">
<arc xlink:type="arc" xlink:from="LU2" xlink:to="LU3" condition="null"/>
<arc xlink:type="arc" xlink:from="LU6" xlink:to="LU3" condition="null"/>
</operator>
<operator type="simple_c">
<arc xlink:type="arc" xlink:from="LU6" xlink:to="LU7" condition="c2"/>
</operator>
<operator type="simple">
<arc xlink:type="arc" xlink:from="LU3" xlink:to="LU4" condition="null"/>
</operator>
<operator type="simple">
<arc xlink:type="arc" xlink:from="LU3" xlink:to="LU5" condition="null"/>
</operator>
<operator type="and">
<arc xlink:type="arc" xlink:from="LU7" xlink:to="LU8" condition="null"/>
<operator type="or">
<arc xlink:type="arc" xlink:from="LU4" xlink:to="LU8" condition="null"/>
<arc xlink:type="arc" xlink:from="LU5" xlink:to="LU8"/>
</operator>
</operator>
</links>
</graph>
```

Figure 15 : Generic XLink file

TABLE I. PROFILE

Profile	
Pedagogical Objective	Engineer
Accomplished units list	LU1,LU2, LU3, LU6
Language	French

```
<xsl:transform
        xmlns:xsl="http://www.w3.org/1999/XSL/Transfor
        m"   version="1.0">
<xsl:variable name="Language" select=""/>
<xsl:variable name="List_LUs" select=""/>
<xsl:variable name="Obj_Pedag" select=""/>
<xsl:template match="/">
<xsl:for-each select="graph">
<xsl:choose>
---------------------------------
</xsl:choose>
</xsl:for-each>
</xsl:template>
</xsl:transform>
```

Figure 17 : Generic XSLT file

Applied to the values of the following profile:

Give the specific XSLT above:

```
<xsl:transform  xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
        version="1.0">
<xsl:transform version="1.0">
<xsl:variable name="Language" select="french"/>
<xsl:variable name="List_LUs" select=" LU1,LU2, LU3, LU6"/>
<xsl:variable name="Obj_Pedag" select=" Engineer ""/>
<xsl:template match="/">
<xsl:for-each select="graph">
<xsl:choose>
---------------------------------
</xsl:choose>
</xsl:for-each>
</xsl:template>
</xsl:transform>
```

Figure 18 : Specific XSLT file

Thus, to generate the SMARTGraph specific to a profile, it is necessary to apply the generic pedagogical graph XLink XSL transformation specific to this profile.

Applied to the specific XSLT of Fig. 18 gives the specific pedagogical Graph below:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<?xml version="1.0" encoding="utf-8"?>
<graph  xmlns:xlink="http://www.w3.org/2000/xlink"
        xlink:type="extended">
<nodes>
<node ID="LU4" xlink:type="locator"/>
<node ID="LU5" xlink:type="locator"/>
<node ID="LU7" xlink:type="locator"/>
<node ID="LU8" xlink:type="locator"/>
    </nodes>
<links>
<operator type="simple">
<arc xlink:type="arc" xlink:from="LU3" xlink:to="LU4"
        condition="null"/>
    </operator>
<operator type="simple">
<arc xlink:type="arc" xlink:from="LU3" xlink:to="LU5"
        condition="null"/>
    </operator>
<operator type="and">
<arc xlink:type="arc" xlink:from="LU7" xlink:to="LU8"
        condition="null"/>
<operator type="or">
<arc xlink:type="arc" xlink:from="LU4" xlink:to="LU8"
        condition="null"/>
<arc xlink:type="arc" xlink:from="LU5" xlink:to="LU8"/>
    </operator>
    </operator>
    </links>
  </graph>
```

Figure 19 : Specific XLink file

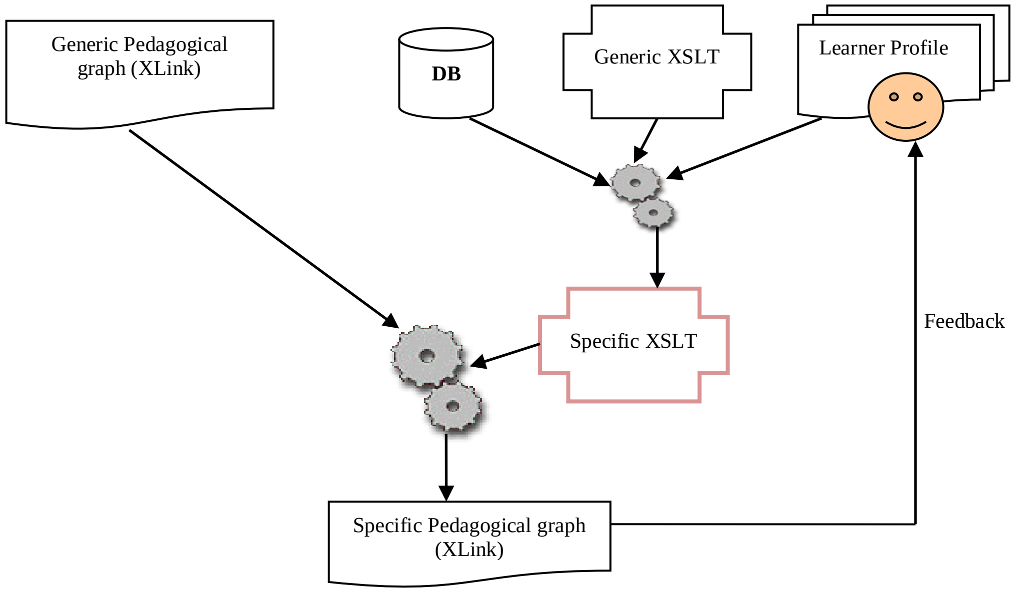


Figure 15. Adaptation process

IV. CONCLUSION

In this paper we have proposed a modelling approach of course sequencing based on graphs. These graphs pedagogical in nature are a projection of the learning paths of the learner defined by the content designer as pedagogical constraints. In our approach the relationship of prerequisites is the translation of these pedagogical constraints. The combination of the prerequisites determines the educational graph SMARTGraph, hence modeling the possible sequencing within the educational content. The SMARTGraph nodes are the learning units and the arcs are the pedagogic constraints between these units.

SMARTGraph allows the real-time modelling of the choice and the succession of the courses, or the parts of a course that a learner operates during his/her training. This modelling consists of presenting the relations between the different parts of a course, or a cursus, by means of algebraic operators. These operators are used, in this sense, as constructors of the sequencing of the parts of a course for a learner's profile at a given time. In fact , this sequencing allows or does not specify access instructions according to the operators result.

The adaptability proposed by our model is not restricted only to the course content, but also to the sequencing of its content. This provides our approach with a high level of adaptability and flexibility.

Based on standard technologies such as XML, XLink and XSLT in our implementation system, we have proved the relevance of the concepts which we have presented.

ACKNOWLEDGEMENT

The authors wish to give their sincere thanks to Professor Ali BOUTOULOUT the director of the research group TSI (Théorie des Systèmes et Informatique) that allows us to conduct the researches.

REFERENCES

[1] A. Benkiran and R. Ajhoun “an Adaptive and cooperative telelearning system: SMART-Learning”, International Journal on E-Learning April-June 2002, Vol.1, No.2 pp: 60-72.

[2] ADL Shareable Courseware Object Reference Model: SCORM Version 1.3, 2004. Available from http://www.adlnet.org/

[3] ADL, “SCORM Sequencing and Navigation (SN) Version 1.3”, http://www.adlnet.gov, 2004.

[4] AICC, “Aviation industry CBT committees”, http://www.aicc.org, 1998.

[5] C. J. Butz, S. Hua, and R. B. Maguire, “A web-based intelligent tutoring system for computer programming”, In WI '04: Proc. Of the 2004 IEEE/WIC/ACM Int.Conf. on Web Intelligence. IEEE Computer Society, 2004.

[6] D. Bouzidi , R. Elouahabi and N. Abghour “SMARTNotes: Semantic annotation system for a collaborative learning”, International Journal of Computer Science Issues, Vol. 8, Issue 6, No 1, November 2011.

[7] IEEE/LTSC, “IEEE Learning Technology Standers Committee”, http://ieeeltsc.org/, 1997.

[8] G. Weber and P. Brusilovsky (2001). “ELM-ART: An adaptive versatile system for Web-based instruction”, International Journal of Artificial Intelligence in Education, 12 (4), Special Issue on Adaptive and Intelligent Web-based Educational Systems, 351-384.

[9] IMS, “IMS Simple Sequencing Information and Behavior Model, Version 1.0 Final Specification”, http://www.imsglobal.org/simplesequencing, 2003.

[10] IMS, “IMS Global Learning Consortium”, http://www.imsglobal.org, 1997.

[11] ISO/IEC JTC1 SC36, “ISO/IEC JTC1 SC36 Standards for: information technology for Learning, Education, and Training”, http://jtc1sc36.org/index.html, 1999.

[12] P. Brusilovsky, and C. Peylo (2003), “Adaptive and Intelligent Web-based Educational Systems”, International Journal of Artificial Intelligence in Education, 13 156-169.

[13] P. Brusilovsky, “Adaptive navigation support in educational hypermedia: The role of student knowledge level and the case for meta-adaptation”, British Journal of Educational Technology, 34(4):487–497, 2003.

[14] P. Brusilovsky and J. Vassileva (2003), "Course sequencing techniques for large-scale Web-based education", International Journal of Continuing Engineering Education and Life-long Learning, Vol.13.

[15] P. Brusilowsky, "Adaptive Hypermedia" Journal of User Modeling and User-Adapted Interaction, Vol. 11, pp. 87-110, 2001.

[16] P. De Bra, G.J. Houben and H. Wu, "AHAM: A Dexter-based Reference Model for Adaptive Hypermedia", Proceedings of the 10th ACM Conference on Hypertext and Hypermedia, ACM Press, 1999, pp. 147–156.

[17] P. De Bra , D. Smits and N. Stash, "Creating and delivering adaptive courses with AHA!" In EC-TEL 2006, pp. 21-33

[18] R. Elouahbi A.Nassir and S.Benhlima, "Design and implementation of adaptive sequencing based on graphs", 11th International Educational Technology Conference, Turkey, 2011, p 233-238.

[19] R. Elouahbi , "Modelisation et mise en œuvre du sequencement dans le systeme de tele-enseignement SMART-Learning via le graphe pedagogique SMARTGraph", Thèse de Doctorat, Spécialité Informatique, Ecole Mohammadia d'Ingénieurs (EMI), Université Mohammed V-Agdal, Rabat, 2005.

[20] S. B. M. Jonas, R. Elouahbi and A. Benkiran, "Installation of the Adaptability in Course Sequencing via the SMARTGraph pedagogic Graph " proceedings of ECEL 2004 the The 3rd European Conference on e-Learning, Paris, France, November 2004, pp. 597-604.

[21] World Wide Web Consortium, Document Object Model DOM Level 2, http://www.w3.org/TR/DOM-Level-2-Core/.

[22] World Wide Web Consortium, eXtensible Stylesheet Language, XSL Specification version 1.0, W3C Recommendation, October 15, 2001. http://www.w3.org/TR/xsl/

[23] World Wide Web Consortium, Extensible Markup Language, XML Specification –version 1.0 (Second Edition), W3C Recommendation, October 6, 2000. http://www.w3.org/TR/2000/REC-xml-20001006

[24] World Wide Web Consortium, XML Linking Language (XLink) Version 1.1, 06 May 2010. http://www.w3.org/TR/2010/REC-xlink11-20100506/.

[25] W3C, "World Wide Web Consortium", http://www.w3.org, 1994.

AUTHORS PROFILE

Rachid ELOUAHBI received his doctorate degree in Computer Science from Mohammadia School of Engineering Morocco in 2005. He is currently an Assistant Professor at the University of Moulay Ismail, Meknès. His research focused on the area of adaptive learning systems, course sequencing and learning technologies. He works on the project Graduate's Insertion and Assessment as tools for Moroccan Higher Education Governance and Management in partnership with the European Union.

Driss BOUZIDI is an Assistance Professor in Computer Science at Hassan II Ain Chock University, Faculty of Sciences Casablanca, Morocco. He received a Ph.D. degree in computer engineering from Mohammed V University in 2004 on "Collaboration in hypermedia courses of e-learning system SMART-Learning". Dr. Bouzidi's research interests are in Distributed multimedia Applications, Collaboration systems, and elearning. He was vice-chair of the international conference NGNS'09 (Next Generation Network and Services) and the treasurer of the two research associations e-NGN and APRIMT.

Noreddine ABGHOUR is currently an Assistant Professor at the Faculty of Sciences Ain Chok, Casablanca, Department of Mathematics and Computer Science. He is also a PhD in Computer Science, area of distributed applications, from The Polytechnic Institute of Toulouse. In addition to teaching, his research interests are in the field of security and collaborative systems.

Mohammed Adil NASSIR is currently a PhD student at the Faculty of Sciences Meknes, Department of Mathematics and Computer Science. He is a Technical Engineer in Computer Systems and a specialist in local area networks from St-Petersburg Technical University - Russia. He works on the project of Graduate's Insertion and Assessment as tools for Moroccan Higher Education Governance and Management in partnership with the European Union.

A Global Convergence Algorithm for the Supply Chain Network Equilibrium Model

Lei Wang
Feixian School
Linyi University
Feixian, Shandong, P.R. China.

***Abstract*—In this paper, we first present an auxiliary problem method for solving the generalized variational inequalities problem on the supply chain network equilibrium model (GVIP), then its global convergence is also established under milder conditions.**

Keywords- Supply chain management; network equilibrium model; generalized variational inequalities; algorithm; globally convergent.

I. Introduction

The topics of supply chain model, analysis, computation, and management are of great interests, both from practical and research perspectives. Research in this area is interdisciplinary by nature since it involves manufacturing, transportation, logistics, and retailing/marketing. A lot of literatures have paid much attention to this area. The interested readers may consult the recent survey papers by Stadtler and Kilger, Poirier, Giannessi and Maugeri (Refs.[1-4]) and references therein. For example, Nagurney et al. ([5]) developed a variational inequality based supply chain network equilibrium model consisting of three tiers of decision-makers in the network. They established some governing equilibrium conditions based on the optimality conditions of the decision-makers along with the market equilibrium conditions in 2002. In 2004, Dong et al.([6]) establish the finite-dimensional variational inequality formulation for a supply chain network model consisting of manufacturers and retailers in which the demands associated with the retail outlets are random.

In 2005, Nagurney et al. ([7]) establish the finite-dimensional variational inequality formulation for a supply chain network model in which both physical and electronic transactions are allowed and in which supply side risks as well as demand side risk are included in the formulation. The model consists of three tiers of decision-makers: the manufacturers, the distributors, and the retailers, with the demands associated with the retail outlets being random. In recent years, variational inequalities have been extended in many directions via innovative techniques to study a wide class of problem arising in pure and applied sciences. A useful and important generalization is called the general variational inequality problem (GVIP). This problem was introduced first by Noor([8]) in 1988, it and related problems have been studied by many researchers(See Refs.[4-15])

In this paper, we consider the solution method for GVIP on supply chain network equilibrium model of finding x^* in R^n such that

$$\langle F(x)-F(x^*), G(x^*)\rangle \geq 0, \forall F(x)\in\Omega, x\in R^n, \quad (1)$$

where R^n be a real Euclidean space, whose inner product and the Euclidean 2-norm are denoted by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$, respectively. Let Ω be a nonempty closed convex set in R^n. Given nonlinear mappings $G: R^n \to R^m$, $F(x) = Mx + p$, $M \in R^{m\times n}$, $p \in R^m$, and F is onto Ω. The solution set of the GVIP is denoted by X^* which is always assumed to be nonempty.

In recent years, many methods have been proposed to solve the GVIP, among various of efficient methods for solving GVIP, projection method is the simplest one([10-15]). Solodov and Svaiter ([11]), He ([10]), Wang ([15]) applied a new class of projection-contraction (PC) methods to monotone GVIP. Different from the algorithm above, we proposed a new method for solving the GVIP under milder conditions, a strictly convex quadratic programming only need to be solved at each iteration.

II. Algorithm and Convergence

In this section, we give a new-type method to solve the GVIP under milder conditions. We first need the definition of projection operator and some relate properties ([16]).

For nonempty closed convex set $\Omega\subset R^n$ and any vector $x\in R^n$, the orthogonal projection of x onto Ω, i.e., $argmin\{\|y-x\| \mid y\in\Omega\}$, is denoted by $P_\Omega(x)$. For (1), $\rho > 0$ is given a constant,

$$e(x) := F(x) - P_\Omega[F(x) - \rho G(x)]$$

is called projection-type residual function, and let $r(x) := \|e(x)\|$. The following conclusion provides the relationship between the solution set of (1) and that of projection-type residual function which is due to Noor([8]).

Lemma 2.1 x is a solution of (1) if and only if $r(x)=0$.

To establish the following algorithm, we also need the following conclusion in [17].

Lemma 2.2 Suppose that the non-homogeneous linear equation system $Hy=b$ is consistent. Then $y=H^{+}b$ is the solution with the minimum 2-norm, where H^{+} is the pesudo-inverse of H.

In this following, we give a description of our proposed algorithm.

Algorithm 2.1

Step1. Take $\varepsilon>0$, parameters $0<\rho<2\mu$, and initial point $x^{0}\in R^{n}$. Set $k\overset{\vee}{=}0$;

Step2. Compute $F(x^{k+1})$ by solving the following problem

$$
\begin{aligned}
\min \quad & (F(x)-F(x^{k}))^{\bullet}\,(F(x)-F(x^{k})) \\
& +2\rho(F(x)-F(x^{k}))^{\bullet}\,G(x^{k}) \qquad (2)\\
\text{s.t.} \quad & F(x)\in\Omega;
\end{aligned}
$$

Step3. If $\|F(x^{k+1})-F(x^{k})\|\le\varepsilon$ go to Step 4, otherwise, go to Step 2 with $k\overset{\vee}{=}k+1$;

Step4. Let $x^{k+1}=M^{+}(F(x^{k+1})-p)$, where M^{+} is the pseudo inverse of M. Stop.

By the definition of projection operator, we can easily get that $F(x^{k+1})$ is a solution of problem (2) if and only if

$$F(x^{k+1})=P_{\Omega}(F(x^{k})-\rho G(x^{k})). \qquad (3)$$

To establish the global convergence of Algorithm 2.1, we will state the following some well-known definitions ([18]).

Definition 2.1 The mapping $G:R^{n}\to R^{m}$ is said to be strongly pseudo monotone with respect to F if there is constant $\mu>0$ such that

$$
\begin{gathered}
\langle G(y),F(x)-F(y)\rangle\ge 0\\
\Rightarrow\langle G(x),F(x)-F(y)\rangle\ge\mu\|G(x)-G(y)\|^{2}, \qquad (4)\\
\forall x,y\in R^{n}.
\end{gathered}
$$

Obviously, If the mapping G is strongly monotone with respect to F ([18]), then The mapping G is strongly pseudo monotone with respect to F, but the converse is not true in general. For example , $G(x)=2-x, F(x)=x$, the mapping G is strongly pseudo monotone with respect to F with constant 1 in interval $[0,1]$, but it is not strongly monotone and even not monotone.

Theorem 2.1 Suppose that the mapping G is strongly pseudo monotone with respect to F , then the sequence $\{x^{k}\}$ globally converges to a solution of the GVIP.

Proof: Since $\rho>0$, (2) has an unique solution, denoted by $F(x^{k+1})$. Obviously, if $F(x^{k+1})=F(x^{k})$, i.e.,

$$r(x^{k})=F(x^{k})-P_{\Omega}(F(x^{k})-\rho G(x^{k}))=0,$$

using Lemma 2.1, then x^{k} is a solution of GVIP.

In the following analysis, we assume that Algorithm 2.1 generates an infinite sequence. Suppose that $F(x^{k+1})\ne F(x^{k})$ holds, and the objective function of (2) is denoted by $H(x)$ with $x^{k}=x^{*}(x^{*}\in X^{*})$. We would prove that the sequence $\{H(x^{k})\}$ is monotone decreasing. To this end, we set

$$
\begin{aligned}
&\Psi(k,k+1)=H(x^{k})-H(x^{k+1})\\
&=(F(x^{k})-F(x^{*}))^{\bullet}\,(F(x^{k})-F(x^{*}))\\
&+2\rho\langle G(x^{*}),F(x^{k})-F(x^{*})\rangle\\
&-(F(x^{k+1})-F(x^{*}))^{\bullet}\,(F(x^{k+1})-F(x^{*}))\\
&-2\rho\langle G(x^{*}),F(x^{k+1})-F(x^{*})\rangle\\
&=(F(x^{k}))^{\bullet}\,F(x^{k})-(F(x^{*}))^{\bullet}\,F(x^{*})\\
&-2\langle F(x^{*}),F(x^{k})-F(x^{*})\rangle\\
&-(F(x^{k+1}))^{\bullet}\,F(x^{k+1})+(F(x^{*}))^{\bullet}\,F(x^{*})\\
&+2\langle F(x^{*}),F(x^{k+1})-F(x^{*})\rangle\\
&+2\rho\langle G(x^{*}),F(x^{k})-F(x^{k+1})\rangle\\
&=(F(x^{k}))^{\bullet}\,F(x^{k})-(F(x^{k+1}))^{\bullet}\,F(x^{k+1})\\
&+2\langle F(x^{*}),F(x^{k+1})-F(x^{k})\rangle\\
&+2\rho\langle G(x^{*}),F(x^{k})-F(x^{k+1})\rangle\\
&=(F(x^{k}))^{\bullet}\,F(x^{k})-(F(x^{k+1}))^{\bullet}\,F(x^{k+1})\\
&-2\langle F(x^{k+1}),F(x^{k})-F(x^{k+1})\rangle\\
&+2\langle F(x^{k+1})-F(x^{*}),F(x^{k})-F(x^{k+1})\rangle\\
&+2\rho\langle G(x^{*}),F(x^{k})-F(x^{k+1})\rangle\\
&=(F(x^{k})-F(x^{k+1}))^{\bullet}\,(F(x^{k})-F(x^{k+1}))\\
&+2\langle F(x^{k+1})-F(x^{k}),F(x^{*})-F(x^{k+1})\rangle\\
&+2\rho\langle G(x^{*}),F(x^{k})-F(x^{k+1})\rangle
\end{aligned}
$$

$$\geq (F(x^k)-F(x^{k+1}))^{\bullet}\,(F(x^k)-F(x^{k+1}))$$
$$-2\rho\langle G(x^k),F(x^*)-F(x^{k+1})\rangle$$
$$+2\rho\langle G(x^*),F(x^k)-F(x^{k+1})\rangle$$
$$=(F(x^k)-F(x^{k+1}))^{\bullet}\,(F(x^k)-F(x^{k+1}))$$
$$+2\rho\langle G(x^k),F(x^k)-F(x^*)\rangle$$
$$-2\rho\langle G(x^k),F(x^k)-F(x^{k+1})\rangle$$
$$+2\rho\langle G(x^*),F(x^k)-F(x^{k+1})\rangle$$
$$\geq (F(x^k)-F(x^{k+1}))^{\bullet}\,(F(x^k)-F(x^{k+1}))$$
$$+2\rho\mu\|\,G(x^k)-G(x^*)\|^2$$
$$-2\rho\langle G(x^k)-G(x^*),F(x^k)-F(x^{k+1})\rangle$$
$$\geq\|\,F(x^k)-F(x^{k+1})\|^2+2\rho\mu\|\,G(x^k)-G(x^*)\|^2$$
$$-2\rho\|\,G(x^k)-G(x^*)\|\;\;\|\,F(x^k)-F(x^{k+1})\|$$
$$\geq\|\,F(x^k)-F(x^{k+1})\|^2+2\rho\mu\|\,G(x^k)-G(x^*)\|^2$$
$$-2\rho\mu\|\,G(x^k)-G(x^*)\|^2-\frac{\rho}{2\mu}\|\,F(x^k)-F(x^{k+1})\|^2$$
$$\geq\|\,F(x^k)-F(x^{k+1})\|^2-\frac{\rho}{2\mu}\|\,F(x^k)-F(x^{k+1})\|^2.$$
$$=(1-\frac{\rho}{2\mu})\|\,F(x^k)-F(x^{k+1})\|^2.$$

Since (2) can be equivalently reformulated as the following variational inequalities

$$\begin{aligned}&\langle 2(F(x^{k+1})-F(x^k)),F(x)-F(x^{k+1})\rangle\\&+2\rho\langle G(x^k),F(x)-F(x^{k+1})\rangle\geq 0,\\&\forall F(x)\in\Omega,\end{aligned}\tag{5}$$

let $F(x)=F(x^*)$ in (5), we have that the first inequality holds. Let $y=x^*\;\;(x^*\in X^*)$ in (4), and

$$\langle G(x^*),F(x^k)-F(x^*)\rangle\geq 0,$$

Combining this with Definition 2.1, we get

$$\langle G(x^k),F(x^k)-F(x^*)\rangle\geq\mu\|\,G(x^k)-G(x^*)\|^2,\tag{6}$$

by (6), we have that the second inequality holds. The third inequality is based on Cauchy-Schwarz inequality. By $0<\rho<2\mu$, we have $\Psi(k,k+1)>0$, the nonnegative sequence $\{H(x^k)\}$ is strictly decreasing.

Combining the definition of $H(x)$, we have

$$\begin{aligned}H(x^k)&=\|\,F(x^k)-F(x^*)\|^2\\&\quad+G(x^*)^{\bullet}\,(F(x^k)-F(x^*))\\&\geq\|\,F(x^k)-F(x^*)\|^2\geq 0.\end{aligned}\tag{7}$$

So $\{H(x^k)\}$ converges, and we get $\Psi(k,k+1)\to 0$ as $k\to\infty$, and

$$\lim_{k\to\infty}\|\,F(x^k)-F(x^{k+1})\|=0.\tag{8}$$

Moreover, $\{H(x^k)\}$ is bounded since it is convergent, and so is $\{F(x^k)\}$ according to (7). Let $\{F(x^{k_i})\}$ be a subsequence of $\{F(x^k)\}$ and converges toward $F(\bar{x})$, by (5), we obtain

$$\langle G(\bar{x}),F(x)-F(\bar{x})\rangle\geq 0,\forall F(x)\in\Omega,\tag{9}$$

by (9), we have $\bar{x}$ is a solution of (1). The $\bar{x}$ can be used as x^* to define the function $H(x)$: denoted $\bar{H}(x)$, we have

$$\begin{aligned}&\|\,F(x)-F(\bar{x}\|^2\leq\bar{H}(x)\\&\leq\|\,F(x)-F(\bar{x}\|^2\\&+\|\,G(\bar{x})\|\;\;\|\,F(x)-F(\bar{x})\|.\end{aligned}\tag{10}$$

and we know that $\{\bar{H}(x^k)\}$ also converges, Substituting $F(x)$ in (10) with $F(x^{k_i})$, we get $\bar{H}(x^{k_i})\to 0$ as $i\to\infty$. Thus, we have $\{\bar{H}(x^k)\}\to 0$ as $k\to\infty$. By using (7) again, we know that the sequence $\{F(x^k)\}$ converges globally toward $F(\bar{x})$. Since F is onto Ω, we have

$$\begin{aligned}&\|\,x^k-\bar{x}\|\\&=\|\,M^+(F(x^{k+1})-p)-M^+(F(\bar{x})-p)\|\\&\leq\|\,M^+\|\;\;\|\,F(x^k)-F(\bar{x})\|\to 0(k\to\infty).\end{aligned}$$

Then the desired result is followed.

III. Conclusion and Prospect

In this paper, we present an auxiliary problem method for solving GVIP, and we also have showed that method has a global convergence, and we needn't the conditions which the mapping G is continuously differentiable and monotone on R^n, and the conditions which the mapping G is Lipschitz continuous is also moved, it is a new result for GVIP. It is uncertain whether the algorithm is global and $R-$ linear convergence, this is a topic for further research.

Acknowledgment

The authors wish to give their sincere thanks to the editor and the anonymous referees for their valuable suggestions and helpful comments which improved the presentation of the paper.

This work was supported by the Natural Science Foundation of China (Grant No. 11171180, 11101303), and Specialized Research Fund for the doctoral Program of Chinese Higher Education(20113705110002), and Shandong Provincial Natural Science Foundation (ZR2010AL005, ZR2011FL017), and the projects for reformation of Chinese universities logistics teaching and research (JZW2012065).

REFERENCES

[1] H. Stadtler and C. Kilger, Supply chain management and advanced planning, Springer-Verlag, Berlin, Germany, 2002.

[2] C. C. Poirier, Supply chain optimization: Building a total business network, Berrett-Kochler Publishers, San Francisco, California, 1996.

[3] C. C. Poirier, Advanced supply chain management: How to build a sustained competitive advantage, Berrett-Kochler Publishers, San Francisco, California, 1999.

[4] F. Giannessi and A. Maugeri, Variational inequalities and network equilibrium problems, Plenum Press, New York, NY, 1955.

[5] A. Nagurney, J. Dong, and D. Zhang, "A supply chain network equilibrium model", Transportation Research E, 38, pp. 281-303, 2002.

[6] J. Dong, D. Zhang, A. Nagurney, "A supply chain network equilibrium model with random demands", European Journal of Operational Research, 156, pp.194-212, 2004.

[7] A. Nagurney, J. Cruz, J. Dong and D. Zhang, "Supply chain networks, electronic commerce, and supply side and demand side risk", European Journal of Operational Research, 164, pp.120-142, 2005.

[8] M. A. Noor, "General variational inequalities", Applied Mathematics Letters, Vol.1, pp.119-121, 1988.

[9] F. Facchinei and J.~S. Pang, Finite-Dimensional variational inequality and complementarity problems. New York: Springer, 2003.

[10] B.S. He, "A class of projection and contraction method for monotone variational inequalities", Appl. Math. Optim., 35, pp.69-76, 1997.

[11] M.V. Solodov and B.F. Svaiter, "A new projection method for variational inequality problems", SIAM J. Control Optim., 37, pp.765-776, 1999.

[12] Y.J. Wang, N.H. Xiu and C.Y. Wang , "A new version of extragradient method for variational inequality problems", Comput. Math. Appl., 42, pp.969-979, 2001.

[13] Y.J. Wang, N.H. Xiu and C.Y. Wang, "Unified framework of extragradient-type methods for pseudomonotone variational inequalities", J. Optim. Theroy Appl., 111, pp. 641-656, 2001.

[14] Y.J. Wang, N.H. Xiu and J.Z. Zhang , "Modified extragradient methods for varitional inequalities and verification of solution existence", J. Optim. Theroy Appl., 119,pp.167-183, 2003.

[15] Y.J. Wang, "A new projection and contraction method for variational inequalities", Pure Math. and Appl., 13(4), pp.483-493, 2002.

[16] E.H. Zarantonello, Projections on convex sets in Hilbert Space and spectral theory, contributions to nonlinear functional analysis, New York: Academic Press, 1971.

[17] Roger A. Horn, Charles R. Johnson, Topics in matrix analysis, Cambridge. Univ. Press, 1991.

[18] N. El Farouq, "Convergent algorithm based on progressive regularization for solving pseudo-monotone variational inequalities", J. Optimization Theory and Applications, 120, pp.455-485, 2004.

A New Approach of Trust Relationship Measurement Based on Graph Theory

Ping He
Department of Information
Liaoning police Academy
Dalian 116036, China

***Abstract*—The certainty trust relationship of network node behavior has been presented based on graph theory, and a measurement method of trusted-degree is proposed. Because of the uncertainty of trust relationship, this paper has put forward the random trusted-order and firstly introduces the construction of trust relations space (TRS) based on trusted order. Based on all those above, the paper describes new method and strategy which monitor and measure the node behavior on the expectancy behavior character for trusted compute of the node. According to manifestation of node behavior and historical information, adjust and predict the trusted-order of node behavior. The paper finally establishes dynamic trust evaluation model based node behavior characters, and then it discusses the trusted measurement method which measures the connection and hyperlink for node behavior of network in trust relationship space.**

Keywords- Trust relations; trust relationship graph; trusted-order; random trust relationship.

I. INTRODUCTION

In recent years, with the wide application of trusted computing in the network security area, the studies of trust relationship in the node behavior of network have been made an important point [1, 2, and 3]. However, conventional environment of trust relationship conduction is "man around the model of experience", which is difficult to deal with the trust featured by procedures of brain thinking. Because of the almighty ability of experience model while processing empiricism information, people depend on experience model to a great extent. When the results from experience model are different from the reality significantly, people will doubt it.

This is the reason of the occurrence of the incompatible problems when traditional information of trust relationship conducting ways is applied to process node behavior of network system.

The trust network model is the prerequisite of trusted computing, how to evaluate the trust in the network, there is not a unity and general method up to now. In fact, the trust network is a kind of network with trust relationships, trust relationship networks can be abstracted as a kind of topological relationships in mathematics. Recently, the research around the model of trust networks is come from different angles [1-6]. But their common point is seeking a formal representation method reasonably.

Studies in literature [3] have shown that a trust relationship can be expressed with a graph. In the paper we study a quantitative expression of trust relationship in network system by using the method of graph theory. Then it measures the trusted degree of each node, and it also presents the trusted measurement of the connection and hyper connection for node behavior of network.

The object of this paper is to establishes dynamic trust evaluation model based on node behavior characters, Through the construction of the relationship between practical node behavior characters and on-the-spot model, it sets up a couple of mapping models of trust relationship, and sketches the skeleton of relationship mapping inversion.

The paper is organized as follows. Section 2 introduces the basic concept of trust relationship based on graph theory, and some properties of the evaluate principle of the trusted network. Section 3 studies measurement of trusted relationship. Section 4 we present dynamic trust evaluation model based on random trusted relationship. Section 5, conclusion puts forward the discoveries of this research and future research direction.

II. BASIC CONCEPTS AND METHODS

A. Certainty trusted relational graphs

Suppose that a trusted network T_N can be expressed as the corresponding graph $G=\{V,E\}$, where $V=\{v_1,v_2,\Lambda\ ,v_n\}$ is a node set of T_N, and $E=\{e_1,e_2,\Lambda\ ,e_m\}$ is an edge set. Moreover, $e_k : v_i \to v_j$ ($k=1,2,\Lambda\ ,m$; $i,j=1,2,\Lambda\ ,n)$, it represents that there is a trust relationship between v_i and v_j, namely v_i trusts v_j. Therefore, the trusted graph G is called a directed graph. For example, given a trusted network with five vertices, based on the analyzing of network behaviors, the trust relationship of vertices is described as follows:

$v_1 \to v_5$, $v_2 \to v_1$, $v_2 \to v_5$,

$v_3 \to v_2$, $v_3 \to v_4$, $v_3 \to v_5$,

$v_4 \to v_1$, $v_4 \to v_3$, $v_4 \to v_5$,

$v_5 \to v_1$, $v_5 \to v_3$.

Its adjacency matrix is

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix},$$

and the trusted relational graph G is shown in Figure 1.

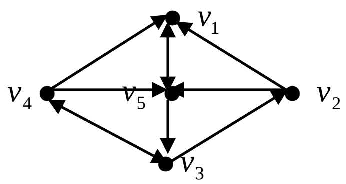


Figure 1. Trusted relational graph

B. Trusted relational trees

Given a trusted network with five vertices (As shown in Figure 1), Let $D(1) = AI = (1, 2, 3, 3, 2)^T$, where $I = (1, 1, 1, 1, 1)^T$, then $D(1)$ is a trusted level vector of each node. Such as the number of trusted vectors about the node v_4 and v_5 is 3 and 2 respectively. According to this a conclusion can be drawn that the trusted level of the node v_4 is taller than the node v_5.

But, the number of trusted vectors about the node v_5 and v_2 are both 2, how to distinguish the difference of v_5 and v_2?

On the analysis of Figure 1 it was found that v_1 and v_5 have a trusted relationship with v_2, and the number of trusted vectors about the node v_1 and v_5 is 1 and 2 respectively. The Node v_1 and v_3 have trusted relationships with v_5, and the number of trusted vectors about the node v_1 and v_3 is 1 and 3 respectively.

It can be seen that the indirect trusted level of v_2 is 3, and the indirect trusted level of v_5 is 4. Thereby, it may be taken for granted that the trusted level of v_5 is taller than the node v_2.

Based on the graph theory, the trusted path determines the trusted level in trusted networks. The above analysis can be representing by the tree in figure 2.

v_1 v_2 v_3

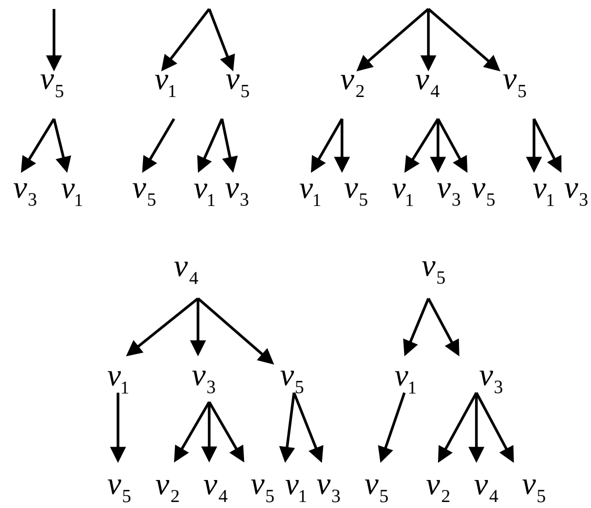


Figure 2. Trusted relational tree

III. Measurement In Trusted Relationship

The indirect trusted vector of each node in the network (when K=2) is as follows: $TD(2) = (2, 4, 7, 6, 4)^T$, then it is certain that $TD(K+1)$ can be measure the trusted level (trusted degree) accurately than $TD(K)$. In most cases, we must think about the limit of $TD(K)$, when $K \to \infty$. In order to ensure the limit convergence, and furthermore, the measuring value of each node should be trusted degree, therefore this paper regard the following limit as the measuring of each node in the trusted relational graph in networks:

$$\lim_{k\to\infty} \frac{TD(K)}{I^T TD(K)}.$$

It will find in fig. 2 that the measurement of the trusted degree about each node is the number of the path in the directed tree, which takes each node for the root. And then the relations are extended to the general case, the definition is as follows:

Definition 3.1 In a network with n nodes, the trusted capability (trusted degree) of a node v_i can be determining by the number of the path, which connects with the K-th path and starts from the node v_i. This number is called the K-th trusted capability, and denoted as $td_k(v_i)$. Vector

$$TD(k) = \{td_k(v_1), td_k(v_2), \Lambda, td_k(v_n)\}$$

is called the *K*-th trusted capability vector of the trusted relational graph G.

Definition 3.2 In a network of n vertices, a limit

$$td(v_i) = \lim_{k\to\infty} \frac{td_k(v_i)}{\sum_{i=1}^{n} td_k(v_i)}$$

is entitled relatively limit trusted degree of the node v_i, it is called a trusted degree for short. For that reason, we have

$$T = (td(v_1), td(v_2), \Lambda, td(v_n))^T = \lim_{k\to\infty} \frac{TD(k)}{I^T TD(k)},$$

which is called a trusted vector of each node in the trusted relational graph, where $I = (1,1,\Lambda,1)^T$.

Theorem 3.1 Let G be a trusted relational graph of n vertices, its adjacency matrix is A, if G is bidirectional connected and $n \geq 4$, α_1 is an eigenvector corresponding to the biggest values of A, then T exists certainly and $T = \alpha_1 / (I^T \alpha_1)$, moreover $\| \alpha_1 / (I^T \alpha_1) \| = 1$.

It can seem from theorem3.1, the *K*-th trusted degree $TD(k)$ of each node is computable in the trusted relational graph G with *n* vertices, and it can be obtain by the following algorithm:

（1）when $k = 0$， $TD(0) = I$；

（2）when $k = 1$，2，…， $TD(k) = ATD(k-1)$； $T\overline{D}(k) = TD(k)/(I^T TD(k))$；

（3）when given precision $e > 0$， calculated until $k = m$， if it is satisfied：

$$\| T\overline{D}(k) - T\overline{D}(k) \| < e,$$

then stopped calculating to choose $T = T\overline{D}(m)$.

The algorithm given in theorem3.1 can be put to use in network according to different trusted levels. Regardless of the connected meaning of network note, it always measures the trusted degree in the trusted relational graph. Meanwhile, the trusted vector T can be regaled as a weighted vector, which expresses the trusted degree of each node in the trusted relational graph.

IV. RANDOM TRUSTED RELATIONSHIP

Thinking about the trusted relational graph of discussion in the previous section, if it has $v_i \to v_j$, then it exists the trusted relationship of completely specified between v_i and v_j

It is called certainty trusted relational graph which has the trusted relationship of completely specified. In fact, trusted relationship is uncertainty in lots of trusted networks. For example, the trusted relationship among people in the Internet, because of the vitality of network activities, the trusted relationship of network is uncertainty, for this reason, the uncertain research methods is used to analyze the trusted relationship in network.

The random trusted relationship is expressed by the random graph for trusted relationship, it has respective trusted relational graph in the basis of different network activities and space-time states. Furthermore, the extent of trusted relationships presents certain probabilistic characteristics with the change of network activities; we can use $P_{ij} (0 \leq P_{ij} \leq 1)$ to express the arisen probability of $v_i \to v_j$. Thereby the trusted relationships in the network consisted of n nodes can be expressed by a family of trusted relational graphs, the family of trusted relational graphs are noted as $G(n,(P_{ij}))$, it is called a random trusted relational graph. When $v_i = v_j$, take for granted, $P_{ij} = 0$; When $i, j = 1,2\Lambda, n$, $i \neq j$, $0 \leq P_{ij} \leq 1$, apparently, $P = [P_{ij}]$ constitutes a square matrix of order n, $G(n,(P_{ij}))$ is called a probability matrix. Suppose the connection of each node be random and independence, then a definition is as follows:

Definition 4.1 A directed and weighted graph, which weighted is a probability matrix P, and it is called a network expression of $G(n,(P_{ij}))$ that is noted as $N(n,P)$.

Definition 4.2 The weighted product of each edge in the directed path L is called a transfer probability in $N(n,P)$. It is called the *k*-th order dispersive degree of the node v_i that the sum of all of transfer probabilities with k connective paths, which starting from the node v_i. Noted as $N_k(v_i)$, and

$$N(k) = (N_k(v_1), N_k(v_2), \Lambda, N_k(v_n))^T.$$

Definition 4.3 The limit $\lim_{k\to\infty} \frac{N_k(v_i)}{I^T N_k(v_i)}$ is called a limit transfer probability of a node v_i.

Based on the probability theory, it is well known that the transfer probability of the path L_{ij} (as dependence), which is from the node v_i to v_j in $N(n,P)$, is the present probability of L_{ij} in $G(n,(P_{ij}))$, that is to say it is a probability of the directed connection (trusted relational chain)between the node v_i and v_j in the random trusted relational graph $G(n,(P_{ij}))$. It is still used $td_k(v_i)$ to express the number of paths, which starting from the node v_i and taking with k paths. We can prove as follows:

Theorem 4.1 Let $N(n,P)$ be disconnected, $n \geq 4$, then the limit transfer probability of each node exists certainly, and equals to the limit

$$\lim_{k\to\infty}\frac{P^k I}{I^T P^k I}.$$

Deduction 4.1 There has $N(k)=P^k I$ in $N(n,P)$.

Theorem 4.2 Let $N(k)$ be the *k*-th order dispersive degree vector of each node in $N(n,P)$, and let $TD(k)$ be the number vector of starting from each node and taking with k paths in $G(n,(P_{ij}))$, then $E(TD(k))=N(k)$ is obtained. This theorem explains that the *k*-th order dispersive degree of the node v_i in $N(n,P)$ is the mathematical expectation of the number of paths, which starting from the node v_i and taking with k paths in $G(n,(P_{ij}))$. Since the measurement of trusted levels for a certain node can be expressed by the number of paths starting from the node. Hereby, we regard the limit transfer probability vector as the weighted vector T of a certain node in the random trusted relational graph. According to theorem 4.2, we can obtain the same arithmetic as theorem 1, so for as changing the adjacency matrix A for the probability matrix P in $N(n,P)$. If and only if $P_{ij}=0$ or $P_{ij}=1$, a random trusted relational graph turns into a certainty trusted relational graph, so the latter is a special case of the former.

V. CONCLUSIONS

In the development of the trusted computing, theoretical research lags behind practical. The trusted measurement is the basic theory of the trusted computing, and is also a key technology in the process of development of the trusted computing. In this paper, a certainty trusted network and a random trusted network were introduced respectively. Then a measurement method of the trusted degree was presented and its arithmetic was described. These theories and methods will help the development of the trusted computing. For future works, the methods will be optimized, which not only depict the fact but also can be used simply and practically.

REFERENCES

[1] Zhuo lingxiao. Discrete Mathematics, Shanghai:Shanghai Scientific and Technical Literature Publishing House, 1989 (Chinese).

[2] Lin chuang, Peng xuemei. **Research on Trustworthy Networks**. Chinese Journal of Computers. 2005, vol. 28, pp.751-758 (Chinese).

[3] Ammann P, Wijesekera D, Kaushik S. Scalable Graph-based Network Vulnerability Analysis, Proceedings of the 9th ACM Conference on Computer and Comm. Security. New York, USA: ACM Press, 2002, pp.217-224.

[4] Swiler L P, Phillips C, Gaylor T. A Graph-based Network Vulnerability Analysis System, Sandia National Laboratories, Albuquerque, USA, Technical Report: SAND97-3010/1, 1998.

[5] Theodorakopoulos G, Baras j s. On trust models and trust evaluation metrics for ad-hoc networks. IEEE Journal on Selected Areas in Communications. 2006, vol.24, pp.318-328.

[6] Sun Y, Yu W. Han Z,Liu K J R. Information theoretic framework of trust modeling and evaluation for ad hoe networks. IEEE Journal on Selected Areas in Communications, 2006, vol.249, pp.305-319.

[7] M. Gómez, J. Carbó, and C. B. Earle. Honesty and trust revisited: the advantages of being neutral about other's cognitive models. Autonomous Agents and Multi-Agent Systems, 2007, vol.15, pp.313-335.

[8] Hofstede, G. J., Jonker, C. M., Meijer, S. and Verwaart, T. Modelling trade and trust across cultures, in Ketil St¿len et al., ed., `Trust Management, 4th Inter-national Conference, iTrust 2006', Vol. 3986 of Lecture Notes in Computer Science, Springer-Verlag Berlin Hei-delberg, Pisa, Italy, 2006, pp. 120-135.

[9] Y. Wang and M. P. Singh. Formal trust model for multiagentsystems. In Proc. 20th International Joint Conference onArtificial Intelligence (IJCAI), 2007, pp. 1551–1556.

[10] Li Xiao-Yong, Gui Xiao-Lin. Research on adaptive prediction model of dynamic trust relationship in Open distributed systems. Journal of Computational Information Systems, 2008, vol.4, pp.1427-1434(Chinese).

AUTHORS PROFILE

He Ping is a professor of the Department of Information at Liaoning Police Academy, P.R. China. He is currently Deputy Chairman of the Centre of Information Development at Management Science Academy of China. In 1986 He advance system non-optimum analysis and founded research institute. He has researched analysis of information system for more than 20 years. Since 1990 his work is optimization research on management information system. He has published more than 100 papers and ten books, and is editor of several scientific journals. In 1992 awards Prize for the Outstanding Contribution Recipients of Special Government Allowances P. R. China

An incremental learning algorithm considering texts' reliability

Xinghua Fan, Shaozhu Wang
College of Computer Science and Technology,
Chongqing University of Posts and Telecommunications,
Chongqing, China

***Abstract*—The sequence of texts selected obviously influences the accuracy of classification. Some sequences may make the performance of classification poor. For overcoming this problem, an incremental learning algorithm considering texts' reliability, which finds reliable texts and selects them preferentially, is proposed in this paper. To find reliable texts, it uses two evaluation methods of FEM and SEM, which are proposed according to the text distribution of unlabeled texts. The results of the last experiments not only verify the effectiveness of the two evaluation methods but also indicate that the proposed incremental learning algorithm has advantages of fast training speed, high accuracy of classification, and steady performance.**

Keywords-text classification; incremental learning; reliability; text distribution; evaluation.

I. INTRODUCTION

Conventional methods of text classification, for example, Centroid, Native Bayes (NB), K-Nearest Neighbor (KNN), Support Vector Machines (SVM) and so on, which are not incremental learning methods, obtain the texts' classification model according to existing labeled training set. However the training set can't be obtained in one time; the methods above are not always effective.

Incremental learning can solve the above problem very well. With the advancement of classification process, in the incremental learning, the scale of training set expands unceasingly; new texts are labeled and added to the training set gradually. Among those text candidates, which texts to select first is the critical point of this classification.

There are two models of selecting texts to add into labeled training set: passive classification model and active classification model.

Passive classification model, which selects the training texts randomly and accepts the text information passively; it believes that the training texts distribute independently in most of classification learning, so passive classification model has obvious deficiency:

- Make the noise spread down, affect the accuracy of classification.
- Ignore the relationship among texts in new incremental training set.

Active classification model selects texts actively. It is a subconscious and higher level learning model, which selects the optimized texts to improve classifier's performance. So compared with passive learning, active learning attracts more researchers' attentions. Reference [1] proposed an algorithm to select a text by calculating the 0-1 loss rate every time, and the algorithm improved the performance of classifier. But large amount of calculation and high time complexity are the algorithm's shortages. Reference [2] proposed an algorithm to select some texts by clustering. This algorithm reduced the training time, but it would be affected by noise data easily and lead to large fluctuations of classifier's performance. No matter the algorithm of selecting one text or that of batch selection [1][2][3][4][5], texts are selected by external evaluation algorithms which need a lot of additional computing, so most of incremental learning algorithms have poor efficiency.

From above, the method to select texts is very important. A good method not only improves the classifier's performance but also reduces the training time. For solving this problem, an incremental learning algorithm considering texts' reliability is proposed in this paper. It includes two evaluation methods named first evaluation method (FEM) and second evaluation method (SEM), which select new texts according to the results in Reference [6], are proposed in this paper. Reference [6] showed that classifier's performance will be improved obviously when the correctly labeled texts are added preferentially. And these two methods are complementary to each other and have low computational complexity, which make full use of useful information among texts and the intermediate data-out in the process of training classifier. For incremental bayesian model [1] can make good use of its prior knowable, it is used to improve the availability of the algorithm proposed. The structure of this paper is organized as follows: the algorithm is introduced in detail in Section II. Section III demonstrates experimental results on artificial and real datasets. We conclude our study in Section IV.

II. AN INCREAMTAL ALGORITHM CONSIDERING TEXTS' RELIABILITY

In this section, a new incremental algorithm will be introduced in detail. The two FEM and SEM methods are important parts of the algorithm .They are inspired from the regularity of texts' distribution, so the corresponding regularity of texts' distribution will be introduced first, and then introduce evaluation methods and their relation. The details of each step of the new algorithm will be given in the end of this section.

A. The first evaluation method (FEM)

Given text vector $d = (W_1, W_2, \ldots, W_n)$ ($W_i = 0$ or 1). If the i-th feature appear in the text, $W_i = 1$, otherwise $W_i = 0$. Supposed that $p_{ki} = \{W_k = 1 | c_i\}$, and $p\{\bullet\}$ is the probability for incident $\{\bullet\}$.The discriminant function[7] of Naive Bayesian classifier can be expressed as:

$$c^* = \arg\max(\log P\{c_i\} + \sum_{k=1}^{D} \log(1-p_{ki}) + \sum_{k=1}^{D} W_k \log\frac{p_{ki}}{1-p_{ki}}) \quad (1)$$

Supposed that:

$$MV_i = \log P\{c_i\} + \sum_{k=1}^{M} \log(1-p_{ki}) + \sum_{k=1}^{M} W_k \log\frac{p_{ki}}{1-p_{ki}}) \quad (2)$$

$$MV_{\max} = \max_{c_i \in C}(MV_i) \quad (3)$$

$$MV_{\sec} = \operatorname*{sec\,\mathit{ond}}_{c_i \in C}\ (MV_i) \quad (4)$$

MV_i is the probability of text vector *d*, which is estimated by feature and belongs to $c_i \in C$, and *C* is the predefined type set. MV_{max} is the maximum of all probabilities in text vector *d*; MV_{sec} is the second maximum of all probabilities in text vector *d*.

The value of rewritten MV_i is negative, normalizing for MV_i:

$$p = MV_{\max} / MV_{\sec} \quad (5)$$

Take the corpus, which will be introduced in section III, as samples. We randomly divide the 6000 texts into 3 groups of datasets. Each group contains a labeled training set of different scales which are 20 texts, 200 texts, 2000 texts, and a common new incremental training set composed of 400 unlabeled texts. Then construct the classifier and classify the new incremental training set. The relationship between the p-value and the number of misclassified texts is shown in table I.

The largest set of the correct texts refers to the texts contained within the p-value, where the misclassified text appears for the first time. Table I shows that the misclassified texts appear and increase gradually with the p-value changing. The greater the p-value is, the more misclassified texts appear. If a set within p-value contains no misclassified texts, it is the correct interval, and names the set of the others texts as fuzzy interval. Table I plus table II, show that with the size of labeled set increasing, more and more texts are distributed in the correct interval. In addition, table I plus table II, show the existence of the correct interval has nothing to do with the scale of labeled texts; the scale only affects the number of texts in correct interval.

TABLE I. THE RELATIONSHIP BETWEEN P-VALUE AND THE NUMBER OF MISCLASSIFIED TEXTS

p's range	The number of misclassified texts		
	Labeled texts(20)	*Labeled texts (200)*	*Labeled texts (2000)*
(0,0.5)	0	0	0
[0.5,0.6)	3	0	0
[0.6,0.7)	4	1	1
[0.7,0.8)	6	1	1
[0.8,0.9)	12	7	3
[0.9,1]	22	6	5

TABLE II. THE RELATIONSHIP BETWEEN LABELED TEXTS' SCALE AND PERCENTAGE OF THE LARGEST SET OF THE CORRECT TEXTS

Labeled texts	20	200	2000
Percentage (%)	40.25	78.75	79.25

Table II shows that when the number of labeled texts is equal or more than 200, nearly 80% of the texts are distributed in the correct interval. As the initial labeled texts are few, in order to maximize the number of the new incremental unlabeled texts falling into the correct interval, the new incremental training set is divided into a number of subsets each containing 100 texts. Carrying out incremental learning among the subset takes advantage of the size and performance of intermediate classifiers.

From the regularity of texts' distribution mentioned, the method of FEM is proposed as follows:

FEM: in the output of classifier, if *p* which is calculated by formula $p = MV_{\max} / MV_{\sec}$ not exceeds a threshold α, corresponding texts are all corrected classified texts.

In order to determine the value of α, take the corpus, which will be introduced in section III, as samples. Take 5 labeled texts each category to construct training set with 20 labeled texts, and classify for 600 new texts by constructed initial classifier, the relationship between the *p-value* and the distribution of misclassified texts is shown in Fig. 1.

Fig. 1 shows that the value of α should be between 0.5 and 0.6, in order to ensure that the texts in this interval are all up to the requirements, α's value should be set to 0.5.

B. The second evaluation method (SEM)

After the FEM assessment, the texts incorrectly labeled by the current classifier concentratedly distribute in fuzzy interval. Deal the texts in fuzzy interval with Affinity Propagation (AP) clustering [8], and get many clusters. In each cluster, the first text is a representative for the others. And most of the texts have the same label as the first text in each cluster. The results of the experiments in Reference [3], which only uses noun as features, show that: more than 90% of the texts have the same label as the representative text. So the result can be used for judging whether the classifier is able to correctly identify the cluster. Take the corpus, which will be introduced in section III, as samples. We randomly get a group of dataset from the 6000 texts.

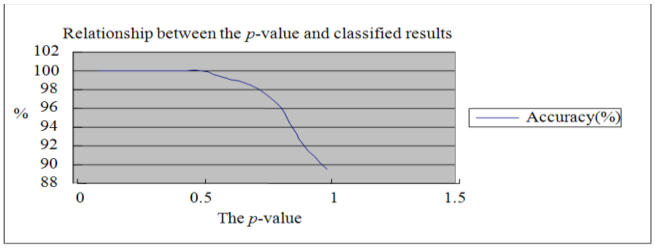


Figure 1. Relationship between the p-value and distribution of misclassified texts

TABLE III. THE ACTUAL LABELS AND OBTAINED LABELS OF CLUSTERS

Texts' actual category	Texts' category of current classifier
2 2 2	2 2 2
1 1 3 1 1 1	1 1 3 1 2 1
3 3 3 3 4 3	3 3 2 1 3 2
4 1 4	4 1 4
3 2 3 3	3 1 2 3

The dataset contains a labeled training set composed of 5 texts each category and a new incremental training set composed of 600 texts. Classify for 600 new texts by initial classifier constructed, α 's value is set to 0.5, do AP clustering for texts in fuzzy intervals. Analyzing the first 5 clusters, their actual labels and obtained labels are shown in table III.

Analyze the label of the third cluster, a conclusion is got, the labeled training set will be introduced four incorrectly labeled texts by the current classifier. In order to avoid this, we only join the texts which have the same label as the representative text into labeled training set, compute the ratio $\beta = num1/num2$, where $num1$ is the number of the texts which have the same label as representative text, $num2$ is the number of the whole cluster. Set a threshold δ , and it means that the current classifier can't correctly identify the cluster if β is less than δ , remove the cluster. And put forward the method of SEM as follows:

SEM: Classify the texts in each cluster by the current classifier, and then calculate the ratio $\beta = num1/num2$. Set a threshold δ **,** if β is not less than δ , it believes that the texts in corresponding cluster can be identified by the current classifier.

In order to determine the value of parameter δ , take the corpus, which will be introduced in section III, as samples. Take 5 texts as labeled texts each category to construct training set with 20 labeled texts, and classify for 600 new texts by initial classifier constructed, the fuzzy intervals are obtained when α =0.5, the relationship between the value of β and learning results of texts in the fuzzy intervals is shown in Fig. 2. As is shown in Fig. 2, the learning performance of classifier is the best when the value of δ near 0.8.

C. Complementarities of FEM and SEM

After the FEM assessment, if continue to do incremental learning for texts in fuzzy intervals by current classifier, the accuracy of learning is not very good. Take the corpus, which will be introduced in section III, as samples. Take 5 texts as labeled texts each category to construct labeled training set

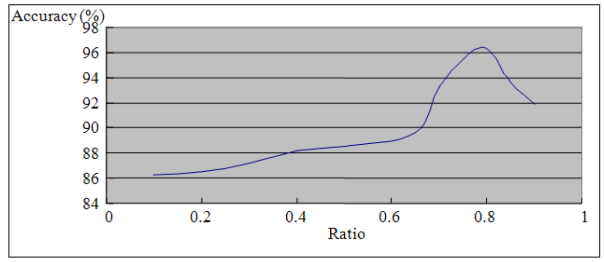


Figure 2. Relationship between the value of β and accuracy of texts' learning

TABLE IV. MACRO_F OF LEARNING FOR TEXTS IN FUZZY INTERVALS

Scale of labeled set	20	200	2000
Macro_F (%)	79.68	82.36	86.72

TABLE V. MACRO_F OF LEARNING BY SEM ONLY

Scale of labeled set	20	200	2000
Macro_F （%）	91.04	94.57	99.18

with 20 labeled texts, and classify for 600 new texts by constructed initial classifier, do incremental learning sequentially for texts in fuzzy intervals when α is equal to 0.5, and take 50 and 500 labeled texts as well. Accuracy is shown in table IV.

As is shown in table IV, with the expansion of labeled training set's scale, the accuracy is better. And combined with figure 2, the accuracy rises by 79.68% to more than 96% by adding SEM.

Current classifier's performance need to be considered in SEM, so it will obtain better results when knowledge of classifier is abundant. If performance of initial classifier is not very good, it will yield big error in calculating the value of β , noise data is introduced and finally lead to the bad performance of classifier. Set the value of δ to 0.8, the results of incremental learning by SEM only (use the same corpus with table IV) are shown in table V.

As is shown in table V, if evaluated by SEM only, the final classifier's performance is obviously affected by initial classifier. The reason is that noise data is introduced into labeled training set in previous iteration. As is shown, the larger scale the initial labeled training set is, the better result the SEM can obtain. And eighty percent of texts are in the correct interval after evaluating by FEM which can lead to obtain a large amount of labeled training set. So FEM and SEM are complementary to each other.

D. Description of algorithm

The two mentioned evaluation methods provide theory basis for the new algorithm proposed in the paper. The algorithm, which uses the two evaluation methods to make the reliable texts join labeled set preferentially, improves the performance of the classifier and reduces the influence by noise data. Because the proportion of texts in correct interval is influenced by the scale of the initial labeled set, divide the unlabeled set into some subsets. So more texts can be in correct interval by intermediate classifies. The algorithm can be described concretely as follows:

Input: Labeled training set $D=\{d_1,d_2,\ldots,d_N\}$

New incremental training set $T=\{t_1,t_2,\ldots,t_m\}$

Output: Classifier *C*

Step1: Use the CHI formula to do the feature selection for training set *D*, and learn a classifier;

Step2: If $T=\Phi$ (Φ is the empty set), go to step5;

Step3: Randomly select 100 texts from *T*, classify each text t_p in new incremental training set *T* by current classifier *C*, select correct texts estimated by FEM to form a new subset $T'\subset T$, and add them into the training set *D*, the rest is added into the untrusting set *U*;

Step4: $T=T-T'$, go to step1;

Step5: If $U=\Phi$, return the classifier, and end the algorithm; else continue;

Step6: Do clustering for the untrusting set *U*, formed *k* subsets $U=\{R_1,R_2,\ldots,R_k\}$, remove the subsets which only have a single text to set U, then select the first text of each cluster respectively to construct a representative text set $r=\{r_1,r_2,\ldots,r_m\}$, $m<k$;

Step7: If $r=\Phi$, go to step5, else for each of the text $r_i\in r$, to repeat the follows:

a) Classify texts r_i by current classifier C, and obtain the label C_p;

b) Classify other texts in subset Ri which ri is in by current classifier C, and calculating the ratio (β) of num to NUM, where NUM is the total number of texts in the cluster which ri is in and num is the number of texts which are classified the same category with ri;

c) If $\beta>\delta$, join the texts including ri in Ri, which are classified the same label with ri, into T'', then update the set $r=r-r_i$;

d) $D=D+T''$, $U=U-T''$, use the CHI formula to select features for training set D, and learn classifier C.

III. Experiments

Five experiments are designed in this paper:

Exp.1: Verify the effectiveness of the correct set division.

Exp.2: Verify the effectiveness of fuzzy data processing.

Exp.3: Verify the effectiveness of subset division.

Exp.4: Verify the high efficiency and steady performance of the proposed method.

Exp.5: A test of training time and learning performance of different scales of new incremental training set.

A. The datasets of experiments

Datasets: The datasets used in experiments are all from netease and sina, which including four categories, and have total 6000 Chinese texts. In the 6000 Chinese texts, category of Olympics, Buddhism, Military and Computer has 1500 texts respectively. Form eight groups of corpus used in Exp.1, Exp.2, Exp.3 and Exp.4. Each group contains 5 initial labeled texts and 100 unlabeled texts each category from the 6000 texts randomly. And form four groups of corpus used in Exp.5. Each group contains a training set with 5 labeled texts each category, and a new incremental training of different scales which are 400 unlabeled texts, 800 unlabeled texts, 1200 unlabeled texts. The same texts mustn't appear in both initial labeled training set and unlabeled training set.

B. The feature selection in experiments

The feature selection method of CHI is used in experiments:

$$\chi^2(w,c)=\frac{N(AD-BC)^2}{(A+C)(B+D)(A+B)(C+D)} \tag{6}$$

Where, *c* is the category, *w* is the feature, *N* is the number of texts, *A* is times of *w* and *c* both appeared, *B* is times of *w* appeared but *c* not appeared, *C* is times of *c* appeared but *w* not appeared. *D* is times of *w* and *c* both not appeared.

C. Performance's assessment

Precision: $P=\frac{N_1}{N_2}\times 100\%$

Recall: $R=\frac{N_1}{N_3}\times 100\%$

Macro average: $\text{Macro\_F}=\frac{2\times P\times R}{P+R}\times 100\%$

Where, N_1 is the number of texts correctly classified in a category, N_2 is the number of texts classified in a category, N_3 is the number of texts in a category of test set.

D. Experimental Results

1) The methods in experiments are defined as:

NBTS: Incremental method considering texts' reliability proposed in this paper.

NBSS: Incremental method with SEM.

NBFS: Incremental method with FEM.

NBS: Incremental method without division subset.

NBKC: Quick clustering based incremental method proposed in reference [4].

EM: The standard Expectation Maximization (EM) algorithm [9].

2) The parameters setting

From the second section, if the classifier's performance is the best, the parameter α is equal to 0.5 and δ is equal to 0.8.

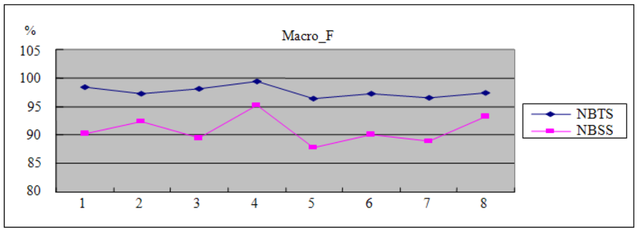


Figure 3. The learning results of NBTS and NBSS

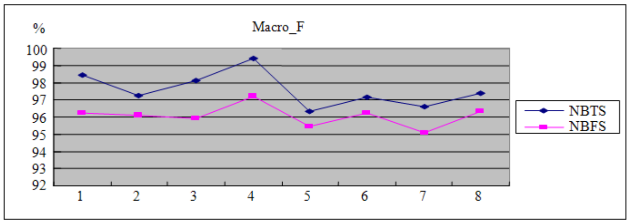


Figure 4. The learning results of NBTS and NBFS

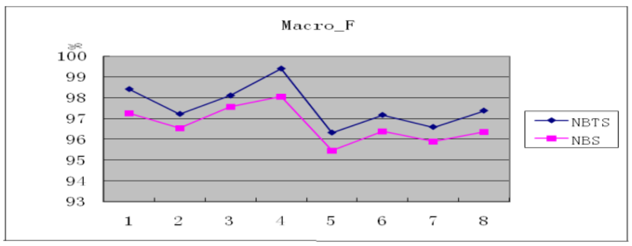


Figure 5. The learning results of NBTS and NBS

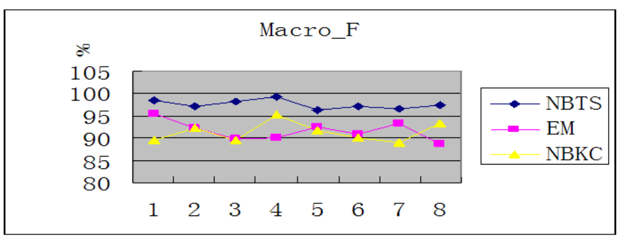


Figure 6. The learning results of the mentioned three incremental methods

TABLE VI. THE AVERAGE TIME CONSUMING OF THE MENTIONED TWO INCREMENTAL METHODS IN EXP. 3

Method	Average time consuming(s)
NBTS	115
NBS	135

TABLE VII. THE AVERAGE TIME CONSUMING OF THE MENTIONED THREE INCREMENTAL METHODS IN EXP. 4

Method	Average time consuming(s)
NBTS	115
EM	200
NBKC	1865

TABLE VIII. THE TRAINING TIME IN DIFFERENT SCALES OF NEW INCREMENTAL TRAINING SET MENTIONED IN THIS PAPER

Group number	The scale of new incremental training set and its training time(s)					
	400	*Time(s)*	*800*	*Time(s)*	*1200*	*Time(s)*
1	98.42	121	96.93	203	97.38	298
2	97.22	94	98.39	215	98.97	307
3	98.11	106	97.74	198	97.12	287
4	99.41	131	98.96	234	96.59	279

3) The results of experiments

Results of Exp. 1-Exp. 4 are shown in Fig. 3-Fig. 6 respectively.

The average time consuming of the methods in Exp. 3 and Exp. 4 are shown in table VI and VII respectively.

Results of Exp.5 are shown in table VIII.

E. Analyses of the experimental results

- Exp.1 shows that the classifier's performance is greatly improved by adding the correctly classified texts to labeled training set, Macro_F increases by about 7% relative to use SEM only. FEM's effectiveness is verified.
- Exp.2 shows that after using SEM to deal with fuzzy data, the classifier's performance increases by 2%. SEM's effectiveness is verified.
- Exp.3 shows that the learning method with division subsets not only improves the classifier's performance, but also shorts the train time. With increase of labeled training set's scale, more and more unlabeled texts lie in the correct interval. The intermediate classifiers are fully used by dividing subsets, more texts are added by FEM, the performance of the classifier is improved, the number of texts in fuzzy interval is reduced and clustering and text selection's time is shorter.
- Exp.4 shows that the classifier trained by proposed algorithm has better and steadier performance, for it decreases the disturbance of noise in the data sets.
- Exp.4 and Exp.5 show that the classifier trained by proposed algorithm has better performance and shorter train time than classifiers trained by other algorithms. The algorithm is more suitable for dealing large data.

IV. CONCLUSIONS

An incremental learning algorithm considering texts' reliability is proposed in this paper. Firstly, the new incremental training set is divided into subsets and the FEM method is used to find out the correct set interval of the subset, which made the number of labeled training set greatly increase.

Then the remaining fuzzy data was dealt by AP classification, and the learning sequence of noise data is further dealt by SEM. The experimental results show that the proposed algorithm is less affected by noise data and the performance of classifier is relatively stable. And the proposed incremental learning algorithm can train a classifier quickly.

ACKNOWLEDGMENT

The research is supported in part by the National Natural Science Foundation of China under grant number 60703010, cultivate project for significant research plan of the National Natural Science Foundation of China under the grant number 90924029 named "Research on the mechanism and related technology of public opinion on internet for abnormal emergency" and the Nature Science Foundation of Chongqing province in China under grant number CSTC, 2009BB2079.

REFERENCES

[1] Xiujun Gong, Shaohui Lin and Zhongzhi Shi, "An Incremental Bayes Classification Model," Chinese Journal of Computers, vol. 25, no. 6, pp. 645-650, 2002.

[2] Houfeng Ma and Xinghua Fan. "Improved incremental Bayes classification algorithm," Chinese Journal of Scientific Instrument, vol.28, no. 8III, pp. 312-316,2007.

[3] Houfeng Ma, Xinghua Fan and Ji Chen, "An incremental Chinese text classification algorithm based on quick clustering," 2008 International Symposiums on Information Processing. IEEE Computer Society, 2008, pp. 308-312.

[4] Xinghua Fan, Zhiyi Guo and Houfeng Ma, "An improved EM-based Semi-supervised Learning Method. 2009 International Joint Conference on Bioinformatics," Systems Biology and Intelligent Computing (IJCBS'09). IEEE Computer Society, 2009, pp. 529-532.

[5] Yimin Wen, Yang Yang and Baoliang Lv, "Study on Ensemble Learning and Its Application in Incremental Learning," Journal of Computer Research and Development, vol. 42 (supplement), pp. 222-227,2005.

[6] M. Li, Z. H. Zhou. SETRED: Self-training with editing[C]. In Proceedings of 9th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'05), Hanoi, Vietnam, LNAI 3518, 2005, pp. 611-621.

[7] X. H. Fan and M. S. Sun, "A High Performance Two-Class Chinese Text Categorization Method," Chinese Journal of Computers, vol. 29, no.1, pp. 124-131, 2006.

[8] Brendan J. Frey and Delbert Dueck, "Clustering by Passing Messages Between Data Points", Science, vol.315, February 2007, pp. 972-976.

[9] Arthur Dempster, Nan Laird and Donald Rubin, "Maximum likelihood from incomplete data via the EM algorithm," Journal of the Royal Statistical Society, Series B, vol.39, no.1, pp.1–38, 1977.

[10] Y. Jiang and Z. H. Zhou, "Editing training data for KNN classifiers with neutral network ensemble," In Proceeding of the 1st International Symposium on Neural Networks, Dalian, China, 2004, pp. 356-361.

[11] F. Sebastiani, "Machine learning in automated text categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1-47, 2002.

[12] Jinshu Su, Bofeng Zhang and Xin Xu, "Advances in Machine Learning Based Text Categorization," Journal of Software, vol. 17, no. 9, pp. 1848-1859, 2006.

[13] A. McCallum and K. Nigam, "A comparison of event models for Naïve Bayes text Classification," AAAI-98 Workshop on Learning for Text Categorization, 1998, pp. 41-48.

[14] D. D. Lewis, R. E. Schapire, J. P. Callan and R. Papka, "Training algoirthms for linear text classifiers," in Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1996, pp. 298-306.

[15] E. H. Han, G. Karypis and V. Kumar, "Text categorization using weight adjusted k-nearest neighbor classification," University of Minnesota, 1999, pp. 11-12.

[16] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in Machine Learning:ECML-98, Tenth European Conference on Machine Learning, 1998, pp. 137-142.

[17] M. Ruiz and P. Srinivasan, "Hierarchical text categorization using neural networks," Information Retrieval, vol. 5, no. 1, pp. 87-118, 2002.

[18] Z. H. Zhou and W. Tang, "Selective ensemble of decision trees," Lecture Notes in Artificial Intelligence 2639, Berlin: Springer, 2003, pp.476-483.

AUTHORS PROFILE

Xinghua Fan was born in 1972, in Chongqing, China. He received his Ph. D. from Chongqing University. He is a professor in Chongqing University of Posts and Telecomunications. His research interests include artificial intelligence, natural language processing and information retrieval.

Shaozhu Wang is a postgraduate in Chongqing University of Posts and Telecomunications. She was born in 1986, in Fuzhou, Fujian province, China. Her research interests include Chinese information processing and text classification.

Applying Social Network Analysis to Analyze a Web-Based Community

Mohammed Al-Taie
Master of computer Science and Communication dept.
Arts, Sciences and Technologies University
Beirut, Lebanon

Seifedine Kadry
Master of computer Science and Communication dept.
Arts, Sciences and Technologies University
Beirut, Lebanon

***Abstract*— this paper deals with a very renowned website (that is Book-Crossing) from two angles: The first angle focuses on the direct relations between users and books. Many things can be inferred from this part of analysis such as who is more interested in book reading than others and why? Which books are most popular and which users are most active and why? The task requires the use of certain social network analysis measures (e.g. degree centrality).**

What does it mean when two users like the same book? Is it the same when other two users have one thousand books in common? Who is more likely to be a friend of whom and why? Are there specific people in the community who are more qualified to establish large circles of social relations? These questions (and of course others) were answered through the other part of the analysis, which will take us to probe the potential social relations between users in this community. Although these relationships do not exist explicitly, they can be inferred with the help of affiliation network analysis and techniques such as m-slice.

Book-Crossing dataset, which covered four weeks of users' activities during 2004, has always been the focus of investigation for researchers interested in discovering patterns of users' preferences in order to offer the most possible accurate recommendations. However; the implicit social relationships among users that emerge (when putting users in groups based on similarity in book preferences) did not gain the same amount of attention. This could be due to the importance recommender systems attain these days (as compared to other research fields) as a result to the rapid spread of e-commerce websites that seek to market their products online.

Certain social network analysis software, namely Pajek, was used to explore different structural aspects of this community such as brokerage roles, triadic constraints and levels of cohesion.

Some overall statistics were also obtained such as network density, average geodesic distance and average degree.

Keywords- Affiliation Networks; Book-Crossing; Centrality Measures; Ego-Network; M-Slice Analysis; Pajek; Social Network Analysis; Social Networks.

I. INTRODUCTION

Social network analysis (SNA) is concerned with realizing the linkages among social entities and the implications of these linkages [33].

It has evolved due to the synergy of three fused (separated, in sometimes) strands. These three strands were formed from the efforts of sociometric analysts who worked on small groups and came up with technical advances in methods of graph theory, the Harvard researchers of the 1930s who discovered patterns of interpersonal relations and the formation of cliques, and the Manchester anthropologists who investigated the structure of community relations in tribal and village societies [28].

The essential goal of SNA is to examine relationships among individuals, such as influence, communication, advice, friendship, trust etc., as researchers are interested in the evolution of these relationships and the overall structure, in addition to their influence on both individual behavior and group performance [29].

As for [23], they conducted a research to measure the growth of SNA field for the period (1963-2000). They consulted three databases that related to three branches of science (namely sociology, medicine and psychology). Among their findings were that the real growth of the field began in 1981 and there was no sign of decline and that the development in the field began in sociology faster than what it was in medicine and psychology. They noticed that the success which SNA has witnessed in the eighties was due to the institutionalization of social network analysis since late seventies and the recent availability of textbooks and software packages.

Today, social network analysts have an international organization called 'The International Network for Social Network Analysis' or INSNA, which holds annual meetings and issues a number of professional journals. Also, a number of centers for network searching and training have opened worldwide [8].

II. APPLICATIONS OF SOCIAL NETWORK ANALYSIS

SNA is involved in a many tasks, such as identifying most important actors in a social network through the use of centrality analysis, community detection, identifying the role associated with each member through conducting role analysis, network modeling for large-scale complex networks, how the information diffuses in a network and viral marketing [31].

A. Semantic Web

The idea of semantic web is to implement advanced knowledge techniques to fill the gap between machine and human. This implies providing the required knowledge that enables a computer to easily process and reason [21].

As for [7], he merged the semantic web frameworks model (which allows representing and exchanging knowledge across web application) and SNA model (which proposes graph algorithms to characterize the structure of a social network and its strategic positions). This combination was necessary in order to go beyond mining the flat link structure of social graphs.

B. Social recommendation systems

The use of SNA in the field of designing recommender systems (RS) is still in primitive stages [36]. However; it is expected that new methods using SNA will be incorporated in recommender system design [24], [36].

For [15], they presented a collaborative-based recommendation system that uses trading relationships to calculate level of recommendation for trusted online auction sellers. They used k-core, center weights algorithms and two social network indicators to create a recommender system that could suggest risks of collusion associated with an account.

C. Software development

Social network analysis in software engineering plays an important role in project support as more projects have to be conducted in globally-distributed settings.

In [16], they developed a method and a tool implementation to apply SNA techniques in distributed collaborative software development, as this provides surpassing information on expertise location, coworker activities and personnel development.

As for [20], they applied SNA to code churn information, as an additional means to predict software failures. Code churn is a software development artifact (common to most large projects) and can be used to predict failures at the file level. Their goal was to examine human factor in failure predicting. They conducted their case study on a large Nortel networking product, comprising more than 11000 files and three million lines of code.

D. Health

Network analysis, more and more, is becoming well-known in infectious disease epidemiology, such as Human Immunodeficiency Virus (HIV) and Sexually Transmitted Diseases (STD). Also, a strong trend is emerging towards using inter-organizational network analysis to detect patterns of health care delivery such as service integration and collaboration [13].

For [26], they conducted a study to know the relationship between SNA and the epidemiology and prevention of STD. They argue that SNA will be of a great utility in the study of STD.

As for [10], they found that the traditional contact tracing (the technique which they used at the beginning of their search to discover the reason behind the spread of tuberculosis in a medium-size community in British Colombia) did not identify the source of the disease. By using whole-Genome sequencing and SNA, they discovered that the cause was related to socio-environmental factor.

E. Cybercrimes

Cybercrimes are offences that are committed against individuals or groups of individuals with a criminal motive using modern telecommunication networks such as Internet and mobile phones [35].

Ref. [37] presented a framework to analyze and visualize weblog social networks. A weblog is a website where the contents are formulated in a diary style and maintained by the blogger. This environment makes a good platform for organizing crimes. With the ability to analyze and visualize weblog social networks in crime-related matters, intelligence agencies will have additional techniques to secure the society.

To investigate hacker community, [18] examined the social structure of an unknown hacker community called 'Shadowcrew'. For the investigation, they used text mining and network analysis to discover the relationships among hackers. Their work showed the decentralized composition of that community. Based on that analysis, they found that this community exhibits features of deviant team organization structure.

F. Business

SNA applies to a wide range of business fields, including human resources, knowledge management and collaboration, team building, sales and marketing and strategy.

Ref. [12] looked at SNA as a tool which can enhance the empirical quality of Human Resource Development (HRD) theory in areas such as organizational development, organizational learning, etc. He argues that SNA will add much to HRD fields by measuring the relations between individuals, and the effect those relations have on human capital output.

For [6], they studied the influence of SNA and sentiment analysis in predicting business trends. They focused on predicting the successes of new movies, in the box office, for the first four weeks. They were trying to predict prices on the Hollywood Stock Exchange (HSE), and the ratio of gross income to the budget of the production. They depended on data posts from Internet Movie Database (IMDb) forums to get sentiment metrics for positivity and negativity based on forum discussions.

Through using a Twitter dataset, [38] tried to predict stock market indicators such as Dow Jones, S&P500 and NASDAQ. They took about one hundredth of the total Twitter data that covered six months of activity. Through analyzing the relationship between data and stock market indicators, they found that emotional tweets displayed negative correlation to NASDAQ and S&P500, but gave positive correlation to VIX. They concluded that Twitter analysis can be used as a tool to predict stock market of the next day.

G. Collaborative Learning

Social network analysis provides meaningful and quantitative insights into the quality of knowledge construction process. It can effectively assess the performance of knowledge building process.

Ref. [27] showed that concepts of SNA, adapted to the collaborative distant-learning, can assist measuring small group cohesion. Their data were taken from distance-learning experiment of ten weeks. They used different ways to measure cohesion in order to highlight active subgroups, isolated people and roles of the members in the group communication structure. They argue that their method can show global attributes at the group level and individual level, and will help the tutor in following the collaboration in the group.

Ref. [25] has investigated the potential use of SNA to evaluate programs that seek to enhance school performance through encouraging greater collaboration among teachers. Through gathering data about teacher collaboration in schools, they mapped the distribution of expertise and resources needed to achieve reforms. One of their findings was that although the majority of teachers consider collecting social network data to be feasible, other teachers show concerns related to privacy and data sharing.

III. GRAPH THEORY

The origins of graph theory can be traced back to Euler's work on the Konigsberg bridges problem (1735), which subsequently led to the concept of an eulerian graph. The study of cycles on polyhedra by the Revd. Thomas Penyngton Kirkman (1800-95) and Sir William Rowan Hamilton (1805-05) led to the concept of a Hamiltonian graph [11].

The simplest definition of a graph is that it is a set of points and lines connecting some pairs of the points. Points are called 'vertices', and lines are called 'edges'. A graph G is a set X of vertices together with a set E of edges and it is written as: G = (X, E).

For a given vertex (x), the number of all vertices adjacent to it is called 'degree' of the vertex x, denoted by d(x). The maximum degree over all vertices is called the maximum degree of G, denoted by □ (G).

The adjacent vertices are sometimes called neighbors of each other, and all the neighbors of a given vertex x are called the neighborhood of x. The neighborhood of x is denoted by N(x). The set of edges incident to a vertex x is denoted by E(x).

One can describe a graph by giving just the list of all of its edges. For graph G, the edge list, denoted by J(G) is the following:

$$J(G) = \{\{x_1,x_2\},\{x_2,x_3\},\{x_3,x_4\},\{x_4,x_5\},\{x_1,x_5\},\{x_2,x_5\},\{x_2,x_4\}\}.$$

A loop is an edge connecting a vertex to it-self. If a vertex has no neighbors, i.e. its degree is 0, then these vertices are said to be isolated. If there are many edges connecting the same pair of vertices, then these edges are called 'parallel' or 'multiple'. A simple adjacency between vertices occurs when there is exactly one edge between them.

In a graph, an ordered pair of vertices is called an 'arc'. If (x,y) is an arc, then x is called the initial vertex and y is called the terminal vertex. A graph in which all edges are ordered pairs is called the 'directed graph', or 'digraph'.

Graphs in which order is not important are called 'undirected graphs'. Undirected graphs without loops and multiple edges are called 'simple graphs' or just simply 'graphs'.

A graph in which all vertices can be numbered x1,x2, . . . , xn in such a way that there is precisely one edge connecting every two consecutive vertices and there are no other edges, is called a 'path', while the number of edges in a path is the 'length'.

A graph is called 'connected' if in it any two vertices are connected by some path; otherwise it is called 'disconnected'. It means that in a disconnected graph there always exists a pair of vertices having no path connecting them. Any disconnected graph is a union of two or more connected graphs; each such connected graph is then called a 'connected component' of the original graph. A 'cycle' is a connected graph in which every vertex has degree 2. It is denoted by Cn where n is the number of vertices.

A simple adjacency between vertices occurs when there is exactly one edge between them. A graph in which every pair of vertices is an edge, is called 'complete', denoted by Kn whereas usually, n is the number of vertices. It is complete because we can't add any new edge to it and obtain a simple graph.

If we have a graph G = (X,E) and a vertex x ϵ X. The deletion of x from G means removing x from set X and removing from E all edges of G that contain x. However, the deletion of an edge is easier than that of the vertex, as it comprises only removing the edge from the list of edges.

Let G = (X,E) be a graph, $x,y \in X$. The distance from x to y, denoted by $d(x,y)$, is the length of the shortest (x,y)-path. If there is no such path in G, then $d(x,y) = \infty$. In this case, G is disconnected and x and y are in different components.

The diameter of G denoted by $diam(G)$ is $\max_{x,y \in X} d(x,y)$, which means it is the distance between the farthest vertices.

A graph $G = (X,D)$ is called 'weighted' if each edge $D \in D$ is assigned a positive real number $w(D)$ called the weight of edge (D). In many practical applications, the weight represents a distance, cost, time, capacity, probability, resistance, etc.

In a graph G, a walk is an alternating sequence of vertices and edges where every edge connects preceding and succeeding vertices in the sequence. It starts at a vertex, ends at a vertex and has the following form: $x_0,e_1,x_1,e_2, \ldots, e_k,x_k$.

A digraph $N = (X,A)$ is called a 'network', if X is a set of vertices (also called nodes), A is a set of arcs, and to each arc $a \in A$ a non-negative real number $c(a)$ is assigned which is called the capacity of arc a. For any vertex $y \in X$, any arc of type (x,y) is called 'incoming, and every arc of type (y, z) is called outcoming.

A digraph is (weakly) connected if its underlying graph is connected. A digraph is strongly connected if from each vertex to each other vertex there is a directed walk.

A cut-vertex (or cutpoint) is a vertex whose removal increases the number of components. A cut-edge is an edge whose removal increases the number of components [32].

IV. Case Study: BX-Dataset Using SNA

A. Data Description

Our dataset, which is available for free download from the internet, has two types of file extension: the (.sql) format and the (.csv) format. Three files are extracted when dealing with the second type of data files: BX-Books, BX-Users and BX-Book-Ratings. The BX-Books file contains information about the books available in the website database. The BX-Users file contains demographic information about registered users, namely location and age. The BX-Book-Ratings file contains the relational data that connect between users and rated items, in addition to the weight of the relationship (expressed as a numerical value on a scale from 0 to 10).

The BX dataset was collected in a 4-week crawl (August/September 2004) by [40] from the Book Crossing, a community where users around the world exchange information about books.

The dataset contains 1,149,780 implicit and explicit ratings on a scale from 0 to 10. Implicit ratings are expressed by 0 on the scale and constitute 716,109 ratings. The remaining 433,681 ratings are regarded as explicit ratings across 1 to 10 on the scale. The total number of users is 278,858 and of the books is 271,379 [30].

Ref. [14] suggest that BX dataset also contains many more implicit preferences, like when users buy books but they do not explicitly rate them, which gives a positive indication towards those books.

BX dataset suffers, like any other public dataset, from a number of drawbacks such as low density of user ratings; a problem makes predictions so noisy in that context. This issue was treated by other researchers through taking only a subset of the BX-dataset [4]. The demographic information contains what it looks erroneous and incomplete data. Also, if the dataset were to have more demographic information (such as gender or occupation) we would have had more deep understanding of users' preference.

Ref. [39] has discretized the BX-dataset into five general domains (based on content):

TABLE I. BOOK DOMAINS IN BX-DATASET

Domain #1	Domain #2	Domain #3	Domain #4	Domain #5
Mystery and Thrillers	Science Fiction and Fantasy	Science	Business and Investing	Religion and Spirituality

B. Data Pre-processing

Removing implicit ratings (those with value=0 on the scale) was necessary since implicit ratings are written reviews rather than numerical values. So, from the original dataset which comprised 1,149,780 ratings, we are left now only with 433,659 ratings (i.e. on a scale from 1 to 10).

C. Software

The specific software which we used in our analysis was Pajek, a program for analysis and visualization of large networks [1]. Several reasons stood behind the use of this software: Pajek is capable of dealing with large networks (several hundred thousand and even millions of nodes), a task not every program can handle successfully. It is freely available to download from the internet. It has a simple GUI, which gives the space for machine resources to function easily and efficiently. It has a well-illustrated user's manual and a lot of free compatible datasets for testing purposes. It has powerful visualization tools and several data analytic algorithms. It has the ability to deal with different types of networks and many networks at the same time. Also, Pajek has the ability to engage with very powerful statistical analysis tools (R and SPSS). The software release we used was 2.05.

D. Two-Mode Network Analysis

A two-mode network data contain measurements on which actors from one of the sets have ties to actors in the other set. Actors in one of the sets are senders, while those in the other are receivers [33]. Examples of two-mode networks include corporate board management, attendance at events, membership in clubs, participation in online groups, membership in production teams and even course-taking patterns of high school students [2].

1) Mother Network Analysis

The first network that we analyzed was the mother network (a name we used to describe the network that covers the entire scale of ratings, i.e. from 1 to 10).

Analyzing this network helped us answering the question: which users have made the highest number of ratings (most active users)? We were also able to answer the question: which books obtained the highest number of ratings (no matter whether they were negative or positive)? Let's take a look at some of the overall statistics, evaluated using Pajek:

TABLE II. OVERALL STATISTICS OF THE MOTHER NETWORK

Metric	Value
Graph Type	Directed
Dimension	263631
Number of Arcs	433660
Network Density	0.00000624
Number of Loops	0
Number of Multiple Lines	0
Average Degree	3.28990142
Connected Components	14684
Single-Vertex Connected Component	0
Maximum Vertices in a Connected Component	229036 (86.877%)

It is a directed two-mode network with density equals 0.00000624, which is very low. Network dimension is 263631 and the number of ties is 433660 (the more number of nodes in a network, the less network density). The network has neither loops nor multiple lines and the average degree is 3.28990142. The number of connected components is 14684, which is very high (due to the high dispersion in users' choices) and the largest component consists of 229036 nodes.

The network has no isolated vertices. The importance of identifying the largest component (also called giant component) in a community is that it helps measuring the effectiveness of the network at doing its job [22].

The highest and lowest out-degrees and out-degree centralization values of the mother-network were as follows:

TABLE III. HIGHEST AND LOWEST OUT-DEGREES AND OUT-DEGREE CENTRALIZATION OF THE MOTHER NETWORK

Metric	Value	Frequency
Highest output degree value	8522	1
Lowest output degree value	1	45375
Network out-degree Centralization	0.03231949	-

We can see that only one node obtained the highest number (8522) of outgoing ties (most active user) from among 263631 nodes, and that 45375 other nodes (approximately 1/6 of network nodes) supplied only 1 vote (least active users). The analysis also gave us 185833 nodes with zero out-degree (not shown in the table above). This is because Pajek analyzed both types of nodes, namely users and books, and that the nodes with out-degree=0 represent books (destination of relation). The highest ten out-degree values (representing most active users) of the mother- network were as follows:

TABLE IV. HIGHEST TEN OUT-DEGREE VALUES (MOST ACTIVE USERS) IN THE BX-DATASET

Rank	Out-Degree	Normalized Out-Degree	User ID	Age	Country
1.	8522	0.0323	11676	Null	N/A
2.	5802	0.0220	98391	52	USA
3.	1969	0.0075	153662	44	USA
4.	1906	0.0072	189835	Null	USA
5.	1395	0.0053	23902	Null	UK
6.	1036	0.0039	76499	Null	USA
7.	1035	0.0039	171118	47	Canada
8.	1023	0.0039	235105	46	USA
9.	968	0.0037	16795	47	USA
10.	948	0.0036	248718	43	USA

Some users have higher out-degree values than others since they have provided a higher number of book ratings; in other word they are more active than their associates. We can see that 70-80% of the people whose outgoing links were probed were from USA, and that the average user age (when the data was crawled) was between 40s and 50s, which gives an indication that older people are more interested in book reading when compared to young ones. Also, it looks that people from USA do more social activities than people from other countries. The same point was pointed out by [19]. In addition to the out-degree measure, we evaluated the in-degree measure. The highest and lowest in-degrees and in-degree centralization values of the mother-network were as follows:

TABLE V. HIGHEST AND LOWEST IN-DEGREES AND IN-DEGREE CENTRALIZATION OF THE MOTHER NETWORK

Metric	Value	Frequency
Highest input degree value	707	1
Lowest input degree value	1	129480
Network in-degree Centralization	0.00267556	-

We can see that only one node has acquired the highest number of incoming arcs (in-degree) from among 263631 nodes, and that 129480 other nodes acquired only 1 incoming arc.

We can see that nodes (which gained only 1 vote from users for each) represent about half the mother-network. The analysis also gave us 77798 nodes with zero incoming ties (not shown in the table above). This is because the analysis comprised both types of nodes, namely users and books, and nodes with in-degree=0 represent users (source of relation). We can determine the ten books that obtained the highest number of ratings (over the entire rating scale) as follows:

TABLE VI. HIGHEST TEN IN-DEGREE VALUES (REPRESENTING THE BOOKS THAT OBTAINED THE HIGHEST NUMBER OF RATINGS) IN THE BX-DATASET

Rank	In-degree	Normalized in-degree	ISBN	Book Title
1.	707	0.0027	0316666343	The Lovely Bones
2.	581	0.0022	0971880107	Wild Animus
3.	487	0.0018	0385504209	The Da Vinci Code
4.	383	0.0015	0312195516	The Red Tent (Bestselling Backlist)
5.	333	0.0013	0679781587	Memoirs of a Geisha*
6.	320	0.0012	0060928336	Divine Secrets of the Ya-Ya Sisterhood
7.	315	0.0012	059035342x	Harry Potter and the Sorcerer's Stone (Harry Potter (Paperback))
8.	307	0.0012	0142001740	The Secret Life of Bees
9.	295	0.0011	0446672211	Where the Heart Is (Oprah's Book Club (Paperback))
10.	282	0.0011	044023722x	A Painted House

The novel 'The lovely bones' has occupied position #1. This is due to the fact that it gained the highest number of users' evaluation and attention. Other books information was taken from the dataset. However, for the ISBN in position 5, we did not find the corresponding information so; we took help from Amazon.com to get the book title and other information. This is an example of the bugs existing in this dataset.

2) User-Preference Network Analysis

This network comprises ratings of users who have rated items with values from 6 to 10 on the scale.

The basic idea behind the formation of this network is that our interest is to know whether a user recommends reading/buying a book or not, which means constructing a network of 'likes' and 'dislikes' [17], [34]. However; [19] considered only ratings with 7 or more on the rating scale as positive. Analyzing the network helped us answering the question: which books were most positively-rated (most popular books)?

Let's have a look at some overall statistics of the user-preference network:

TABLE VII. OVERALL STATISTICS OF THE USER-PREFERENCE NETWORK

Metric	Value
Graph Type	Directed
Dimension	228970
Number of Arcs	363258
Network Density	0.00000693
Number of Loops	0
Number of Multiple Lines	0
Average Degree	3.17297463
Connected Components	13979
Single-Vertex Connected Component	0
Maximum Vertices in a Connected Component	196180 (85.679%)

It is a two-mode network consisting of 228970 nodes and 363258 arcs with no edges, since it is a relationship between a user and the book that he/she evaluates. Even though the network density is low (0.00000693), it is still higher than the mother network. This is because the current network has a less number of nodes, as the largest the number of nodes is, the lowest the density. The largest component in this network occupies about 85.679% of the total size of the network. The highest and lowest in-degree values and the network in-degree centralization were as follows:

TABLE VIII. HIGHEST AND LOWEST IN-DEGREES AND IN-DEGREE CENTRALIZATION OF THE USER-PREFERENCE NETWORK

Metric	Value	Frequency
Highest input degree value	663	1
Lowest input degree value	1	112010
Network in-degree centralization	0.00288867	

We can see that nearly half of the user-preference network nodes (i.e. 112010 nodes) obtained only 1 vote, and that only one node obtained the highest number of votes, namely 663.

We can also calculate the highest ten in-degree values (representing most popular books) as follows:

TABLE I. TOP TEN MOST POPULAR BOOKS IN THE BX-DATASET

Rank	In-degree	Normalized in-degree	ISBN	Book Title
1.	663	0.0029	0316666343	The Lovely Bones
2.	452	0.0020	0385504209	The Da Vinci Code
3.	344	0.0015	0312195516	The Red Tent (Bestselling Backlist)
4.	307	0.0013	0679781587	Memoirs of a Geisha
5.	305	0.0013	059035342x	Harry Potter and the Sorcerer's Stone (Harry Potter (Paperback))
6.	292	0.0013	0142001740	The Secret Life of Bees
7.	285	0.0012	0060928336	Divine Secrets of the Ya-Ya Sisterhood
8.	274	0.0012	0446672211	Where the Heart Is (Oprah's Book Club (Paperback))
9.	260	0.0011	0452282152	Girl with a Pearl Earring
10.	250	0.0011	0671027360	Angels & Demons

The table above lists the ten most popular books. The more in-degree value is, the more prestigious the book. With this metric, we can say that the most preferred (popular) book (at the time when the data was crawled) by users was "The Lovely Bones: A novel".

3) User Non-Preference Network Analysis

The third network that we analyzed was the user non-preference network. It comprised users who have rated books with values from 1 to 5 on the rating scale. Analyzing the network helped us answering the question: which books were most negatively-rated (most un-popular books)?

TABLE IX. OVERALL STATISTICS OF THE USER NON-PREFERENCE NETWORK

Metric	Value
Graph Type	Directed
Dimension	73716
Number of Arcs	70403
Network Density	0.00001296
Number of Loops	0
Number of Multiple Lines	0
Average Degree	1.91011449
Connected Components	10865
Single-Vertex Connected Component	0
Maximum Vertices in a Connected Component	45008 (61.056%)

It is a two-mode network consisting of 73716 nodes and 70703 arcs with no edges or loops. Network Density = 0.00001296 which is very low (however, it is still higher than the two previous networks since this network has only 73716 nodes). We notice that the number of nodes here exceeds the number of arcs, which indicates users' less interest to evaluate books if they did not like. The number of connected components and the average degree are less than its two previous networks (Tables II, VII). It has less average degree value because the number of arcs here is less than the number of nodes.

The highest and lowest in-degree values and the network in-degree centralization were as follows:

TABLE X. HIGHEST AND LOWEST IN-DEGREES AND IN-DEGREE CENTRALIZATION OF THE USER NON-PREFERENCE NETWORK

Metric	Value	Frequency
Highest input degree value	389	1
Lowest input degree value	1	41447
Network Input Degree Centralization	0.00526420	-

More than half of the network nodes (books) obtained only 1 vote for each, while the highest in-degree value in the user-non preference network was 389, which means that the corresponding book was rated by the users as the most unpopular book.

By implementing the in-degree measure, we get the following ten results which represent most unpopular books:

TABLE XI. TOP IN-DEGREE VALUES (REPRESENTING MOST UNPOPULAR BOOKS) OF THE USER NON-PREFERENCE NETWORK

Rank	In-degree	Normalized in-degree	ISBN	Book Title
1.	389	0.0053	0971880107	Wild Animus
2.	51	0.0007	044023722x	A Painted House
3.	44	0.0006	0316666343	The Lovely Bones
4.	41	0.0006	0316601950	The Pilot's Wife
5.	41	0.0006	0316769487	The Catcher in the Rye
6.	39	0.0005	0312195516	The Red Tent (Bestselling Backlist)
7.	39	0.0005	0446605239	The Notebook
8.	38	0.0005	0425182908	Isle of Dogs
9.	36	0.0005	0140293248	The Girls' Guide to Hunting and Fishing
10.	35	0.0005	0375727345	House of Sand and Fog

We notice that two of the books in the table above (positions 3 and 6) have also been seen in the user-preference network (Table I). This may reflect the fact that users' choices covered a wide range of ratings over a scale (from 1 to 10), and that peoples' opinions towards these books largely scattered between "good" and "bad".

E. Affiliation Network Analysis

The term Affiliation refers to membership or participation data such as when we have data on which actors have participated in which events. It can be represented as a bipartite graph (*V1, V2, E*), where *V1* and *V2* are two different sets of nodes, while *E* is an affiliation relation between elements of *V1* and *V2* [2]. Usually, we can extract two one-mode networks from one a two-mode network as follows: the first one is the network of interlocking events (if two books share the same event i.e. being read by the same two or more readers) and the second one is the network of actors (if two users or more like the same books). The idea behind inducing co-affiliation network from affiliation network is that a co-affiliation network provides the ground for the development of social relationships between the actors of one set. For example, the more the number of times people come at the same event, the more likely those people are going to interact and develop some type of relationship. It has been reported that persons whose activities are focused around the same point, frequently become connected over time.

1) User-User Network Analysis

For the purpose of affiliation network analysis, we made use of the lately generated user-preference network to generate this new network which will help us later on probing the potential social relations among users. It is a network with connections between users only.

We restricted ourselves here to extract this network from the user-preference network (rather than other networks), because what makes people develop friendships depends mainly on the things they share and the things they like.

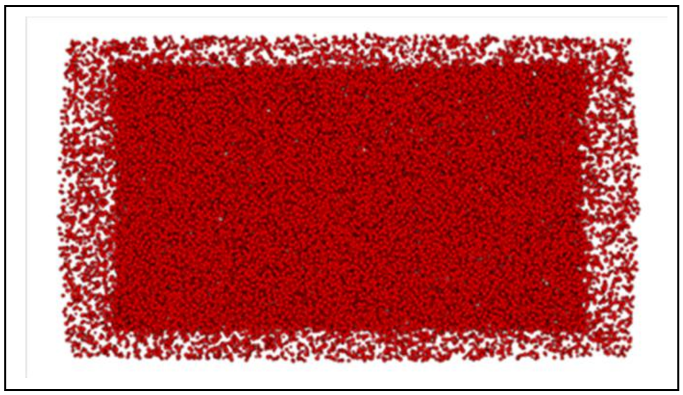


Figure 1. 2-D representation of the user-user network.

Figure (1) gives a 2-D representation of the user-user network. The network was energized using Fruchterman-Reingold algorithm [9]. Edges and vertex labels have been eliminated. Nodes in the middle are the core nodes, while nodes around the core are the periphery nodes.

Some overall statistics of the user-user network are as follows:

It is a one-mode undirected sub-network consisting of 69768 vertices and 3176585 weighted edges. Network density =0.001305 which is higher than the earlier networks. This is because a 1-mode network has higher density than its equivalent a 2-mode network since in 1-mode network; vertices can have ties with any other nodes in the network, while this is not true for 2-mode networks.

The results showed that for (n>=2), the network consisted of 883 components. The size of the largest component is 54701

(78.404%), while the size of the next largest component is 13096 (18.770%, not shown here) and the rest of components constitute approximately 10% of the network. Network diameter, which is the longest shortest path in the network, is 10. This geodesic distance exists only between two users, namely 150578 and 112131. The first guy is 43 years old from Milano, Italy while the other guy is 12 years old from Sydney, Australia. This could be due to variation in age and the geographic locations of both.

We can see that at this time, the network not only having connected components of two or more vertices, but also having single-vertex connected components = 13096, which means that it includes 'isolates'. This is because the user-user sub-network emerges from a larger network, namely the user-preference network, which already contains books having in-degree value=1. When extracting a one-mode subnetwork from a two-mode network, these nodes become 'isolates'. The network has 1875356164 unreachable pairs, which expresses the number of pairs of nodes that do not have a connection between them.

TABLE XII. OVERALL STATISTICS OF THE USER-USER NETWORK

Metric	Value
Graph Type	Undirected
Dimension	69768
Number of Edges	3176585
Network Density	0.00130522
Number of Loops	0
Number of Multiple Lines	0
Connected Components	883
Single-Vertex Connected Component	13096
Maximum Vertices in a Connected Component	54701(78.404%)
Maximum Geodesic Distance (Diameter)	10
Average Geodesic Distance (Among Reachable Pairs)	2.80782
Average Degree	91.06137484
Number of Unreachable Pairs	1875356164

2) Applying Centrality Measures

We want to infer the most potential central people in the user–user network. So, we are going to implement the three measures of centrality, namely degree, closeness and betweenness centrality measures. Research has proved that these three measures are highly correlated and give similar results in identifying most important actors in a network [3]. The importance behind identifying most important actors is that it reflects how active an actor is. Also, active actors are more likely to establish social ties with a large number of other actors and can affect how the network works.

First, we are going to find top-degree centrality users using the in-degree measure. Figure (1) was built based on node (circle) size. The larger the node is, the more central a user in the network in regard to degree centrality. The user with the highest degree centrality was #11676. However, we didn't find any demographic information related to him/her, as it seems he/she preferred to keep identification information dim. That guy has already occupied position #1 in terms of people with the highest number of outgoing ties in the mother-network (Table IV). That guy has the largest potential social network, as he/she is connected in a direct path to 24026 other actors (neighbors) in the network, which means that he/she shares common opinions about a specific number of book(s) with other 24026 users in the user-user network. This high number of connections reflects the fact that a 1-mode network has a higher density than a 2-mode network as nodes can freely connect to any other nodes in the same network.

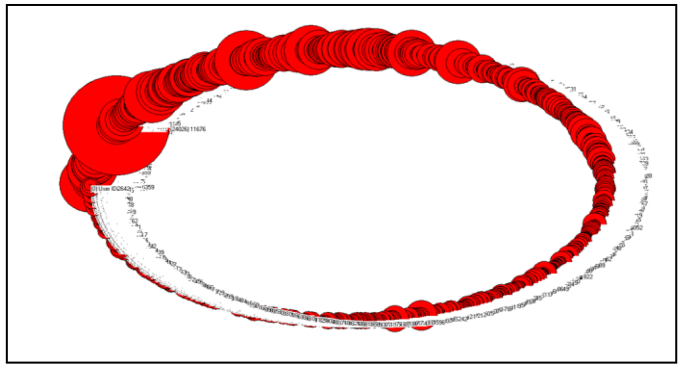


Figure 2. Circular 2-D representation of the degree-centrality measure in the user-user network

Degree centrality statistics of the user-user network were as follows:

TABLE XIII. DEGREE CENTRALITY STATISTICS OF THE USER-USER NETWORK

Metric	Value
Dimension	69768
Highest degree centrality value	24026
Lowest degree centrality value	0
Network Input Degree Centralization	0.34307946

We can see that we have nodes with degree centrality =0 because these are 'isolates'. Highest ten degree centrality values in the user-user network were as follows:

TABLE XIV. HIGHEST TEN DEGREE CENTRALITY VALUES IN THE USER-USER NETWORK

Rank	User ID	Degree Centrality	Demographic Info
1.	11676	24026	N/A
2.	16795	8614	Mechanicsville, Maryland, USA, 47 Years Old
3.	95359	8110	Charleston, west Virginia, USA, 33 Years Old
4.	60244	6493	Alvin, Texas, USA, 47 Years Old
5.	204864	6104	Simi valley, California, USA, 47 years Old
6.	104636	5533	Youngstown, Ohio, USA
7.	98391	5480	Morrow, Georgia, USA, 52 years old
8.	35859	5409	Duluth, Minnesota, USA
9.	135149	5283	ft. Pierce, Florida, USA
10.	153662	5281	ft. Stewart, Georgia, USA, 44 years old

The user in position #2, namely user ID 16795 (47 years old of Maryland, USA), has the second largest potential social network consisting of 8614. However, he/she came only in rank #9 in a previous statistics about users with the highest number of outgoing ties (Table IV). This might mean that even though that guy had fewer number of outgoing ties than the other eight guys, his/her choices were more focused and that he/she could share book preferences with more other people, which makes him/her more candidate to establish social relations than others (of course behind our top-user). The second measure of centrality, we use here, is closeness. The concept of closeness centrality depends on the total distance between one vertex and all other vertices, as large distances show lower closeness centrality. Closeness centrality values range from 0 (for isolated vertices) to 1. For a specific vertex, it results from the number of all other vertices in the network divided by the sum of distances between that vertex and all other vertices in the network. Therefore, closeness centrality values are continuous rather than discrete [5].

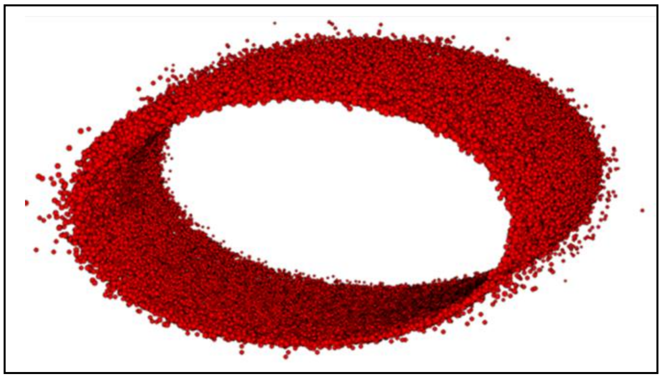


Figure 3. Circular 3-D representation for the closeness centrality measure of the user-user network. All nodes are equally sized

It took about 15 hours to calculate closeness centrality values of all vertices in the network however; it depends mainly on the device specifications. Some overall statistics for closeness centrality are as follows:

TABLE XV. OVERALL STATISTICS FOR CLOSENESS CENTRALITY MEASURE IN THE USER-USER NETWORK

Metric	Value
Dimension	69768
Highest closeness centrality value	0.4926
Lowest closeness centrality value	0.0000
Arithmetic mean	0.2233
Median	0.2661
Standard deviation	0.1220
Network closeness centralization cannot be computed since the network is weakly connected	-

The closeness centrality values of the first ten actors were as follows:

TABLE XVI. CLOSENESS CENTRALITY VALUES OF THE FIRST TEN ACTORS IN THE USER-USER NETWORK

Rank	Closeness Centrality	User ID	Demographic Info
1.	0.4926	11676	N/A
2.	0.4021	16795	Mechanicsville, Maryland, USA, 47 years old
3.	0.4011	95359	Charleston, west Virginia, USA, 33 years old
4.	0.3919	60244	Alvin, Texas, USA, 47 years old
5.	0.3906	204864	Simi valley, California, USA, 47 years old
6.	0.3862	35859	Duluth, Minnesota, USA
7.	0.3860	135149	ft. Pierce, Florida, USA
8.	0.3852	104636	Youngstown, Ohio, USA
9.	0.3850	153662	ft. Stewart, Georgia, USA, 44 years old
10.	0.3838	98391	Morrow, Georgia, USA, 52 years old

It is easy to notice that the user (id=11676) is the top-closeness centrality user. This is mainly true because he/she is the top out-degree user (Table: IV), and the top in-degree user (Table: XIV). The rest of actors in the table also appeared in the study, which reflects their importance at the social level, alongside the ultimate importance of the top-user (namely user ID= 11676). Network closeness centralization cannot be computed if the network was not strongly connected since there are no paths between all vertices so; it is impossible to compute the distances between some vertices [5].

While degree and closeness centrality are based on the concept of the reachability of a person, betweenness centrality is based on the idea that a person is more important if he/she was more intermediary in the network. The more a person is a go-between, the more central her/his position in that network. This reflects the importance of a person being in the middle of social communications of a network and to what extent he/she is needed as a link in the chains of contact in the society. On the other hand, a vertex has betweenness centrality = 0 if it was not located between any other vertices in the network, which points out to a weak social role that he/she plays. Many vertices may not appear in the figure below (Figure: 4) because they do not mediate between any two vertices, so their betweenness centralities equal zero. The drawing was built based on node size. The larger the node (circle) is, the more central the user in the network, in regard to betweenness centrality concept.

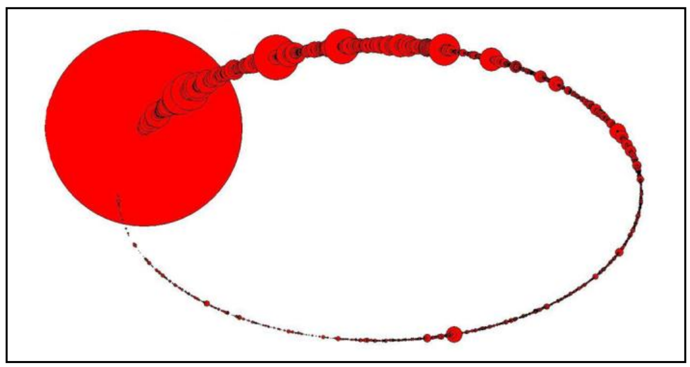


Figure 4. Circular representation for betweenness-centrality measure of the user-user network

It took about 15 hours to calculate all betweenness centrality values. Some overall statistics of the betweenness centrality measure were as follows:

TABLE XVII. SOME OVERALL STATISTICS OF THE BETWEENNESS CENTRALITY MEASURE IN THE USER-USER NETWORK

Metric	Value
Dimension	69768
Highest betweenness centrality value	0.1735
Lowest betweenness centrality value	0.0000
Arithmetic mean	0.0000
Median	0.0000
Standard deviation	0.0007
Network betweenness centralization	0.17345915

The betweenness centralities of the first ten actors in the network were as follows:

TABLE XVIII. TOP BETWEENNESS CENTRALITY MEASURE VALUES IN THE USER-USER NETWORK

Rank	Betweenness Centrality	User ID	Demographic Info
1.	0.1735	11676	N/A
2.	0.0121	98391	Morrow, Georgia, USA, 52 Years Old
3.	0.0094	16795	Mechanicsville, Maryland, USA, 47 Years Old
4.	0.0085	95359	Charleston, west Virginia, USA, 33 Years Old
5.	0.0065	153662	ft. Stewart, Georgia, USA, 44 Years Old
6.	0.0055	204864	Simi valley, California, USA, 47 years old
7.	0.0055	60244	Alvin, Texas, USA, 47 years old
8.	0.0053	23902	London, England, United Kingdom
9.	0.0047	135149	ft. Pierce, Florida, USA
10.	0.0045	104636	Youngstown, Ohio, USA

The top-user (user id=11676) is still in rank #1 in the table which means that he/she lies at the geodesic distances between other pairs, more than any other vertex in the network. This nominates him/her (more than others) to be a candidate person to play many potential brokerage roles in the future.

As we notice, all the three measures (degree, closeness and betweenness centrality) have showed similar (not identical) results, which support the notion that all these measures collectively are used to measure most important individuals in a community

3) Ego-Network Analysis

After conducting a comprehensive analysis using some important measures in SNA, we turn our eyes to the top-user (ID = 11676) who occupied the first position in all the previous tests (Tables: IV, XIV, XVI, XVIII), and try to analyze his/her sub-network (which is called ego-network or ego-centric approach as opposed to the socio-centric approach).

A very useful way to understand complicated networks is to see how they arise from the local connections of individual actors.

Ego-network (which consists of ego, its neighbors and ties among them) was extracted the from the user-user network. We show below a 2-D representation of the ego-network:

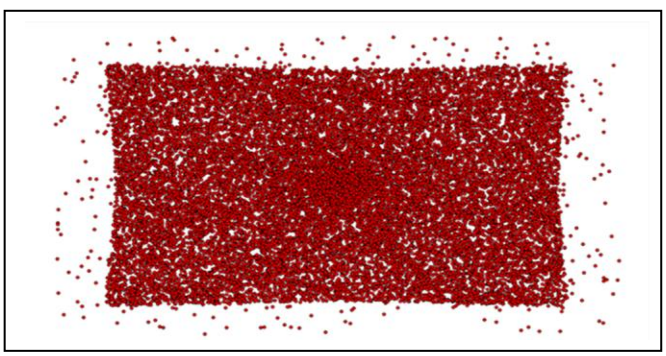


Figure 5. 2-D Representation of the ego-network, extracted from the original user-user network

Let's take a look at a short summary of some metrics, evaluated using Pajek:

TABLE XIX. SUMMARY OF THE EGO-NETWORK STATISTICS

Metric	Value
Graph Type	Undirected
No. of Neighbors	24026
Number of Edges	2278058
Ego-network Density	0.00789314
Number of Loops	0
Number of Multiple Lines	0
Maximum Geodesic Distance (Diameter)	8
Average Geodesic Distance	2.49241
Average Degree	189.63273121
Ego-network Betweenness Centralization	0.02163385

The network consists of 24026 neighbors. Those neighbors are only the direct ones, i.e. who are located at distance one from ego. Also, the number of edges is 2278058. This number represents the relations among vertices around ego.

The density of ego-network expresses the density of ties among its neighbors. The result is 0.00789314 which is relatively high and at the same time higher than the densities of our earlier networks (namely the mother, the user-preference, the user non-preference and the user-user networks), which means that ego-network is quite embedded in dense local substructure. This is because this network is the local network of the top-user (who occupied the 1st position in all the previous four tests). The ego network diameter and the average geodesic distance are a little slighter than the user-user network (Table XII). This is intuitive since the current network is a dense fragment of the user-user network.

Ego-network diameter, that is the maximum geodesic distance between two vertices, is 8. This geodesic exists between User ID=47534 (45 years from Luzern, Switzerland) and User ID=240418 (34 years old from Barcelona, Spain). Also, we notice that ego-network betweenness centralization is 0.02163385 which is lower than what it is in the user-user network (Table XVII).

This is because the variation in vertex betweenness centrality in the user-user network is higher than what it is in ego-network (Figure: 6).

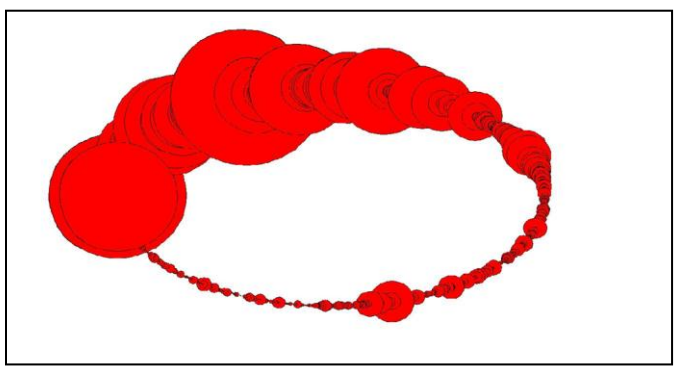


Figure 6. Ego-network betweenness centralization

We can calculate the geodesics from the top-user to all other vertices in the user-user network as follows:

TABLE XX. GEODESIC DISTANCES FROM TOP-USER TO ALL OTHER USERS IN THE USER-USER NETWORK

Cluster	Frequency
0	1
1	24026
2	29088
3	1490
4	86
5	10
Sum	54701
Unknown	15067
Total	69768

We can see that the top-user can reach 24026 users with only one hop. He/she can reach 29088 other users with two hops, 1490 others with three hops and so on. However, there are other 15067 vertices that can't be reached by ego; in other word they are unreachable in our ego-network, and hence are given the value 999999997 in Pajek's report of distances. In the ego-network, there are unreachable nodes because our user-user network is split up into smaller parts (883 connected components. Table: XII). Cluster (0) means that ego doesn't need any hop to get to that node, as it is the ego him/her-self.

We can also calculate the potential brokerage roles practiced by ego in the user-user network. Brokerage expresses the ability to induce and exploit competition between the other two actors of the triad (a triad consists of a focal person, alter and a third person in addition to the ties among them), and also expresses his/her qualifications to play a subversive role through creating or exploiting conflict between the other two actors in order to control them [5].

Brokerage can be calculated by using the 'aggregate constraint' concept which is the sum of the dyadic constraint on all of a vertex's ties. However, the aggregate constraint has an opposite effect, i.e. the more the aggregate constraint, the less the brokerage role an actor can play. The implementation gave us the distribution table of aggregate constraints. We put down here the two extremes:

TABLE XXI. THE TWO EXTREMES OF AGGREGATE CONSTRAINT IN THE USER-USER NETWORK

Aggregate Constraint	Value	Representative
Highest Value	1.3203	44726
Lowest Value	0.0007	11676

We see that the top-user (ego) has the lowest aggregate constraint in the user-user network because he/she has the highest out-degree in the mother network (Table: IV), in-degree, betweenness and closeness values in the user-user network (Tables: XIV, XVI, XVIII). In other words; he/she can perfectly play brokerage.

4) M-Slice Analysis

A one-mode network induced from a two-mode network creates the atmosphere to discover many dense structures. One way to detect cohesive subgroups in one-mode networks is to detect m-slice sub-networks. M-slice can be defined as the maximal sub-network in which line multiplicity is equal or greater than m. It was first introduced by John Scott as 'm-core'. This technique puts into consideration line multiplicity rather than the number of neighbors (which is defined by the k-core concept). M-slice method comprises allocating values to network nodes based on m-slice, i.e. the highest tie these nodes are incident (connected) with. The importance of conducting this type of analysis is that it helps us identify the strongest potential social relations in the network based on 'participation rate' between each pair of nodes. It has been found that the larger the number of interlocks between two users, the stronger their tie (or relationship) and the more similar they are [5]. We first examine the network in order to find out the distribution of tie weights, as these weights control how m-values are allocated to nodes:

TABLE XXII. DISTRIBUTION OF TIE WEIGHTS IN THE USER-USER NETWORK

I	Tie Weights	Frequency
1	36.0000	24834
2	36.0000 - 8470.3333	3151746
3	8470.3333 - 16904.6667	4
4	16904.6667 - 25339.0000	1
	Total No. of Links	3176585

The results above show that the lowest line multiplicity is 36 (achieved in 24834 ties) and the highest line multiplicity is 25339 (achieved in only 1 tie). From the m-slice frequency tabulation values of the user-user network, we display the highest five values in addition to the lowest five values:

TABLE XXIII. M-SLICE VALUES IN THE USER-USER NETWORK

M-slice	Value	Number of Nodes	Representative
Lowest five values	0	13096	-
	36	160	-
	42	525	-
	48	774	-
	49	621	-
Highest five values	25339	2	98391, 235105
	16129	1	11676
	9864	1	153662
	9466	1	16795
	7781	1	104636

The results above show that 13096 of the nodes belong to the 0-slice, which means that these nodes are not connected to any nodes in the network and that the users do not share book preference among them or with any other users. In other words; they are 'isolates'. They represent the weakest potential social components in the user-user network. In fact, they constitute no social components at all (in the context of our measures). It is not likely that those users in the future establish relationships among them by any means or of any type, since there is nothing they can gather on. We can see also that the strongest potential social component (which belongs to the 25339-slice) consists of two nodes: 98391 (52 years old from Georgia, USA) and 235105 (46 years old from Missouri, USA). This pair can formulate the most powerful, everlasting, and fast-shaping relationship.

Many reasons may stand behind that, for instance: age, occupation, level of education, environment, gender, past experience, marital status and so on.

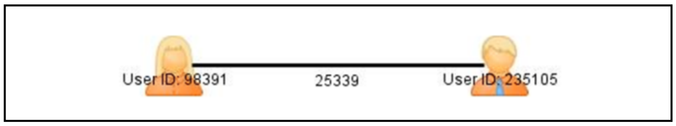


Figure 7. The strongest potential social component in the user-user network which comprises two vertices (98391 and 235105)

Using the m-slice concept, we can extract stronger and stronger subgroups by removing undesired lines and nodes that do not satisfy our goals. The process will raise the minimum m-slice threshold, which in turn forms more cohesive groups. For example, if we remove the 0-slice nodes, the resulting network will consist of 56672 and 3176585 lines. Next, we need to eliminate unnecessary lines. Thus, we obtain more cohesive components. If we keep going on that process, we will end up with the highest m-slice component, namely 25339-slice that consists of only two nodes (98391, 235105).

V. Conclusion And Future Work

The purpose of this study was to present an in-depth analysis of one of the most celebrated social sites frequently visited by individuals who are interested in exchanging information about the books they have already experienced. The research questions of interest were addressed via analyzing the relational data to detect social interaction schemes and to find out most celebrated features that characterize this community.

In order to narrow the results above, we calculated the number of appearance of each user (in each category). For example, the top-user (user id=11676) appeared in four places so, he/she is in category #4 (A category represents the number of appearance for each user within the list of ten top-users). Also, he/she has occupied the first positions in all these four tests so; he/she gets four points (by multiplying 1*4). This user is in a better position as compared to his subsequent fellow (in the below table), namely user id=16795, who belongs to category #4 also but obtained 16 points (9+2+3+2=16). As a rule of thumb, we shall suppose: the less the number of points is, the higher the rank of a user. The overall top ten users within the Book-Crossing (at the time when that data was crawled) were as shown in table XXIV.

The results show that 8 to 9 of actors were from USA and that only 1 to 2 of actors was from a country rather than USA, namely United Kingdom. The results also show that almost half of the actors were in 40s. The lack of more demographic information has stopped us from knowing more about the implications behind users' choices. For the books that earned the highest number of ratings, whether they were negative or positive (from 1 to 10 on the rating scale), we obtained the results showed in Table: (VI). Although these books obtained a large number of users' evaluations, they are not necessarily considered the most preferable books to users. We can say these books took a wide range of users' interest, and that users had different impressions about these books which in turn pushed them to take different perspectives.

TABLE XXIV. TOP TEN USERS WITHIN THE BOOK-CROSSING DATASET

Rank	User ID	Category	Points	Demographic information
1.	11676	4	4	Null
2.	16795	4	16	Mechanicsville, Maryland, USA, 47 Years Old
3.	98391	4	21	Morrow, Georgia, USA, 52 years old
4.	153662	4	27	ft. Stewart, Georgia, USA, 44 years old
5.	95359	3	10	Charleston, west Virginia, USA, 33 years old
6.	60244	3	15	Alvin, Texas, USA, 47 years old
7.	204864	3	16	Simi valley, California, USA, 47 years old
8.	104636	3	24	Youngstown, Ohio, USA
9.	135149	3	25	ft. Pierce, Florida, USA
10.	23902	2	13	London, England, United Kingdom

The books that earned the highest number of positive ratings (from 6 to 10 on the rating scale) were showed in Table: (I). Eight of these books also appeared within the list of the books that earned the highest number of ratings (Table: VI) and that some of them have become the story of cinema movies (e.g. the Da Vinci Code). Also, the author "Dan Brown" had two books within this list, namely the book in position #2 and in position #10. This may reflect his significance as a key author in the world of books.

For the books that earned the highest number of negative ratings (i.e. the unpopular books), we obtained the results showed in Table: (XI). We can see that 2 of these books also appeared in the list of books that earned the highest number of positive ratings (namely books in position #3 and position #6). This gives an indication that users' opinions towards these books scattered across the entire scale and that people were inconsistent about them. The research also took us to dig out the most powerful and the weakest social relationships within the hypothetical user-user network by using m-slice type of analysis (Table: XXIII). We can see that the weakest relationships have weight=0, which means that these entities represent isolated nodes. The number of weakest relationships is 13069 relationships (nodes). Also, the strongest potential relationship has a weight=25339 and that only one entity represents this relationship, which exists between two nodes (98391 & 235105). The research methodology of this study can be further extended to other online social networks rather than the book-Crossing community. Any website where people are able to rate items on a specific scale (e.g. from 1 to 5 or 10) will be a good place to induce potential social relations from that community. Many websites, these days, give the space for their visitors to rate the materials they have bought or only checked. We can build map of user preferences which will help us further predict user behaviors and even give recommendations to similar associates (based on either most important people in the network or through the help of m-slice analysis). We can further make the process more autonomous and develop an agent that can automatically visit a specific website and recursively extract huge amount of data (maybe bigger than the current one). But we should keep in mind at the same time preserving user privacy.

REFERENCES

[1] Batagelj, V. & Mrvar, A. (2003). Pajek: Analysis and visualization of large networks. Graph Drawing Software, 77-103.

[2] Borgatti, S.P. and Halgin, D. (2011). Analyzing Affiliation Networks. In Carrington, P. and Scott, J. (Eds.), Handbook of Social Network Analysis. SAGE Publications.

[3] Cheng, B. (2006). Using social network analysis to investigate potential bias in editorial peer review in core journals of comparative/international education (Doctoral dissertation).

[4] Cortes, C., Mohri, M. & Rastogi, A. (2007). An alternative ranking problem for search engines. Proceedings of the 6th International Conference on Experimental Algorithms, 4525, 1-22.

[5] De Nooy, W., Mrvar, A. & Batagelj, V. (2005). Exploratory social network analysis with Pajek. New York, NY: Cambridge University Press.

[6] Doshi, L., Krauss, J., Nann, S. & Gloor P. (2009). Predicting movie prices through dynamic social network analysis. Elsevier Science Journal, 2(4), 6423-6433.

[7] Ereteo, G. (2011). Semantic social network analysis (Doctoral dissertation).

[8] Freeman, L. C. (2004). The development of social network analysis: A study in the sociology of science. South Carolina: BookSurge, LLC.

[9] Fruchterman, T. M. J. & Reingold, E. M. (1991), Graph drawing by force-directed placement. Software: Practice and Experience, 21, 1129–1164.

[10] Gardy, J. L., Johnston, J. C., Sui, S. J. H., Cook, V. J., Shah, L., Brodkin, E., Rempel, S., Moor, R., Zhao, Y., Holt, R., Varhol, R., Birol, I., Lem, M., Sharma, M. K., Elwood, K., Jones, S. J. M., Brinkman, F. S. L., Brunham, R. C. & Tang, P. (2011). Whole-Genome sequencing and social network analysis of Tuberculosis outbreak. The New England Journal of Medicine, 364(8), 730-739.

[11] Gross, J. L. & Yellen, J. (Eds.). (2004). Handbook of Graph Theory. New York, NY: CRC Press.

[12] Hatala, J. P. (2006). Social network analysis in human resources development: A new methodology. Human Resources Development Review, 5(1), 45-71.

[13] Hawe, P., Webster, C. & Shiell, A. (2004). A glossary of terms for navigating the field of social network analysis. J Epidemiol Community Health, 58, 971-975.

[14] Huang, B. & Jebara, T. (2010). Exact graph structure estimation with degree priors. International Conference on Machine Learning and Applications, 111–118.

[15] Jyun-Cheng, W. & Chui-Chen, C. (2010). Recommending trusted online auction sellers using social network analysis. Expert Systems with Applications 34, 1666–1679.

[16] Kramer, T., Hildenbrand, T. & Acker, T. (2009). Enabling social network analysis in distributed collaborative software development. Software Engineering (Workshops), 255-266.

[17] Lowd, D. & Davis, J. (2010). Learning Markov network structure with decision trees. Proceedings of the 10th IEEE International Conference on Data Mining (ICDM).

[18] Lu, Y., Polgar, M., Luo, X. & Cao, Y. (2010). Social network analysis of a criminal hacker community. Journal of Computer Information Systems, 51(2), 31.

[19] Lucas, J. P., Segrera, S. & Moreno, M. N. (2008). Comparing the use of traditional and associative classifiers towards personalized recommendations. In Proceedings of SEKE'2008, 607-612.

[20] Meneely, A., Williams, L., Snipes, W. & Osborne J. (2008). Predicting failures with developer networks and social network analysis. Proceedings of the 16th ACM SIGSOFT International Symposium on Foundations of software engineering ACM.

[21] Mika, P. (2007). Social networks and the semantic web. New York, NY: Springer.

[22] Newman, M. E. J. (2003). The structure and function of complex networks. SIAM Review 45, 167-256. arXiv:cond-mat/0303516v1.

[23] Otte, E. & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. Journal of Information Science, 28(6), 441-453.

[24] Park, D. H., Kim, H. K., Choi, Y. & Kim, J. K. (2011). A review and classification of recommender system research. International Conference on Social Science and Humanity, 5, V1290-V1294.

[25] Penuel, W. R., Sussex, W., Korbak, C. & Hoadley, C. (2006). Investigating the potential of using social network analysis in educational evaluation. American Journal of Evaluation, 27(4), 437-451.

[26] Perisse, A. R. S. & Nery, J. A. da C. (2007). The relevance of social network analysis on the epidemiology and prevention of sexually transmitted diseases. Cad. Saude Publica, 3, 5361-5369.

[27] Reffay, C. & Chanier, T. (2003). How social network analysis can help to measure cohesion in collaborative distance-learning. Designing for change in networked learning environments In international conference on computer support for collaborative learning, 343-352.

[28] Scott, J. (2000). Social Network Analysis: A handbook (2nd). London: SAGE publications, Ltd.

[29] Stockman, F. N. (2004). What binds us when with whom? Content and structures in social network analysis. Extended version of keynote at the SUNBELT XXIV. International social network conference, Portoroz (Slovenia).

[30] Sun, H., Peng, Y., Chen, J., Liu, C. & Sun, Y. (2011). A new similarity measure based on adjusted Euclidean distance for memory-based collaborative filtering. Journal of Software, 6(6).

[31] Tang, L. & Liu, H. (2010). Graph mining applications to social network analysis. The Kluwer International Series on Advances in Database Systems, 40, 487-513.

[32] Voloshin, V. I. (2009). Introduction to graph theory. New York, NY: Nova Science Publishers, Inc.

[33] Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications. Cambridge: Cambridge University Press.

[34] Weng, L., Zhang, Y., Zhou, Y., Yang, L. T., Tian, P. & Zhong, M. (2009). A joint web resource recommendation method based on category tree and associate graph. Journal of Universal Computer Science, 15(12), 2387-2408.

[35] Wikipedia, the free encyclopedia. Computer crime. Retrieved, January 15, 2012, from http://en.wikipedia.org/wiki/Computer_crime.

[36] Xu, Y., Ma, J., Sun, Yonghong, Hao, J., Sun, Yongqiang, & Zhao, Y. (2009). Using social network analysis as a strategy for e-commerce recommendation. PACIS Proceedings, 106.

[37] Yang, C. C., & Ng, T. D. (2007). Terrorism and crime related weblog social network: Link, content analysis and information visualization. Intelligence and Security Informatics, IEEE, 55-58.

[38] Zhang, X., Fuehres, H. & Gloor P. A. (2010). Predicting stock market indicators through twitter "I hope it was not as bad as I fear". Anxiety Journal, 1-8.

[39] Zhang, Y., Cao, B. & Yeng, D. Y. (2010). Multi-Domain collaborative filtering. Proceedings of the Twenty-Sixth Conference Annual Conference on Uncertainty in Artificial Intelligence, 725–732.

[40] Ziegler, C. N., McNee, S. M., Konstan J. A. & Lausen, G. (2005). Improving recommendation lists through topic diversification. WWW '05 Proceedings of the 14th international conference on World Wide Web.

Conception of a management tool of Technology Enhanced Learning Environments

Sérgio André Ferreira
School of Education and Psychology
Universidade Católica Portuguesa
Porto, Portugal

António Andrade
School of Economics and Management
Universidade Católica Portuguesa
Porto, Portugal

***Abstract*—This paper describes the process of the conception of a software tool of TELE management. The proposed management tool combines information from two sources: i) the automatic reports produced by the Learning Content Management System (LCMS) Blackboard and ii) the views of students and teachers on the use of the LCMS in the process of teaching and learning. The results show that the architecture of the proposed management tool has the features of a management tool, since its potential to control, to reset and to enhance the use of an LCMS in the process of teaching and learning and teacher training, is shown.**

Keywords-Learning Content Management System; Management Tool; Technology Enhanced Learning Environments.

I. INTRODUCTION

The introduction of change and educational innovation through technology in Higher Education Institutions (HEI) is a hot topic in research and in policies of international organizations and countries [1, 2]. The Learning Management Systems (LMS) and the Learning Content Management Systems (LCMS) are the most visible faces of the penetration of technology in HEI, and in many cases they are the only technological platforms to support the training activity, institutionalized and of common use by the different agents of the organization. These technological platforms are often associated with a financial investment, whose cost/ benefit ratio has to be justified.

The management of these new learning environments is critical to their success. The main factors that might endanger the success of any initiative in the introduction of technology in organizations have been identified. Several of these factors are related to aspects of management, including: introduction of technology without strategy, incipient evaluation, weak involvement of decision makers, attitudes of resistance [3].

In this study, based on the case of the Universidade Católica Portuguesa - Porto Regional Center (Católica - Porto), the process of the conception of a software tool of TELE management is described. During the school year 2003/2004, this University introduced the Blackboard LMS to support classroom teaching and to offer distance learning courses. In 2011/2012 a new investment was made for the provision of an LCMS, also Blackboard. These investments have a significant financial impact and are expected to provide a return in the educational field.

In a previous study, it was concluded that the current statistical reports produced by Blackboard neither provide critical information, nor provide a degree of disaggregation that allows the positioning of each CU, department and school/ university on the levels of integration of the LMS in the learning process [4]. The limitations of the reports affect their role as management tools, as they do not favor the dissemination of good practices or the removal of barriers, hence resulting in a slower penetration of the new culture [5, 6]. The goal of this article is to devise a management tool of the TELE that allows the possibility to give an answer to these limitations.

Planning a management tool requires a clear definition of the desired future, identifying the information subsystems required for this. In other words, it must meet the information needs of the organization and users. A tool developed without proper planning will result in a high degree of dissatisfaction among its users and will fall into disuse [7].

The approach to this problem was made by adopting a methodology of action research type, in which the researchers are actively involved in the cause of the research [8].The researchers, as users of Católica's TELE - Porto, have identified gaps in the information provided by the LCMS reports and have contributed to the information needs of the organization and users. This contribution was the basis of the work for which the representative of Blackboard/ service provider would introduce the changes required in the reports.

In addition to the objective data of the automatic LCMS reports, it is essential for management to obtain information about the users' views on the various dimensions of the TELE. This justifies the development of two questionnaires (one for teachers and one for students), so that this information could be compared with the data from the LCMS reports.

Beyond this introduction, this paper is divided into four chapters: in chapter 2 the TELE of Católica - Porto is contextualized and some data that demonstrates the dynamics of the use of the institutional TELE are presented; in chapter 3 the methodological approach is clarified and the information requirements of the organization and end-users that the management tool must address are explained; in chapter 4 the description of the whole process of the conception of the tool is presented; in chapter 5 the conclusions and proposals for future work are presented.

II. DYNAMICS OF CATÓLICA - PORTO'S TELE

Today's society, based on information and knowledge, requires a transformation with regard to educational infrastructure. Adopting an innovative approach to the education system is mainly related to the incorporation of technology in teaching practice [9]. Fig. 1 shows a possible architecture of a TELE. Currently, the learning environment is a result of the institutional vision associated with the personal vision of each student, hence a Hybrid Institutional Personal Learning Environment. The students' Personal Learning Environment (PLE) consists of the exploration of a multiplicity of skills available in the Cloud Learning Environment, which can go beyond the institutional vision. In fact, learning takes place increasingly through social media, institutional communities, exploring web tools, libraries of digital resources, repositories of Learning Objects, and other environments, tools and resources, which together result in the construction of the student's PLE student outside of the HEI.

The institutional environment will gradually incorporate this new way of learning, strongly supported by technology. Successful experiences with the integration of web 2.0 tools in contexts as unchanging and conservative in the use of technology as are lectures to dozens of students in auditoriums [10]; the ubiquitous presence of computer labs; the use of notebooks, netbooks, tablets and smartphones for teaching purposes; the availability of CU online via LCMS, are examples of reliable indicators of the construction of TELE fostered by the institution. It is, therefore, noticed a growing interest in the integration of the technological component in the educational field.

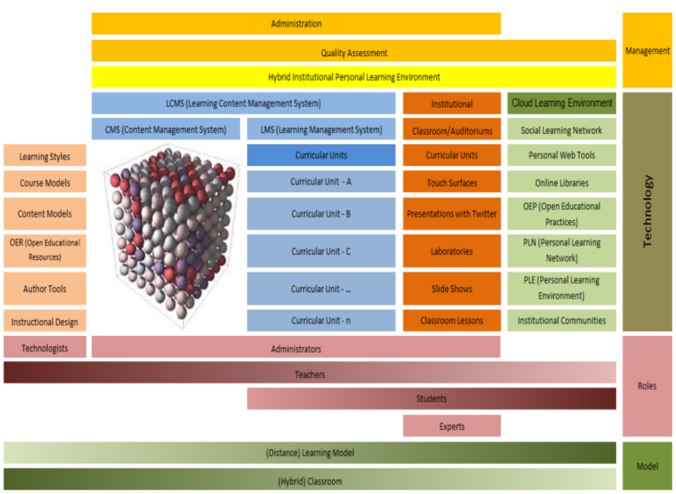


Figure 1. A possible architecture of a Hybrid Institutional Personal Learning Environment

Fig. 2 shows the LCMS as a central element of the campus of Católica - Porto, integrating the services of Learning System, Community System and Content System and maintaining communication with the administrative services – The Sophia Academic Management. The campus, supported by Windows Server and SQL Server, supports teachers, students, academic services and the public. Católica - Porto's TELE is the product of this context. However, the consistency in the management of these environments at the institutional level is not easy to achieve, due to lack of critical and meaningful information. In this paper, we try to make a contribution on this topic. The analysis is focused on the institutional TELE, which in this case is supported by LCMS Blackboard.

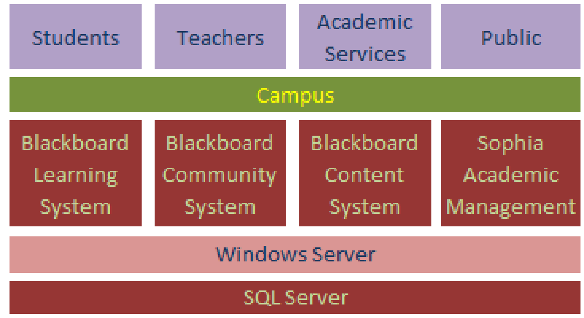


Figure 2. LCMS as a central element of the campus

Católica - Porto's TELE is currently a dynamic system. Between 17th October and 14th November 2011 there were 5,864 registered users of the system and 666 active CU. During this period, the maximum open sessions per hour hit 470. Fig. 2-5 show some data that proves the dynamics of the system.

The number of daily visits to the campus reaches, in the highest peaks, close to 4000 (Fig. 3), and the average time per visit is 7'53"; the number of visitors on the days of greater access exceeds 2000 (Fig. 4); there are many days when the peak of pages viewed/ day is located close to 60,000 (Fig. 5) and on average 16 pages are seen in each visit; the mobile access also reaches Fig. close to 150 hits on the days of higher peaks (Fig. 6).

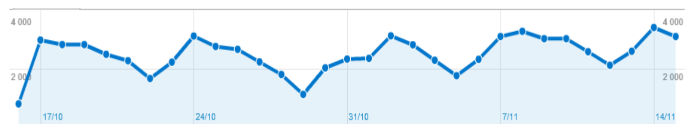


Figure 3. Daily visits to the campus

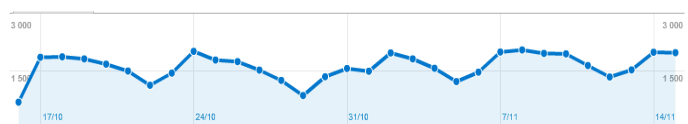


Figure 4. Visitors/day to the campus

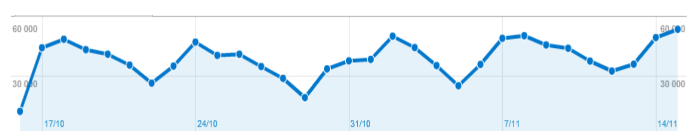


Figure 5. Pages seen/day

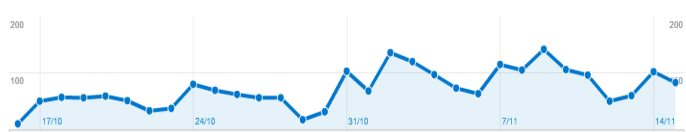


Figure 6. Mobile access

The financial investment made in acquiring the product and the services of the LCMS Blackboard and this dynamic of accesses justify the development of management tools that make available information on the integration of this TELE in the training process, guide educational policies and teacher training.

III. METHODOLOGY

The methodological approach followed is part of the action research. Such research can be defined as follows: "Action research aims to contribute to both the practical concerns of people in an immediate problematic situation and to further the goals of social science simultaneously. Thus, there is a dual commitment in action research to study the system and concurrently to collaborate with members of the system in changing it in what is together regarded as a desirable direction" [11]. In Fig. 7 the main methodological steps followed are systematized.

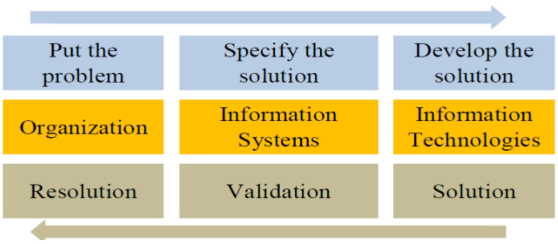


Figure 7. Methodological steps

1st *Put the problem*: The first step was to identify the problem at the organizational level. As already mentioned, there was a financial investment by the institution in the implementation of a TELE and the indicators provided by the Blackboard reports and via Google Analytics point to widespread use. It is important to justify the investment made and to understand the degree of integration of the LCMS in the formative process. The problem we face is to devise a management tool that delivers relevant information that can be aggregated according to the organizational plan of the institution.

2nd *Specify the solution*: Once the problem was identified [4], a solution, in which the way to use Information Systems (IS) is determinate and the architecture of the management tool is defined, was proposed in order to address the problem.

3rd *Develop the solution*: The implementation of the specified solution demands the improvement of the Blackboard reports. The inputs related to the definition of new information requirements, which respond to the needs of the institution and the users, were given by the investigators to the supplier of the service and to the representative of Blackboard. In a dialectical process, the provider of the services has made progressive approaches to the proposals presented by the researchers, which in turn have reoriented their solutions in an effort to reconcile their goals with the requirements of technological feasibility, presented by the engineers at Blackboard.

In addition to the reports of Blackboard, it was considered important to gather the views of users (teachers and students) on aspects related to the integration of the TELE in the teaching activity. This way, objective data on the exploitation of the features of TELE were compared with feedback from users. It has often been highlighted the importance of user involvement so that they understand and feel the usefulness of the IS [7]. In this way of conception of a management tool, this requirement is fully met in two ways: i) participation of researchers, who are also users, in the definition of the TELE automatic reports; ii) consultation of teachers and students who are TELE's users via a questionnaire.

TABLE I. PLANNING INFORMATION SYSTEMS

Activities	Goals
Strategic Analysis	To identify the current situation of the organization and SI (Where are we?)
Strategic Definition	To identify the vision and strategies to achieve it (Where do we want to go?)
Strategic Implementation	To plan, oversee and review the strategy (How will we get there?)

Each of these three methodological steps have correspondence in the three phases of the IS planning, synthesized by Varajão [7], as shown in Table I.

IV. THE MANAGEMENT TOOL

Fig. 8 presents a simplified diagram of the main entities involved in the teaching process in Católica - Porto.

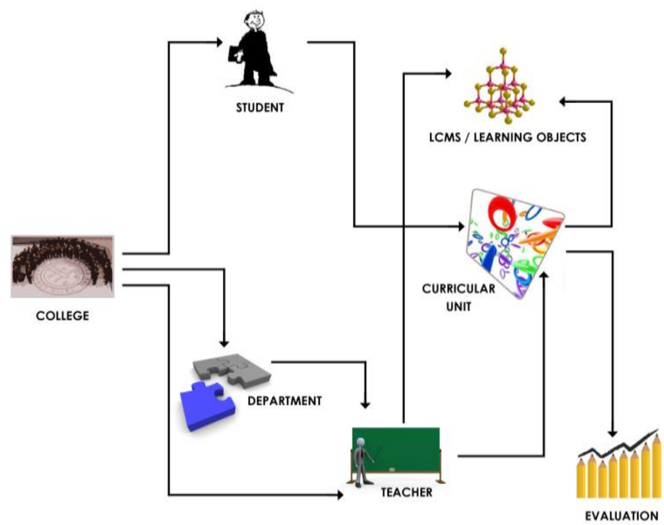


Figure 8. Simplified diagram of the main entities involved in the teaching process

Like many HEI, Católica - Porto has an organization on multiple levels: the university is divided into various colleges/ schools, that have under them several departments, in which the teachers are integrated (who can serve in different schools), who teach several CU (each CU is automatically created in the LCMS, virtual environment where part of the teaching activity takes place and where Learning Objects are made available). The institution's students are enrolled in certain CU and are automatically enrolled in these CU on LCMS (there is communication between the administrative management system and the Blackboard).

In the conception of the management tool, we tried to reflect this working model. In this context, the CU is the atom of information, from which the aggregation to higher levels is done, as it is shown in Fig. 9. One of the limitations identified in automatic reports from Blackboard was precisely the impossibility of making the aggregation of information across multiple levels [4].

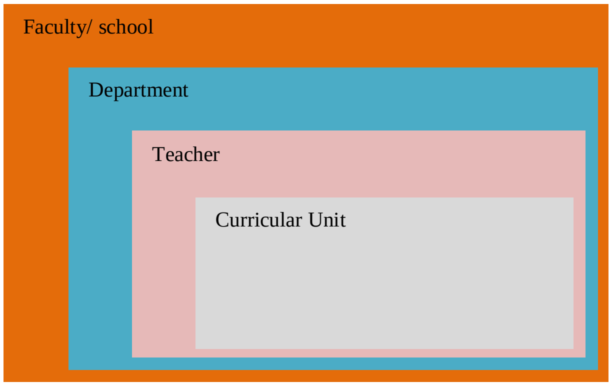


Figure 9. Aggregation of information reflecting the operating model

Another conclusion that was reached in the analysis phase was that the LCMS reports did not provide information on critical aspects to the understanding of the TELE's integration in teaching and learning. To address these limitations, seven dimensions that cover the main valences offered by the Blackboard were identified and indicators to characterize each of these dimensions were defined (table II). For each one of the indicators, metrics were established with five levels of integration in the formative process.

TABLE II. DIMENSIONS AND INDICATORS PROPOSED FOR THE AUTOMATIC LCMS REPORTS

Dimensions	Indicators
Dynamics of accesses	Hits per week / active user
Information	Information relating to the CU through notices, messages, CU programme (or summary) and calendar
Synchronous Communication	Number of open forums and nr of posts/ active user
Asynchronous Communication	% of users who use one or more asynchronous communication tools within the LCMS
Digital Content	Number of digitally rich content (it is considered to be rich digital content, all that goes beyond text and static image. Example: podcasts, electronic presentations, games, animations,...)
Delivery of work	Use of features relating to the delivery of individual and group papers, progress monitoring of the work group, detection of plagiarism
Evaluation	Number of tests performed in LCMS

Based on the indicators and metrics defined, a matrix with five levels of integration of the LCMS in the process of teaching and learning (Fig.10) was drawn up. This matrix of integration of technology in organizations and in the educational process through five stages of evolution is often present in studies [eg. 12, 13].

- *Entry*: Very low number of hits user/week. Lack of relevant information about the CU. No use or very little use of synchronous and asynchronous communication tools. Poor digital content. No delivery of works. No evaluation tests. The LCMS has a very limited impact on the teaching and learning process. It is possible to be successful at the CU without accessing the LCMS.

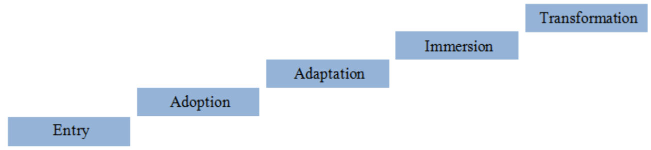


Figure 10. Levels of integration of the LCMS in the teaching and learning process

- *Adoption*: Low number of accesses user/week. Not much basic information about the CU. The synchronous and asynchronous communication tools are little explored. The presence of rich digital content is small, covering a small part of the key issues. The delivery of works via the platform is in its infancy, which limits the work of monitoring and detection of plagiarism. The assessment is only sporadically done and/or it is not important for the regulation of the study. The LCMS has limited impact, but it is visible in the process of teaching and learning. The student struggles to be successful in the CU without accessing the LCMS.

- *Adaptation*: The access to the CU is done regularly throughout the week. There is some basic information about the CU. The use of the synchronous and asynchronous communication tools is quite important in the construction of knowledge. The presence of rich digital content is visible, but they do not cover an important part of key issues. The delivery of works via the platform is sometimes associated with forms of communication, monitoring and detection of plagiarism. There are some key issues with assessment tests that are important for the regulation of the study. The LCMS has a clear impact on the teaching and learning process. It is extremely difficult for the student to be successful in CU without accessing the LCMS.

- *Immersion*: Access to CU is done on daily or almost daily basis. The majority of relevant information about the CU is available. The use of the synchronous and asynchronous communication tools is important in the construction of knowledge. There is digitally rich content that brings added value in comparison to printed material, covering most of the key themes of the CU. The delivery of works via the platform is usually associated with forms of communication, monitoring and detection of plagiarism. There are tests for most of the key issues that are important for the regulation of the study. The LCMS has a great impact on the teaching and learning process. The student cannot succeed without access to the LCMS.

- *Transformation*: The access to the CU is done daily or several times a day. All relevant information about the CU is available. The use of the tools synchronous and asynchronous communication is very important in the construction of knowledge. There is digitally rich content that brings great added value in comparison to printed material, covering most of the key themes of the CU. The delivery of works via the platform is always papers. There are tests for most of the key issues and they are very important for the regulation of the study. The LCMS is vital and has a transforming power in the process of teaching and learning.

In order to achieve this automatic placement via the LCMS reports, a back office system that allows the parameterization of the factors analyzed for each level of evolution was drawn up. To complement the information from the reports, which are translated into objective data on how the LCMS are used, it was considered important to ascertain the views of teachers and students about the same dimensions. Thus, the report data is compared with information from two questionnaires, which aims to understand the perspectives of teachers and students about key aspects of each dimension. Fig. 11 shows a possible form of representing the information: the radar charts.

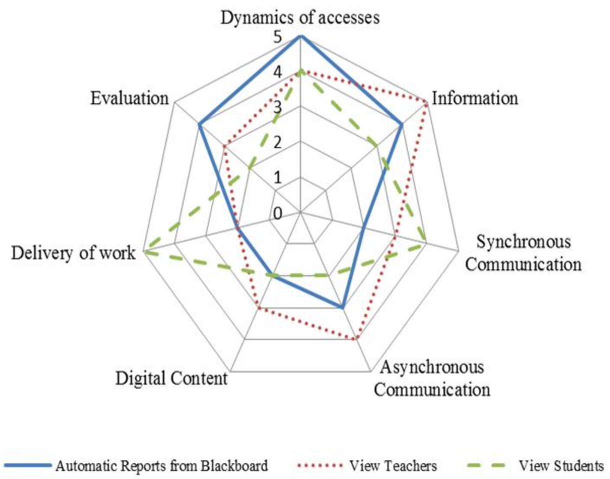


Figure 11. Radar Charts: Versatile forms of knowledge presentation

This type of chart, because of its versatility in knowledge representation, is often used in the analysis of organizational development and measurement of quality [14]. In an academic environment still in transition, where there is an increasing penetration of technology associated with pedagogical changes, radar charts are suitable and adaptable. In Fig. 11 the versatility of the radar chart is shown: appear represented in the same graph the dimensions of the TELE and the levels of integration of technology, where a comparative reading of data provenance can be made (Automatic reports from Blackboard, teacher's view, students' view).

The processing and the intelligent and versatile presentation of data are key features of management tools. Computer applications like OLAP (Online Analytical Processing) are useful as they offer these essential valences for decision-making [15].

The multidimensional OLAP functionality is based on structures called "cubes." The term "cube" is an analogy with the geometric object that implies three dimensions, but in real use, the OLAP cube can have more than three dimensions. The OLAP cube is comparable to a database, in which relationships between different dimensions and categories are established. It is a sophisticated technology that uses multidimensional structures to provide fast access to data for analysis. This organization facilitates the display of high-level of summaries. For the analysis of the TELE it can allow, for example, to extract data for each CU, for the dimensions in question, crossing automatic reports with the view of users.

The OLAP analysis system helps to organize data by many levels of detail (the information can be aggregated by CU, department, school/college or university), it also allows selecting and listing only the dimensions that we really want to consider in a given time. This strength allows conditional access to information, if this is the goal of the institution. In this case, each teacher will only have access to the information on the CU that he/she teaches, the coordinator of the department to all the CU of his/her department, the director of college/school to all the CU in the institution he/she manages, the Service Quality Management (SIGIQ) and the direction of Cattólica - Porto to all the information.

The data can also be selected by time periods or to view only a few variables. This way of presentation of information facilitates the interpretation of the results, since it reduces the entropy associated with high volumes of information.

In Fig. 12 the overview of the architecture proposed for evaluation of IS TELE is represented in a schematic form. The data comes from two sources: automatic reports from LCMS and the points of view of students and teachers (via questionnaire). The outputs of the management tool of information result in a matrix with five levels of integration of the LCMS in the process of teaching and learning. Currently, we are studying a way of providing information through an OLAP type application that allows a multidimensional analysis of data is under way.

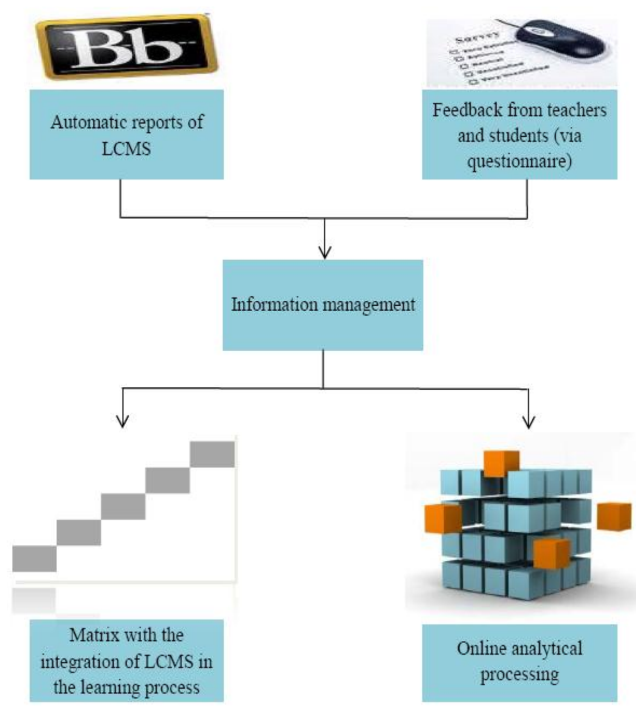


Figure 12. Architecture Overview of the management tool

V. CONCLUSIONS

In this paper the process of conception of a tool for managing a TELE is described, and the three steps of the IS Planning model synthesized by Varajão [7] were followed:

Strategic Analysis: The delimitation of the problem was made. The shortcomings in the current IS when compared with the demands of information by the organizations in order to manage the TELE were identified.

Strategic Definition: In the conception of the management tool a response to the identified problems was given, by defining the dimensions of analysis (i - dynamic access; ii - CU information; iii - synchronous communication; iv - asynchronous communication; v – digital contents; vi - delivery of papers, vii - evaluation) and the way of aggregating data, so that an analysis at various levels, in accordance with the organizational plan of the institution, is possible.

Solution Development: Through an action research methodology, researchers (LCMS users) presented the inputs for the improvement of the automatic reports and, in a working process in partnership with the supplier of the technical services, successive approximations have been made to the solutions presented. In this dialectical process, the proposals have been adequate to the requirements of technological feasibility, presented by the company's technical representative of Blackboard. In developing the tool it is expected that this information will be complemented with data collected, via questionnaire, reflecting the views of users (teachers and students) on the various dimensions of the TELE.

The main advantages of the proposed management tool and the methodology used are: i) the involvement of users in the conception of the tool, a fact that enhances its usefulness, ii) the comparison between data from the LCMS automatic reports with the view of users, a factor that potentially increases the effectiveness of the tool as a management tool; iii) the possibility of aggregation for hierarchical levels, which reflect the organizational plan of the institution.

VI. FUTURE WORK

As future work, it is predicted: i) to continue the improvement of the LCMS automatic reports and of the matrix to the position of the CU; ii) to develop a way to process and present OLAP data type, which articulates the results of the LCMS with the data from the questionnaires to teachers and students; iii) to integrate the OLAP component as a subsystem of quality management. Subsequently, it will be necessary to implement the management tool and carry out successive tests, in order to technically stabilize the system and refine the information output.

REFERENCES

[1] S. A. Ferreira and A. Andrade, "Models and instruments to assess Technology Enhanced Learning Environment in Higher Education," eLearning Papers, pp. 1-10, 2011.

[2] SNAHE, "E-learning quality: Aspects and criteria for evaluation of e-learning in higher education," Luntmakargatan2008.

[3] M. Rosenberg, Beyond e-learning. San Francisco: Pfeiffer, 2006.

[4] S. A. Ferreira and A. Andrade, "Aproximação a um modelo de análise da integração do LMS no processo formativo," in II CIDU 2011 - II Congreso Internacional da Docencia Universitária, Vigo, Spain, 2011.

[5] J. Kotter, Leading change. Massachusetts: Harvard Business School Press, 1996.

[6] J. Kotter and D. Cohen, The heart of change. Massachusetts: Harvard Business School Press, 2002.

[7] J. Varajão, A arquitetura da gestão de sistemas de informação, 2 ed. Lisboa: FCA - Editora Informática, 1998.

[8] R. Bogdan and S. Biklen, Investigação qualitativa em Educação. Porto: Porto editora, 1994.

[9] A. McCormack, The e-Skills Manifesto. A Call to Arms. Brussels, Belgium: European Schoolnet, 2010.

[10] S. A. Ferreira, C. Castro, and A. Andrade, "Cognitive Communication 2.0 in the Classroom - Resonance of an Experience in Higher Education," in 10th European Conference on e-Learning (ECEL 2011), University of Brigthon, Brigthon, United Kingdom, 2011, pp. 246-255.

[11] A. Dill, D. Gerwin, D. Mitchell, L. Russell, and B. Staley, Research Guidelines. Hadley - Massachusetts: WHSRP Planning Group, 2003.

[12] T. Teo and Y. Pian, "A model for Web adoption," Information & Management, vol. 41, pp. 457-468, 2004.

[13] T. Florida Center for Instructional. (2011). Technology Integration Matrix. Available: http://fcit.usf.edu/matrix/matrix.php

[14] D. Kaczynski, L. Wood, and L. Harding, "Using radar charts with qualitative evaluation: Techniques to assess change in blended learning," Active Learning in Higher Education, vol. 9, pp. 23-41, 2008.

[15] OLAP.COM. (2010). OLAP software and education wiki. Available: http://olap.com/w/index.php/OLAP_Education_Wiki

AUTHORS PROFILE

Sérgio André Ferreira Master in Sciences of Education, specialization in Education Computer Sciences.PhD student in Sciences of Education, specialization in Education Computer Sciences. His research is concerned with Technology Enhanced Learning Environments.

Webpage: http://www.sergioandreferreira.com/

António Andrade Senior Lecturer at the School of Economics and Management of the Catholic Universidade Católica Portuguesa. PhD in Technologies and Information Systems, MSc in Information and Management, Director of the MSc in Information and Documentation. His research is concerned with Technology Enhanced Learning Environments. Webpage: http://www.porto.ucp.pt/feg/docentes/aandrade/

E-Learning Methodologies and Tools

[1]Oye, N.D.
Faculty of Computer Science and Information System
Universiti teknologi Malaysia
81310 Skudai, Johor
+607-5532070

[2]Mazleena Salleh
Faculty of Computer Science and Information System
Universiti Technologi Malaysia
81310 Skudai, Johor
+607-5532006

[3]N. A. Iahad
Faculty of Computer Science and Information systems
Universiti Teknologi Malaysia
81310 Skudai Johor
+60129949511

Abstract— **E-learning is among the most important explosion propelled by the internet transformation. This allows users to fruitfully gather knowledge and education both by synchronous and asynchronous methodologies to effectively face the need to rapidly acquire up to date know-how within productive environments. This review paper discusses on e-learning methodologies and tools. The different categories of e-learning that includes informal and blending learning, network and work-based learning. The main focus of e-learning methodologies is on both asynchronous and synchronous methodology. The paper also looked into the three major e-learning tools which are (i) curriculum tools (ii) digital library tools and (iii) knowledge representation tools. The paper resolves that e-learning is a revolutionary way to empower workforce with the skill and knowledge it needs to turn change to an advantage. Consequently, many corporations are discovering that e-learning can be used as a tool for knowledge management. Finally the paper suggests that synchronous tools should be integrated into asynchronous environments to allow for "any-time" learning model. This environment would be primarily asynchronous with background discussion, assignments and assessment taking place and managed through synchronous tools.**

Keywords: E-learning; Synchronous; Asynchronous; Tools; Methodology; Knowledge management.

I. INTRODUCTION

In the last century, we have moved from the Industrial Age through the Information Age and now to the Knowledge Age. Knowledge and its efficient management constitute the key to success and survival for organizations in the highly dynamic and competitive world of today. Efficient acquisition, storage, transfer, retrieval, application, and visualization of knowledge often distinguish successful organizations from the unsuccessful ones. The ability to obtain, assimilate, and apply the right knowledge effectively will become a key skill in the next century. Learning is the key to achieving our full potential. Our survival in the 21first century as individuals, organizations, and nations will depend upon our capacity to learn and the application of what we learn to our daily lives.

E-learning has the potential to transform how and when employees learn. Learning will become more integrated with work and will use shorter, more modular, just-in-time delivery systems. E-learning delivers content through electronic information and communications technologies (ICTs). According to [2], the use of these facilities, involves various method which includes systematized feedback system, computer-based operation network, video conferencing and audio conferencing, internet worldwide websites and computer assisted instruction. This delivery method increases the possibilities for how, where and when employees can engage in lifelong learning. Employers are especially excited about the potential of e-learning for just-in-time learning delivery.

By leveraging workplace technologies, e-learning is bridging the gap between learning and work. Workers can integrate learning into work more effectively because they use the same tools and technology for learning as they use for work. Both employers and employees recognize that e-learning will diminish the narrowing gap between work and home, and between work and learning. E-learning is an option to any organization looking to improve the skills and capacity of its employees. With the rapid change in all types of working environments, especially medical and healthcare environments, there is a constant need to rapidly train and retrain people in new technologies, products, and services found within the environment. There is also a constant and unrelenting need for appropriate management and leveraging of the knowledge base so that it is readily available and accessible to all stakeholders within the workplace environment.

II. DEFINITION OFE-LEARNING

E-learning is not only about training and instruction but also about learning that is tailored to individuals. Different terminologies have been used to define learning that takes place online, a fact that makes it difficult to develop a generic definition. Authors agree that a single definition for e-learning has not yet been found. Terms that are commonly used to define online learning include e-learning, Internet learning, distributed learning, networked learning, tele-learning and telematics distributed learning [4], [1], virtual learning, computer-assisted learning, Web-based learning, and distance learning. It includes the delivery of content via Internet, Intranet, and Extranet, satellite broadcast, audio-video tape, interactive TV and CD-ROM [15]. Nonetheless, the different terminologies point to a similarly conceived educational experience. All of these terms imply that the learner is at a distance from the tutor or instructor, that the learner uses some form of technology (usually a computer) to access the learning material, and that the learner uses technology to interact with the tutor or instructor and other learners, and that some form of support is provided to learners[1].

E-learning refers to the use of information and communication technology (ICT) to enhance and/or support learning in tertiary education. However this encompasses an ample array of systems, from students using e-mail and accessing course materials online while following a course on campus to programmes delivered entirely online. E-learning can be different types, a campus-based institution may be offering courses, but using E-learning tied to the Internet or other online network (Lorraine M.2007). What is E-learning? E-learning is an education via the Internet, network, or standalone computer. E-learning is basically the network-enabled convey of skills and knowledge. E-learning refers to using electronic applications and processes to learn. E-learning applications and processes include Web-based learning, computer-based learning, virtual classrooms and digital collaboration. EL is when content is delivered via the Internet, intranet/extranet, audio or video tape, satellite TV, and CD-ROM. E-learning was first called "Internet-Based training" then "Web-Based Training" Today you will still find these terms being used, along with variations of E-learning. EL is not only about training and instruction but also about learning that is tailored to individual. Different terminologies have been used to define learning that takes place online [1, 2].

A. Categories of e-learning

These are considered as follows:-

1) Courses

Most discussion of e-learning focuses on educational courses. Educational course materials or courseware are usually modified and added with various different media and are uploaded to a networked environment for online accessing. Today, there are several popular learning management systems (LMS) such as WebCT and Blackboard which are commonly used by educational institutions. In achieving a more motivating courseware, courseware designers have began to add innovative presentation such as simulations, storytelling and various unique traits into the materials. E-learning has distinct similarities with classroom environment whereby both of the learners and the instructors are together related to the common course arrangement and flow.

2) Informal Learning

Information learning can be said to be one of the most dynamic and adaptable features of learning but nevertheless it is least recognized. Our need for information (and how we intend to use it) drives our search. Search engines (like Google) coupled with information storage tools (like Furl) and personal knowledge management tools like wikis and blogs present a powerful toolset in the knowledge workers portfolio. Cross [4] opined that in workplace we acquire more knowledge during break time than in a formal learning environment. We progress more in our jobs through informal learning, sometimes using trial and error and other times through conversations.

3) Blended Learning

Integrated learning provides a good transition from classroom learning to e-learning. Integrated learning which is also referred to as blended learning is a combination of a face to face and online learning. The productiveness of this method cannot be over emphasized. It encourages educational and information review beyond the classroom settings. Blended learning combines several different delivery methods, such as collaboration software, web-base courses and computer communication practices with face to face instruction [15]. Integrated learning utilizes the best of classrooms with the best of online learning.

4) Communities

Learning is social [1].The frequent challenges we battled with in our business milieu are sophisticated and unstable. Because we are in the global era, our methods of problem solving are changing daily. Therefore people dialogue with other members of the same organization or network globally to other organization. Communities strongly contribute to the flow of tacit knowledge.

5) Knowledge Management

Globalization is focused on e-learning because e-learning technology has the potential to bring improved learning opportunities to a larger audience than has ever previously been possible. [3] Suggested that a nation's route to becoming a successful knowledge economy is its ability to also become a learning society. Early KM technologies included online corporate yellow pages as expertise locators and document management systems. Combined with the early development of collaborative technologies (in particular Lotus Notes), KM technologies expanded in the mid-1990s. Subsequent KM efforts leveraged semantic technologies for search and retrieval and the development of e-learning tools for communities of practice. Knowledge management is an essential process which is concern with how to create atmosphere for people to share knowledge on distribution, adoption and information exchange activities in an organization[7], [16], [17].The semblance of knowledge management and the theory of e-learning reveals powerful relationship which is causing disarray between the two fields.

6) Learning Networks

Learning network is a procedure of developing and preserving relationship with people and information and communicating to support each other's learning. Therefore (LN) is enhancing and it offers chances to its members to engage online with each other, sharing knowledge and expertise. [13] States that, the use of pen and paper in our educational system today is producing inadequacy and challenges in the global era that we are in today where subject matter is changing speedily. The application of personal learning networks will create connections and develop knowledge for workers to remain current in their field.

III. WHAT IS REQUIRED FOR E-LEARNING TO BECOME AN EFFECTIVE KNOWLEDGE MANAGEMENT TOOL?

Several trends are spurring the momentum behind e-learning. There is the need for firms to keep up with the ever-changing businesses environment and shorter product lifecycles. Another trend is the growing importance of information sharing. E-learning can be taken outside of company firewalls and can be used to educate firm partners, customers, and suppliers, in addition to the firm's employees. In return, the firm can generate new knowledge through the use of chat rooms, surveys, etc. Knowledge partner's benefit from the information gained through e-learning, while the firm in turn benefits from the capture of new information from knowledge partners. Once information is captured and categorized as useful knowledge, its sources become irrelevant in terms of value. Cisco Systems, one of the many companies that promotes e-learning as part of its knowledge management strategy, defines the benefits of e-learning as follows (Cisco Systems, 2001): "E-learning provides a new set of tools that can add value to all the traditional learning modes – classroom experiences, textbook study, CD-ROM, and traditional computer-based training." Old-world learning models do not scale to meet the new world learning challenges. E-learning can provide the tools to meet that challenge.

IV. E-Learning Methodologies

E-learning exploits Web technology as its basic technical infrastructure to deliver knowledge. As the current trend of academic and industrial realities is to increase the use of e-learning, in the near future a higher demand of technology support is expected. In particular, software tools supporting the critical task of instruction design should provide automated support for the analysis, design, documentation, implementation, and deployment of instruction via Web.

A. Interaction in Learning

Learner(s) - Tutors(s) Interaction, and Learner(s) – Learner(s) Interaction: these two types of interactions are among humans, and they are the interaction forms that people are most familiar with. Therefore, most research studies are focusing on these two types of interaction, especially in the research of Computer Supported Collaborative Learning (CSCL). According to [13], if collaboration rather than individual learning designs were used in an online class, students should be more motivated to actively participate and should perceive the medium as relatively friendly and personal as a result of the online social interactions. This increased active group interaction and participation in the online course, hence, resulted in higher perceptions of self-reported learning. Whereas individuals working alone online tended to be less motivated, perceive lower levels of learning, and score lower on the test of mastery.

In CSCL, researchers usually distinguish two types of interactions between learner- tutor and learner- learner. The first one, synchronous interaction, requires that all participants of interaction are online at the same time. Examples include Internet voice telephone, video teleconferencing, text-based chat systems, instant messaging systems, text-based virtual learning environments, graphical virtual reality environments, and net based virtual auditorium or lecture room systems. Synchronous interaction promotes faster problem solving, scheduling and decision making, and provides increased opportunities for developing.

In 2000, Heron et al. studied the interaction in virtual learning groups supported by synchronous communication. They found that learning in virtual environments can be greatly enhanced by content-related dialogues with minor off-task talk, coherent subject matter discussion with explanation, and equal participation of students supported by synchronous interaction [14].. However, the cost of synchronous interaction is usually very high, and synchronous interaction is more constricted due to time differences. The second one is asynchronous interaction, in which learners or tutors have freedom of time and location to participate in the interaction, examples including interaction using e-mail, discussion forums, and bulletin board systems. It has been reported that by extending interactions to times outside of classes, more persistent interaction and closer interpersonal bonds among students can occur [12]. Thus, while one cannot totally simulate a real classroom with synchronous interaction, one can offer asynchronous interaction that provides time for better reflection, and allows global communication un-bounded by time zone constraints. Asynchronous interaction thus is more commonly provided in CSCL systems than the costly synchronous interaction.

V. E-Learning Tools

Here we discuss three types of e-learning tools: (i) curriculum tools,(ii) digital library tools and

(iii) knowledge representation tools. We can generally say that each type of tool emphasizes different parts of the process. Curriculum tools provide a systematic and standard environment to support classroom learning; their functions are particularly helpful in the initiation and selection stages. Digital library tools facilitate effective and efficient access to resources to support exploration and collection while knowledge representation tools focus on formulation and representation.

A. Curriculum Tools.

Curriculum tools are widely used in high school and college of education. Materials are selected and organized to facilitate class activities. Additional tools, such as discussion forums and online quizzes, are integrated to support collaboration and evaluation. A typical commercial curriculum tool includes three integrated parts: instructional tools, administration tools, and student tools. Instructional tools include curriculum design and online quizzes with automated grading. Administration tools include file management authentication, and authorization. Student tool functions include:

- Browsing class material: readings, assignments, projects, other resources
- Collaboration and sharing: asynchronous and synchronous bulletin boards and discussion forums.
- Learning progress scheduling and tracking: assignment reminders and submission, personal calendars, and activity logs.
- Self-testing and evaluation: tests designed by instructors to evaluate student performance
- WebCT and Blackboard are the most popular commercial curriculum tools. A review comparing these two tools suggests that Blackboard's flexible content management and group work support [3] make it more suitable for independent and collaborative learning. WebCT's tighter structure and fully embedded support tools make it more appropriate for guided, less independent learning. In general, these tools are tailored more to support class activities than independent research or self-study.

B. Digital library Tool

While curriculum tools support class functions, digital library tools focus on locating resources. These functions support the exploration and collection phases of information search. Digital library tools help users find the right information amidst a huge amount of digital material. Digital library features usually include search, browsing, and discovering special collections or exhibits. Search and browsing are used to locate resources and explore related topics. Special collections or exhibits contain organized materials representing a unique treasure for interested users.

C. 5.3 Knowledge Representation Tool

Knowledge representation tool help learners to visually review, capture, or develop knowledge. Curriculum tools rely primarily on a text-based, syllabus approach to describing course content. This approach often fails to delineate the relationship of concepts and skills covered in one course to those covered in another. It also fails to show the knowledge base that a learner will have acquired at the end of his/her course of study. A visualization tool can engage both learners and instructors in an active learning process when they construct spatial semantic displays of the knowledge, concepts, and skills that the learner possesses and acquires [22].

The e-Learning evolution proposes a good number of tools assisting the instructional designer during the analysis, design, implementation, and delivery of instruction via the Web [5]. If on one side an automated support should be provided by authoring tools [6],[16],[19], on the other side these tools should implement suitable e-learning process design methodologies [11],[21].

VI. CONCLUSION

E-learning is among the most important explosion propelled by the internet transformation. This allows users to fruitfully gather knowledge and education both by synchronous and asynchronous methodology to effectively face the need to rapidly acquire up to date know-how within productive environments. E-learning delivers content through electronic information and communications technologies (ICTs). According to [2], the use of these facilities, involves various methods which includes systematized feedback system, computer-based operation network, video conferencing and audio conferencing, internet worldwide websites and computer assisted instruction. This delivery method increases the possibilities for how, where and when employees can engage in lifelong learning. Finally we conclude that synchronous tools should be integrated into asynchronous environments to allow for "Any-time" learning model. This environment would be primarily asynchronous with background discussion, assignments and assessment taking place and managed through synchronous tools that integrate into the asynchronous environment.

ACKNOWLEDGEMENT

The authors would like to thank all authors that their papers are quoted. In addition, the authors gratefully acknowledge UTM, Research Universiti Teknologi Malaysia for their support and encouragement.

REFERENCE

[1] Ally, M. (2004). Foundations of Educational Theory for Online Learning. In T. Anderson & F. Elloumi (Eds.), Theory and Practice of Online Learning (pp. 3-31): Athabasca University & Creative Commons.

[2] Ajayi, I.A. (2008). Towards effective use of information and communication technology for teaching in nigerian colleges of education. Asian J. Inf. Technol. 7(5): 210 – 214.

[3] Bayne, S. and Cook, J.(2006). "WebCT vs BlackBoard? An Evaluation of Two Virtual Learning Environments",http://www.ltss.bris.ac.uk/interact21/in21p04.htm,

[4] Bates, A. W. (2005). Distance Education in a Dual Mode Higher Education Institution: A Canadian Case Study [Electronic Version]. Centre for Distance Education, Korean National Open University. Retrieved Nov. 13, 2005 from http://www.tonybates.ca/papers/KNOUpaper.htm.

[5] Bruce, L. R. and Sleeman, P. J.(2000). "Instructional Design: a primer". In Greenwich, CT: Information Age Publishing, 2000

[6] Campbell, J. D. and Mahling, D. E. (1998). "A Visual Language System for Developing and Presenting Internet-Based Education". In Proceedings of IEEE Symposium on Visual Languages. Nova Scotia, Canada, 1998, pp.: 66-67.

[7] Capozzi, M. M. (2007). "Knowledge Management Architectures Beyond Technology". First Monday 12 (6). http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1871/1754

[8] Czerniewicz, L .,and Brown, C. (2009). A virtual wheel of fortune? Enablers and constraints of ICTs in Higher Education in South Africa. In S. Marshall, W. Kinuthia and W. Taylor (Eds.), Bridging the Knowledge Divide (pp57-76). Charlote, NC: Information Age Publishing.

[9] Demetriadis, S., Triantfillou, E., and Pombortsis, A. (2003)."A phenomenographic study of students' attitudes toward the use of multiple media for learning," Proceedings of the 8th annual conference on Innovation and technology in computer science education, 2003.

[10] Deshpande, S.G., and Hwang, J.-N. (2001)."A Real-Time Interactive Virtual Classroom Multimedia Distance Learning System," IEEE TRANSACTIONS ON MULTIMEDIA (3:4), DECEMBER 2001, pp 432-444.

[11] Douglas, I.(2001). "Instructional Design Based on Reusable Learning Object: Applying Lessons of Object-Oriented Software Engineering to Learning System Design". In Proceeding ASEE/IEEE Frontiers in Education Conference, Reno, NV, 2001. Vol. 3. pp.: F4E1-F4E5.

[12] Haythornthwaite, C.(1999). "Collaborative Work Networks among Distributed Learners,"Proceedings of the 32nd Hawaii International Conference on System Sciences, IEEE Computer Society Press, 1999.

[13] Hiltz, S.R. and Turoff, M. (2002). " What makes learning networks effective?",Communicationsof the ACM(45:5) 2002, pp56-59

[14] Hron, A., H., F., C., U., and G., C.(2000). "Implicit and Explicit Dialogue Structuring in Virtual Learning Groups," British Journal of Educational Psychology (70) 2000, pp 53-64.

[15] Kaplan-Leierson, E.(2006). E-learning glossary. Retrieved on January 10, 2006 from http://www.learningcircuits.org/glossary.html.

[16] Kasowitz, A.(1998). "Tool for Automating Instructional Design. ERIC elearning house in Information Technology in Education". Available at http://ericit.org/digests/EDO-IR- 1998-01.shtml, Syracuse, NY, 1997.

[17] M. Lorrain, "Strategies to Engage Online students and Reduce Attrition Rates.," The Journal of Educator Online.2007.

[18] Maier R (2007): Knowledge management systems. Information and communication technologies for knowledge management. 3rd edn, Springer, Berlin

[19] Maier, R. and Thalmann, S. (2007). Describing learning objects for situationoriented knowledge management applications. In: Gronau N (ed) 4th Conference on Professional Konwledge Management Experiences and Visions, vol 2, GITO, Berlin, pp 343-351

[20] N.D. Oye , M. Salleh, and N.A. Iahad, "Challenges of E-learning in Nigerian University Education Based on the Experience of Developed Countries," International Journal of Managing Information Technology, vol. 3, no. 2, 2011, pp. 39-48.

[21] R. Sharma, M.S. Ekundayo, and E. Ng, “Beyond the digital divide: policy anaylysis for knowledge societies.,” Journal of Knowledge management, vol. 13, no. 5, 2009, pp. 373-386.

[22] Thomson, J. R. and Cooke, J.(2000). “Generating Instructional Hypermedia with APHID”. In Hypertext 2000. pp. 248-249.

[23] Vaishnavi, V., and Kuechler, W. (2004). "Design Research in Information Systems," 2004. Last update August, 2009.

[24] Vrasidas, C.(2002). “A Systematic Approach for Designing Hypermedia Environments for Teaching and Learning”. In International Journal of Instructional Media. Available at http://www.cait.org/vrasidas/hypermedia.pdf, 2002.

[25] Zhang, D. (2004)."Virtual Mentor and The LBA System ? Towards Building An Interactive, Personalized, and Intelligent E-Learning Environment," Journal of Computer Information Systems (XLIV:3) 2004, pp 35-43.

AUTHORS – BIBLIOGRAPHY

N.D.Oye, receive his M.Tech OR (Operations Research) degree from the Federal University of Technology Yola- Nigeria in 2002. He is a lecturer in the department of Mathematics and Computer Science in the same University (for the past 15yrs). At the moment he is a Phd student in the department of Information Systems in the Faculty of computer Science and Infor-mation systems at the Univeristi Teknologi Malaysia, Skudai, Johor, Malaysia. +60129949511 **oyenath@yahoo.co.uk**

Mazleena, Salleh received her PhD in Computer Science at Universiti Teknologi Malaysia in the field of computer networking. Presently she is the Head of Department of Computer Systems and Communication, Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia. Her research area include computer security, cryptograph and networking.

Noorminshah A. Iahad, PhD : She did her PhD at the School of Informatics, The University of Manchester. She worked with Professor Linda Macaulay from the Interactive Systems Design research section in the same school and Dr George Dafoulas from the School of Computing Science, Middlesex University. Her research is on investigating interaction patterns in asynchronous computer-mediated-communication. Her work includes analysing threaded discussion transcripts from the discussion feature of a well known Leaning Management System: WebCT. FSKSM, UTM 81310 Skudai, Johor, Malaysia.Email : minshah@utm.my This e-mail address is being protected from spambots. You need JavaScript enabled to view it , noorminshah@gmail.com,This e-mail address is being protected from spambots. You need JavaScript enabled to view it Office : +607 5532428

Fairness Enhancement Scheme for Multimedia Applications in IEEE 802.11e Wireless LANs

Young-Woo Nam, Sunmyeng Kim, and Si-Gwan Kim
Department of Computer Software Engineering
Kumoh National Institute of Technology
Gumi, South Korea

***Abstract*—Multimedia traffic should be transmitted to a receiver within the delay bound. The traffic is discarded when breaking its delay bound. Then, QoS (Quality of Service) of the traffic and network performance are lowered. The IEEE 802.11e standard defines a TXOP (Transmission Opportunity) parameter. The TXOP is the time interval in which a station can continuously transmit multiple data packets. All stations use the same TXOP value in the IEEE 802.11e standard. Therefore, when stations transmit traffic generated in different multimedia applications, fairness problem occurs. In order to alleviate the fairness problem, we propose a dynamic TXOP control scheme based on the channel utilization of network and multimedia traffic quantity in the queue of a station. The simulation results show that the proposed scheme improves fairness and QoS of multimedia traffic.**

Keywords- fairness; multimedia traffic; EDCA; QoS; TXOP.

I. INTRODUCTION

The IEEE 802.11 wireless LAN is widely used for wireless access due to its easy deployment and low cost. The IEEE 802.11 standard defines a medium access control (MAC) protocol for sharing the channel among stations [1]. The distributed coordination function (DCF) was designed for a contention-based channel access.

The widespread use of multimedia applications requires new features such as high bandwidth and small average delay in wireless LANs. Unfortunately, the IEEE 802.11 MAC protocol cannot support quality of service (QoS) requirements [2, 3]. In order to support multimedia applications with tight QoS requirements in the IEEE 802.11 MAC protocol, the IEEE 802.11e has been standardized [4]. It introduces a contention-based new channel access mechanism called enhanced distributed channel access (EDCA). The EDCA supports the QoS by introducing four access categories (ACs). To differentiate the ACs, the EDCA uses a set of AC specific parameters, which include minimum contention window $CWmin[i]$, maximum contention window $CWmax[i]$, and arbitration interframe space (AIFS) AIFS[i] for AC i ($i = 0, \ldots, 3$). Furthermore, the EDCA introduces a transmission opportunity (TXOP). The TXOP is the time interval in which a station has the right to initiate transmission. In other words, a station can transmit multiple data packets consecutively until the duration of transmission exceeds the specific TXOP limit. The TXOP provides not only service differentiation among various ACs, but also improves the network performance.

In the original TXOP of the EDCA, the TXOP limits at stations are fixed and generally allocated among stations with identical traffic load. Under this condition, fair bandwidth allocation is expected. However, if stations transmit data packets with different traffic load, fairness problem arises. This problem is explained in detail in Section II.

In order to support multimedia traffic, many schemes have been proposed in the literature. However, the previous schemes still have several problems. First, some of them need modifications to the IEEE 802.11e standard [5-8]. Therefore, they are not backward compatible with the legacy EDCA. For example, Deng et al. proposed a surplus TXOP diverter (STXD) scheme to define the TXOP limit for per-flow but not for per-ACs [6]. However, the standard is on a per-AC basis. Second, some use analytical models to calculate the QoS metrics which are usually derived based on a few impractical hypotheses. They do not reflect the characteristics of multimedia traffic. Therefore, they are always inaccurate and clearly not applicable to realistic environments [9, 10]. Third, some require feedback information from stations to consider the dynamic behavior of multimedia flows, but the feedback cannot provide an appropriate indication to the current network load conditions in a real-time manner [11, 12]. Finally, others proposed very simple schemes to allocate the TXOP limit. A threshold-based dynamic (TBD) TXOP scheme dynamically adjusts the TXOP limit according to the queue length and the pre-setting threshold [13]. Each station has two TXOP limit values: a low and a high TXOP. If the queue length is below the threshold, the TXOP limit is fixed at the low value; otherwise, the TXOP limit is set to the high value. A distributed optimal (DO) TXOP scheme proposed in [14] uses the throughput information instead of the queue length. In the DO TXOP scheme, each station measures its throughput and compares it with the target throughput. If the measured throughput is higher than the target value, the station reduces its TXOP limit; otherwise, it increases its TXOP. It is hard for the stations in the TBD and DO TXOP schemes to have adequate TXOP limit since the both schemes allocate the TXOP limit based on only one parameter: the pre-setting threshold in the TBD TXOP scheme and the target throughput in the DO TXOP scheme. And the TBD and DO TXOP schemes do not take into account the channel utilization. Therefore, at high loads, all the stations in the both schemes have large TXOP limit. On the contrary, at light loads, all the stations have small TXOP limit.

In this paper, we propose a simple and effective scheme for alleviating the fairness problem. The proposed scheme dynamically adjusts the TXOP limit based on the local information such as the channel utilization at QoS access point (QAP) and current network load at stations without any feedback information. Therefore, we call the proposed scheme DTC (Dynamic TXOP Control) scheme.

The paper is organized as follows. The fairness problem is presented in Section II. In Section III, the proposed DTC scheme is explained in detail. In Section IV, we discuss simulation results. Finally, we conclude in Section V.

II. FAIRNESS PROBLEM

All stations use the same TXOP limit in the IEEE 802.11e EDCA. If traffic quantity of each station is same, no problem occurs, because bandwidth is allocated fairly. If each station supports multimedia application service with different traffic generation rate, fairness problem occurs. As traffic generation rate is different, each station has mutually different traffic quantity. If all stations use the same TXOP limit value in this situation, a station with less multimedia traffic quantity can promptly transmit data packets in its queue. Thus, the station acquires good performance by satisfying its delay bound. However, a station with more multimedia traffic quantity performs backoff process in several times, which lengthens waiting time to transmit packets. As the delay bound of packets is not satisfied, a receiver discards them, thereby lowering performance of multimedia traffic. Thus, stations with less traffic quantity always have better performance than those with more traffic quantity. This causes fairness problem among stations with different traffic quantity.

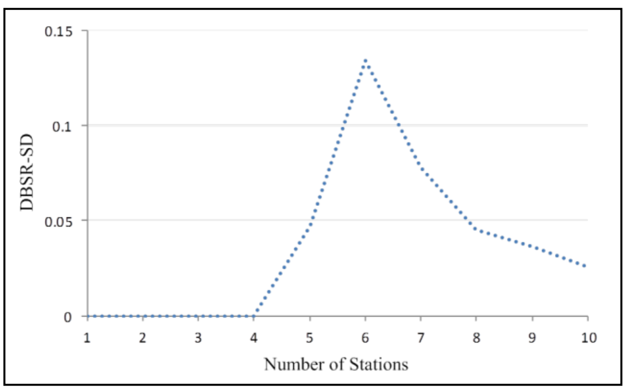


Figure 1. Standard deviation of delay bound success ratio according to the number of stations

Fig. 1 shows the standard deviation of delay bound success ratio for the IEEE 802.11e EDCA where all stations have the same TXOP limit. We simulated by using the simulation parameters in Tables I and II of Section IV. Delay bound success ratio is the number of data packets successfully transmitted to a receiver over the total number of transmitted data packets. The standard deviation hardly varies when there are few stations in the figure, since small traffic quantity can be immediately transmitted, regardless of traffic quantity in each station. When there are more than 4 stations, however, the standard deviation rapidly increases. This is because the delay bound success ratio becomes lower for elongated waiting time of packets in the queue of stations with more multimedia traffic quantity. When there are more than 6 stations, the standard deviation decreases, because the delay bound success ratio of all stations is lowered due to channel congestion. Like this, the provided QoS varies, depending on the multimedia traffic quantity of each station.

III. DTC (DYNAMIC TXOP CONTROL) SCHEME

The proposed DTC scheme dynamically adjusts TXOP limit value in consideration of channel utilization and queue utilization to alleviate the fairness problem depending on the traffic quantity. When the channel utilization of network is low, the DTC scheme can transmit more data packets. Then channel contention gets lower, if higher TXOP limit is allocated to a station with many packets in the queue. Thus, the possibility of packet collision becomes lower and channel waste can be reduced to improve overall network performance. The QoS of multimedia traffic may be lowered because a station with less data packets in the queue uses relatively lower TXOP limit. However, the DTC scheme is still effective since fair QoS can be provided irrespective of difference in traffic quantity.

The proposed DTC scheme is made up of two processes. First, QAP calculates the TXOP limit based on the channel utilization of network and then transmits it to stations through a beacon frame. Second, a station calculates the TXOP limit to be actually used to transmit data packets based on its queue utilization and the TXOP limit obtained from the beacon frame. Hereinafter the former is referred to as $TXOP_{QAP}$ and the latter, as $TXOP_{STA}$ to distinguish between TXOP limit calculated by QAP and TXOP limit calculated by a station.

A. Process to Calculate TXOP limit at QAP

The channel utilization of network is used to calculate TXOP limit at QAP. Channel utilization is calculated by dividing the busy time of channel by a beacon frame transmission period. Busy time means time when channel is used, whether packets are successfully transmitted or not.

QAP measures channel busy time (*Busy*) using the carrier sensing during a beacon frame transmission period. Channel utilization (C_Util) is calculated as follows.

$$C\_Util = \frac{Busy}{BeaconPeriod} \quad (1)$$

where *BeaconPeriod* indicates the period of a beacon frame.

As the channel utilization calculated in (1) fluctuates very irregularly in each calculation, performance fluctuates considerably if the calculated value is used as it is. Since it cannot be used in TXOP limit calculation as it is, moving average window is used as follows.

$$C\_Util_n = (1-\alpha) \cdot C\_Util_{n-1} + \alpha \cdot C\_Util_{current} \quad (2)$$

where $C\_Util_{current}$ is the channel utilization measured in the nth beacon frame period, C_Util_{n-1} and C_Util_n are the average channel utilizations at the end of n-1th and nth beacon frame period, respectively. α is a smoothing factor.

We introduce 4 new parameters to calculate $TXOP_{QAP}$: TXOP limit maximum value ($TXOP_{QAPmax}$) and minimum value ($TXOP_{QAPmin}$) at QAP, and upper threshold of channel utilization (C_Util_{high}) and lower threshold (C_Util_{low}).

Channel utilization increases and converges to the maximum utilization as the number of stations increases as shown in Fig. 2. As the number of stations increases, so does the possibility of packet collision. Thus, channel utilization increases due to frequent retransmission of packets.

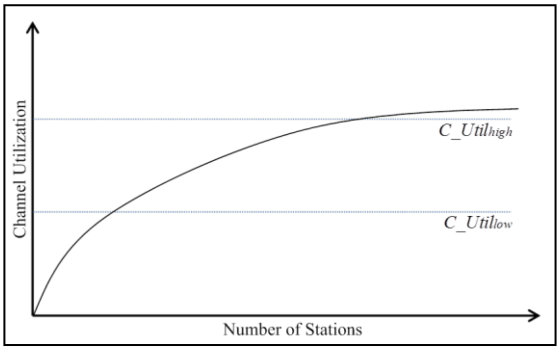


Figure 2. Channel utilization according to the number of stations

High channel utilization means that data packets are continuously transmitted as many stations contend for channel. Therefore, network performance needs to be improved by reducing TXOP limit. If channel utilization is low, TXOP limit should be increased so that a station can transmit many packets in a backoff process. Thus, $TXOP_{QAP}$ is set to $TXOP_{QAPmin}$, if the measured channel utilization is larger than the upper threshold. If it is smaller than the lower threshold on the contrary, $TXOP_{QAP}$ is set to $TXOP_{QAPmax}$. When it is between the upper and lower thresholds, QAP calculates $TXOP_{QAP}$ value as follows by using the channel utilization obtained from (1) and (2).

$$TXOP_{diff} = TXOP_{QAPmax} - TXOP_{QAPmin}. \tag{3}$$

$$TXOP_{QAP} = TXOP_{diff} \cdot \frac{C\_Util_{high} - C\_Util}{C\_Util_{high} - C\_Util_{low}} + TXOP_{QAPmin}. \tag{4}$$

$TXOP_{QAPmin}$ is added to (4) so as to ensure that the calculated $TXOP_{QAP}$ is always larger than $TXOP_{QAPmin}$.

In Fig. 2, the upper threshold of channel utilization similar to maximum value is selected to fully use channel. The lower threshold is selected at medium value instead of low value, because no fairness problem occurs when channel utilization is low since traffic of all stations can be transmitted within their delay bound. We will explain how to decide these values in Section IV.

QAP transmits $TXOP_{QAP}$ value obtained in the above process to all stations through a beacon frame. The process to calculate $TXOP_{QAP}$ value based on the channel utilization at QAP is described in Fig. 3.

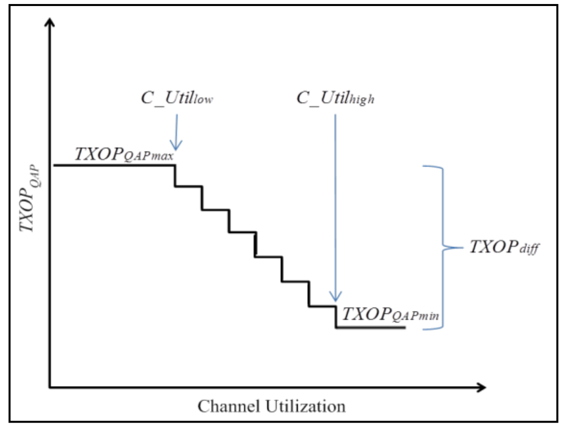


Figure 3. Calculation process of TXOP limit at QAP

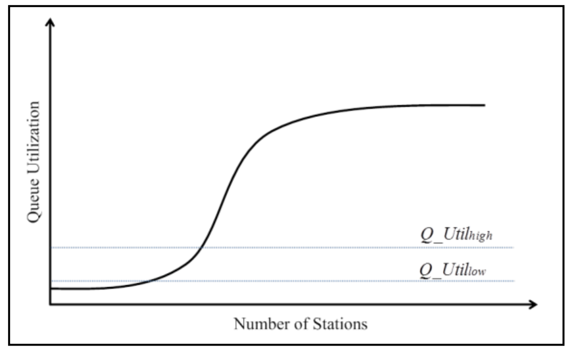


Figure 4. Queue utilization according to the number of stations

B. Process to Calculate TXOP limit at a Station

This subsection explains the process to calculate $TXOP_{STA}$ to be used by each station. $TXOP_{STA}$ is calculated based on the queue utilization of a station and $TXOP_{QAP}$ obtained from a beacon frame transmitted by QAP.

Each station calculates queue utilization (Q_Util) after receiving a beacon frame. The utilization shows the quantity of data packets in the queue of a station. It is calculated as follows.

$$Q\_Util = \frac{Q_{packet}}{Q_{size}} \tag{5}$$

where Q_{size} is the maximum number of packets that can be kept in the queue and Q_{packet} is the number of packets in the queue.

Similar to the channel utilization calculation, queue utilization is calculated as follows by using moving average window to reflect traffic pattern.

$$Q\_Util_n = (1-\alpha) \cdot Q\_Util_{n-1} + \alpha \cdot Q\_Util_{current} \tag{6}$$

where $Q\_Util_{current}$ is the queue utilization measured after receiving nth beacon frame, Q_Util_{n-1} and Q_Util_n are the average queue utilizations after receiving n-1th and nth beacon frame, respectively.

Fig. 4 shows the queue utilization depending on the increase of station number. In the figure, as the number of stations increases, queue utilization rapidly increases and converges to maximum size of queue. When there are many stations, the collision probability and the time to wait for transmission become bigger. Then, newly generated traffic exceeds the quantity of packets completely transmitted. Thus, the quantity of packets waiting for transmission in queue increases. Further, packets generated in excess of maximum size of the queue are discarded.

We introduce 3 new parameters to calculate $TXOP_{STA}$: upper threshold (Q_Util_{high}) of queue utilization and lower threshold (Q_Util_{low}), and TXOP limit minimum value ($TXOP_{STAmin}$) at a station.

Each station calculates $TXOP_{STA}$ value by using queue utilization acquired from (5) and (6). This process is similar to the process where QAP calculates $TXOP_{QAP}$ value by using channel utilization. Since high queue utilization means that queue currently has many data packets, the delay bound of packets should be satisfied by transmitting the packets fast by increasing TXOP limit. Thus, $TXOP_{STA}$ is set to $TXOP_{QAP}$ if the measured queue utilization is larger than the upper threshold. If it is smaller than the lower threshold on the contrary, $TXOP_{STA}$ is set to $TXOP_{STAmin}$. If it is between the upper and lower thresholds, it is calculated as follows.

$$TXOP_{diff} = TXOP_{QAP} - TXOP_{STAmin}. \quad (7)$$

$$TXOP_{STA} = TXOP_{diff} \cdot \frac{Q\_Util_{high} - Q\_Util}{Q\_Util_{high} - Q\_Util_{low}} + TXOP_{STAmin}. \quad (8)$$

In (8), $TXOP_{STAmin}$ is added to ensure that the calculated $TXOP_{STA}$ is always larger than $TXOP_{STAmin}$.

In Fig. 4, there is not large difference between the upper and lower threshold values of queue utilization. Unless a station transmits all the packets in its queue within a given TXOP limit, queue utilization increases continuously. Thus, the upper threshold value should be set low to enable a station to have large TXOP limit. Then, the station can transmit packets fast. We will explain how to decide these values in Section IV.

The process where a station calculates $TXOP_{STA}$ depending on queue utilization is described in Fig. 5.

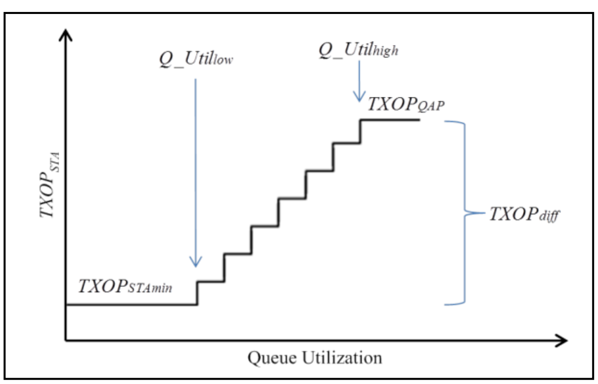


Figure 5. Calculation process of TXOP limit at a station

IV. PERFORMANCE EVALUATION

Let us discuss the simulation results of the proposed DTC scheme. To validate the proposed scheme, we compare them to the results of the IEEE 802.11e standard EDCA. The parameters used in the simulation are listed in Table I. The average is calculated after repeating simulation for 30 seconds 10 times. Although TXOP limit value is given as time interval in the IEEE 802.11e standard, here it is represented in terms of the number of data packets. Thus, $TXOP_{STAmin}$ value for the proposed DTC scheme is set to 2 data packets, $TXOP_{QAPmax}$ value is set to 10 data packets and $TXOP_{QAPmin}$ value is set to 8 data packets. On the contrary, fixed 5-data packet time is set for the IEEE 802.11e EDCA. 0.9 is used as smoothing factor. The delay bound of multimedia data packets is set to 33ms. Unless a packet is transmitted to a receiver within 33ms, it is discarded.

TABLE I. SIMULATION PARAMETERS

Parameter	Value
Simulation Time	30 s
Beacon Period	100 ms
$TXOP_{QAPmin}$	8
$TXOP_{QAPmax}$	10
$TXOP_{STAmin}$	2
Delay Bound	33 ms
Q_Util_{high}	0.20
Q_Util_{low}	0.05
Queue Size	100
Smoothing Factor	0.9
C_Util_{high}	0.95
C_Util_{low}	0.75

A constant data packet size of 1500 bytes is used. We use the negative exponential distribution to get the lengths of the data packet inter-arrival times. The average inter-arrival time of the distribution with arrival rate parameter λ is 1/λ.

TABLE II. MULTIMEDIA DATA RATE PER STATION

	Inter-arrival Time(μs)	Data Rate(Mbps)
Group 1	2326	5.16
Group 2	1587	7.56

The number of stations used in the simulation is 1~10 and the stations are divided into 2 groups by half. λ is set as shown in Table II to differently set multimedia traffic quantity to be transmitted by stations belonging to each group. The average inter-arrival time of stations in group 1 is set to 2326μs ($\lambda = 0.00043$). Thus, these stations generate data packets at a rate of 5.16 Mbps. The average inter-arrival time of stations in group 2 is 1587μs ($\lambda = 0.00063$) and data generation rate is 7.56 Mbps.

The following performance metrics are used to compare and analyze the results of simulation.

- Channel Utilization: the fraction of time that the channel is used to transmit data packets.
- Normalized Throughput: the amount of useful data successfully transmitted divided by the capacity of the medium.
- DBSR (Delay Bound Success Ratio): the ratio of the number of packets which are successfully transmitted without breaking their delay bound to the number of all data packets transmitted to a receiver.
- DBSR-SD (Standard Deviation of Delay Bound Success Ratio): the standard deviation of DBSR.
- Queue Utilization: the ratio of the number of packets in the queue to the queue size.

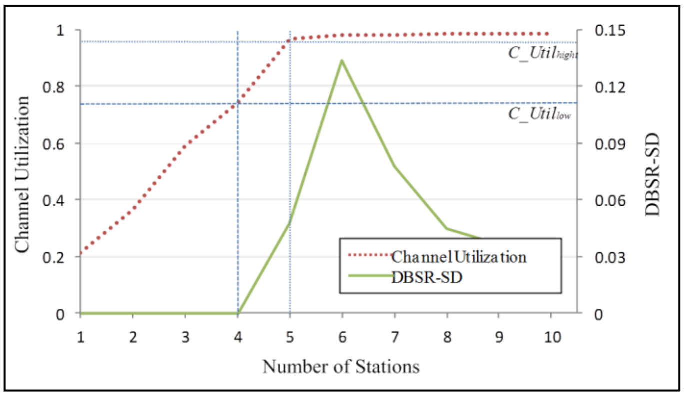


Figure 6. Channel utilization and DBSR-SD according to the number of stations

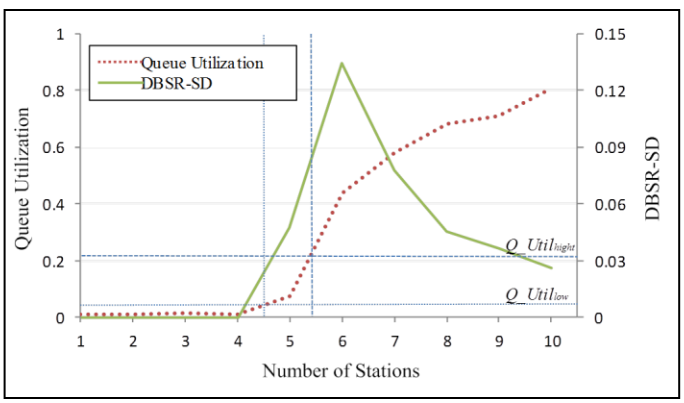


Figure 7. Queue utilization and DBSR-SD according to the number of stations

The values of C_Util_{low} and C_Util_{high} are decided by using the simulation results for TXOP limit of the IEEE 802.11e EDCA. Fig. 6 shows the results of channel utilization and DBSR-SD according to the number of stations. C_Util_{low} value is set to 0.75, the point where DBSR-SD increases and the fairness problem occurs when there are more than 4 stations. C_Util_{high} value is set to 0.95, the point where channel utilization begins to converge to maximum value.

The values of Q_Util_{low} and Q_Util_{high} are decided by using the results acquired in the same environment as Fig. 6.

Fig. 7 shows the queue utilization and the DBSR-SD caused by increasing the number of stations. Q_Util_{low} value is set to 0.05, the point where a station cannot transmit all the packets in its queue during one TXOP limit, and Q_Util_{high} is set to 0.2, because DBSR-SD rapidly increases in the interval where the number of stations change from 5 to 6.

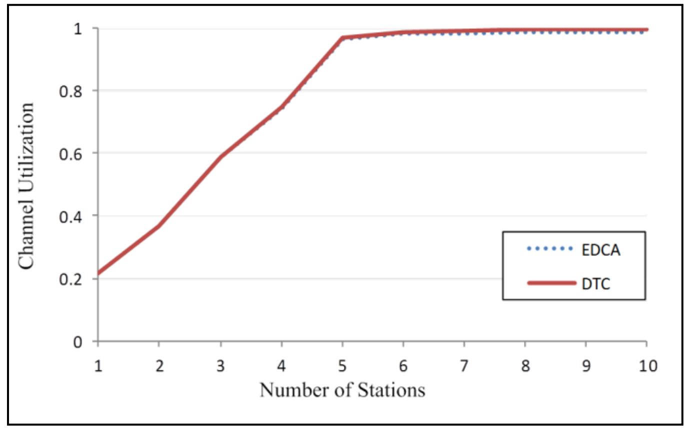


Figure 8. Channel utilization according to the number of stations

Fig. 8 shows channel utilization of the IEEE 802.11e EDCA and the proposed DTC scheme according to the number of stations. The figure shows that both of EDCA and DTC have similar channel utilization irrespective of the number of stations. The transmitted traffic does not fully use the capacity of media until the number of stations reaches 5.

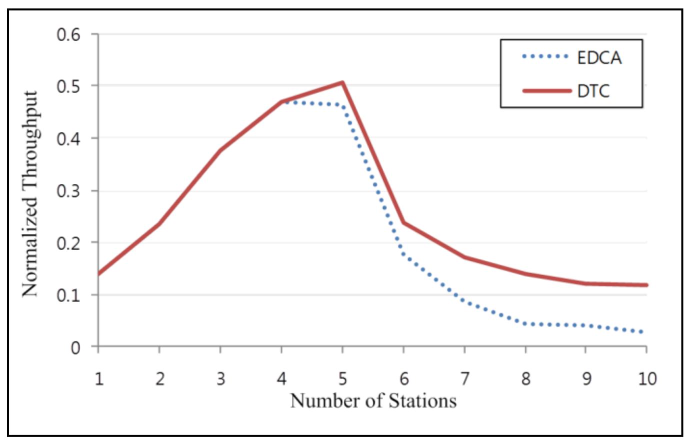


Figure 9. Normalized throughput according to the number of stations

Fig. 9 indicates normalized throughput depending on the number of stations. The throughput is calculated by considering only packets successfully transmitted within their delay bound and it is indicated in the figure through normalization. Although throughput increases similarly in both of DTC and EDCA until the number of stations reaches 4, the throughput of the DTC scheme is higher thereafter as the number of stations increases.

As shown in Fig. 8, although channel utilization is nearly same, the proposed DTC scheme further satisfies delay bound of transmitted data packets. Fig. 9 shows that our scheme is better scheme than the IEEE 802.11e EDCA.

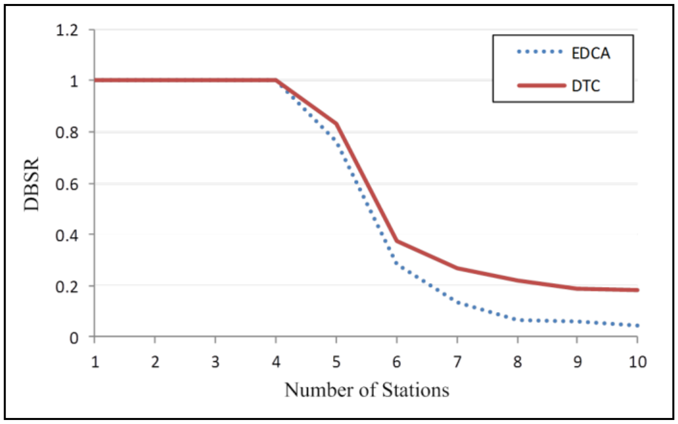


Figure 10. DBSR according to the number of stations

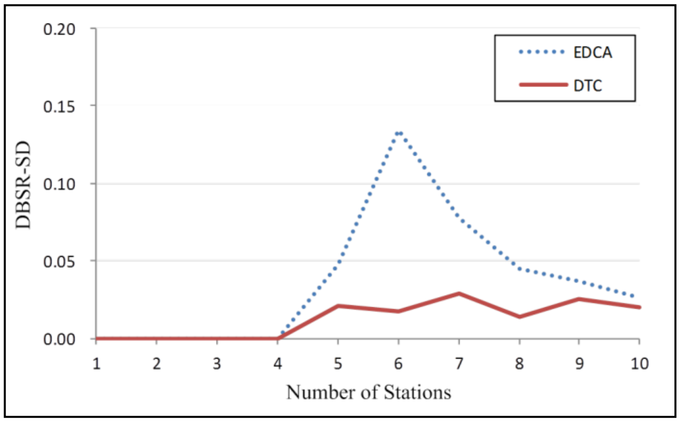


Figure 11. DBSR-SD according to the number of stations

Fig. 10 indicates the ratio of data packets which is successfully transmitted to a receiver within its delay bound. As shown in the figure, when there is small number of stations and channel utilization is low, the proposed DTC and the IEEE 802.11e EDCA have success ratio of almost 100%. However, as the number of station increases, DBSR rapidly decreases due to channel congestion. The proposed DTC scheme, however, has higher success ratio than the EDCA. The high DBSR means that the quantity of packets discarded due to breaking their delay bound is low. We can see that the DTC scheme has higher performance than the EDCA when the number of stations increases.

Fig. 11 shows the results of DBSR-SD according to the number of stations. From the figure, we can see that the DBSR-SD of the proposed DTC scheme is up to 10% lower than that of the EDCA. Since low standard deviation among stations implies that DBSR of each station is similar, the proposed scheme is found to be fairer in providing QoS to stations regardless of the quantity of multimedia traffic.

Fig. 12 shows the results of average queue utilization. The average queue utilization of the DTC scheme is lower than that of the IEEE 802.11e EDCA. Since the increased number of stations increases queue utilization, the DTC scheme uses bigger TXOP limit to transmit more data packets. Therefore, in the DTC scheme, multimedia traffic is transmitted faster than the IEEE 802.11e EDCA and the performance of overall network is improved.

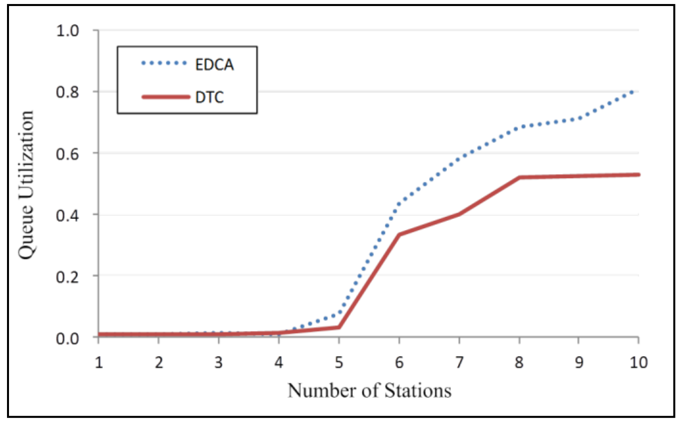


Figure 12. Queue utilization according to the number of stations

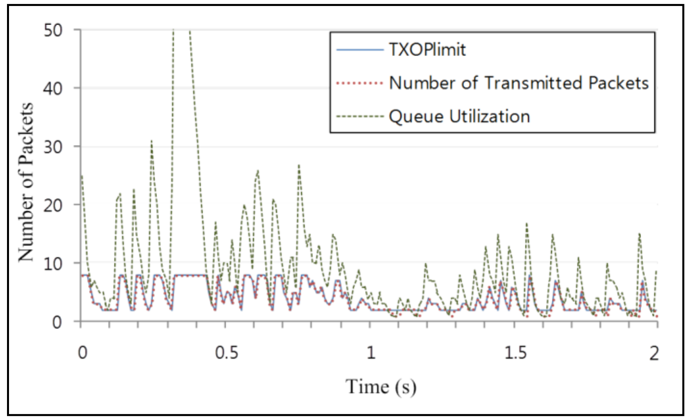


Figure 13. TXOP limit, number of transmitted packets, and queue utilization according to time in the DTC scheme

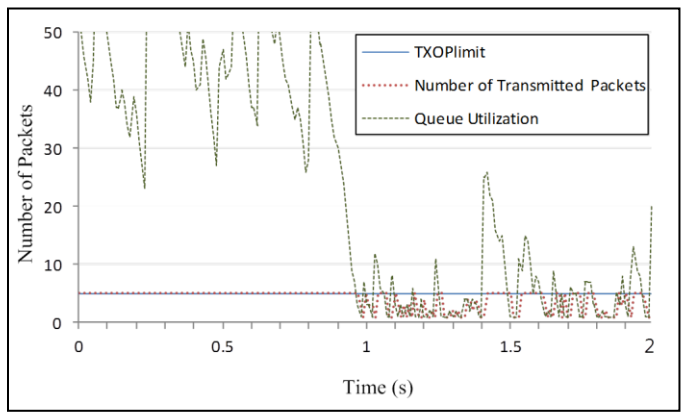


Figure 14. TXOP limit, number of transmitted packets, and queue utilization according to time in the EDCA

Figs. 13 and 14 show the results of TXOP limit value, queue utilization, and the number of transmitted packets depending on the time in one station for the DTC and the IEEE 802.11e EDCA. In the figures, the unit of queue utilization is converted to the number of packets in the queue in order to keep consistency of unit on Y axis. When there are 10 stations, one station is randomly chosen to be measured for 2 seconds. The EDCA always has fixed TXOP limit of 5 and its queue utilization is large. The DTC scheme has TXOP limit of from 2 to 8 and its queue utilization is low in average. Although $TXOP_{QAPmax}$ value is set to 10 in Table 1, the DTC has TXOP limit value greater than or equal to 9 only when channel utilization is very low and queue utilization is very high. Thus, it cannot have such values in the situation shown in Fig. 13. The average TXOP limit value of the DTC scheme is very close to 5 and the difference in transmission possibility may be low among stations, when compared to the IEEE 802.11e EDCA.

V. Conclusion

As the IEEE 802.11e EDCA applies the same TXOP limit to stations where multimedia traffic is generated in different quantity, QoS is discriminatorily provided to each station. To alleviate this problem, we propose the DTC scheme which dynamically adjusts TXOP limit. In the DTC scheme, QAP uses channel utilization to calculate TXOP limit value, which is transmitted to each station through a beacon frame. And then, a station calculates the final TXOP limit value based on its own queue utilization information. Considering the channel utilization and queue utilization, the proposed DTC scheme adaptively allocates TXOP limit value depending on the network state to provide stations with QoS which is fairer than that provided by the IEEE 802.11e DECA. It is confirmed that the overall network performance is improved as transmission success ratio within the delay bound becomes larger.

Acknowledgment

This paper was supported by Research Fund, Kumoh National Institute of Technology.

References

[1] IEEE, “Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications,” *IEEE Standard 802.11*, June 1999.

[2] Y. Xiao, “A simple and effective priority scheme for IEEE 802.11,” *IEEE Communications Letters*, vol. 7, no. 2, pp. 70-72, 2003.

[3] H. Zhai, X. Chen, Y. Fang, “How well can the IEEE 802.11 wireless LAN support quality of service?,” *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 3084-3094, 2005.

[4] IEEE, “Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment: medium access control (MAC) quality of service enhancements,” *IEEE Standard 802.11e*, 2005.

[5] S. Kim, R. Huang, and Y. Fang, “Deterministic priority channel access scheme for QoS support in IEEE 802.11e wireless LANs,” *IEEE Transactions on Vehicular Technology*, vol. 58, no. 2, pp. 855-864, 2009.

[6] Q. Deng and A. Cai, “A TXOP-based scheduling algorithm for video transmission in IEEE 802.11e networks,” *Proceedings of the 6th International Conference on ITS Telecommunications*, pp. 573-576, 2006.

[7] N. Cranley and M. Davis, “An experimental investigation of IEEE 802.11e TXOP facility for real-time video streaming,” *Proceedings of IEEE Global Telecommunications Conference*, pp. 2075-2080, 2007.

[8] N.S. Shankar and M. Schaar, “Performance analysis of video transmission over IEEE 802.11a/e WLANs,” *IEEE Transactions on Vehicular Technology*, vol. 56, no. 4, pp. 2346-2362, 2007.

[9] J. Zhu, A. Fapojuwo, “A new call admission control method for providing desired throughput and delay performance in IEEE802.11e wireless LANs,” *IEEE Transactions on Wireless Communications*, vol. 6, no. 2, pp. 701-709, 2007.

[10] X. Chen, H. Zhai, and Y. Fang, “Supporting QoS in IEEE 802.11e wireless LANs,” *IEEE Transactions on Wireless Communications*, vol. 5, no. 8, pp. 2217-2227, 2006.

[11] G. Boggia, P. Camarda, L.A. Grieco, and S. Mascolo, “Feedback-based control for providing real-time services with the 802.11e MAC,” *IEEE/ACM Transactions on Networking*, vol. 15, no. 2, pp. 323-333, 2007.

[12] S. Kim and Y.J. Cho, “Channel time allocation scheme based on feedback information in IEEE 802.11e wireless LANs,” *Elsevier Computer Networks*, vol. 51, no. 10, pp. 2771-2787, 2007.

[13] J. Hu, G. Min, and M.E. Woodward, “A threshold-based dynamic TXOP scheme for intra-AC QoS differentiation in IEEE 802.11e networks,” *Proceedings of the 14th IEEE International Conference on Computational Science and Engineering*, pp. 526-531, 2011.

[14] J.Y. Lee, H.Y. Hwang, J. Shin, and S. Valaee, “Distributed optimal TXOP control for throughput requirements in IEEE 802.11e wireless LAN,” *IEEE PIMRC*, 2011.

Authors Profile

Young-Woo Nam received the B.S. and M.S. degrees in software engineering from Kumoh National Institute of Technology, Gumi, Korea, in 2010 and 2012, respectively. His research interests include wireless LANs, wireless mesh networks, quality of service enhancement.

Sunmyeng Kim received the B.S., M.S., and Ph.D. degrees in information and communication from Ajou University, Suwon, Korea, in 2000, 2002, and 2006, respectively. From May 2006 to February 2008, he was a Postdoctoral Researcher in electrical and computer engineering with the University of Florida, Gainesville. In March 2008, he then joined the Department of Computer Software Engineering, Kumoh National Institute of Technology, Gumi, Korea, as an assistant professor. His research interests include resource management, wireless LANs and PANs, wireless mesh networks, and quality of service enhancement.

Si-Gwan Kim received the B.S. degree in Computer Science from Kyungpook Nat’l University in 1982 and M.S. and Ph.D. degrees in Computer Science from KAIST, Korea, in 1984 and 2000, respectively. He worked for Samsung Electronics until 1988 and then joined the Department of Computer Software Engineering, Kumoh National Institute of Technology, Gumi, Korea, as an associate professor. His research interests include sensor networks, mobile programming and parallel processing.

Hybrid Feature Extraction Technique for Face Recognition

Sangeeta N. Kakarwal
Department of Computer Science and Engineering
P.E.S. College of Engineering
Aurangabad, India

Ratnadeep R. Deshmukh
Department of Computer Science and IT
Dr. Babasaheb Ambedkar Marathwada University
Aurangabad, India

***Abstract*— This paper presents novel technique for recognizing faces. The proposed method uses hybrid feature extraction techniques such as Chi square and entropy are combined together. Feed forward and self-organizing neural network are used for classification. We evaluate proposed method using FACE94 and ORL database and achieved better performance.**

Keywords-Biometric; Chi square test; Entropy; FFNN; SOM.

I. INTRODUCTION

Face recognition from still images and video sequence has been an active research area due to both its scientific challenges and wide range of potential applications such as biometric identity authentication, human-computer interaction, and video surveillance. Within the past two decades, numerous face recognition algorithms have been proposed as reviewed in the literature survey. Even though we human beings can detect and identify faces in a cluttered scene with little effort, building an automated system that accomplishes such objective is very challenging. The challenges mainly come from the large variations in the visual stimulus due to illumination conditions, viewing directions, facial expressions, aging, and disguises such as facial hair, glasses, or cosmetics [1].

Face Recognition focuses on recognizing the identity of a person from a database of known individuals. Face Recognition will find countless unobtrusive applications such as airport security and access control, building surveillance and monitoring Human-Computer Intelligent interaction and perceptual interfaces and Smart Environments at home, office and cars [2].

Within the last decade, face recognition (FR) has found a wide range of applications, from identity authentication, access control, and face-based video indexing/ browsing; to human-computer interaction. Two issues are central to all these algorithms: 1) feature selection for face representation and 2) classification of a new face image based on the chosen feature representation. This work focuses on the issue of feature selection. Among various solutions to the problem, the most successful are those appearance-based approaches, which generally operate directly on images or appearances of face objects and process the images as two-dimensional (2-D) holistic patterns, to avoid difficulties associated with three-dimensional (3-D) modeling, and shape or landmark detection [3]. The initial idea and early work of this research have been published in part as conference papers in [4], [5] and [6].

A recognition process involves a suitable representation, which should make the subsequent processing not only computationally feasible but also robust to certain variations in images. One method of face representation attempts to capture and define the face as a whole and exploit the statistical regularities of pixel intensity variations [7].

The remaining part of this paper is organized as follows. Section II extends to the pattern matching which also introduces and discusses the Chi square test, Entropy and FFNN and SOM in detail. In Section III, extensive experiments on FACE94 and ORL faces are conducted to evaluate the performance of the proposed method on face recognition. Finally, conclusions are drawn in Section IV with some discussions.

II. PATTERN MATCHING

A. Pattern Recognition Methods

During the past 30 years, pattern recognition has had a considerable growth. Applications of pattern recognition now include: character recognition; target detection; medical diagnosis; biomedical signal and image analysis; remote sensing; identification of human faces and of fingerprints; machine part recognition; automatic inspection; and many others.

Traditionally, Pattern recognition methods are grouped into two categories: structural methods and feature space methods. Structural methods are useful in situation where the different classes of entity can be distinguished from each other by structural information, e.g. in character recognition different letters of the alphabet are structurally different from each other. The earliest-developed structural methods were the syntactic methods, based on using formal grammars to describe the structure of an entity [8].

The traditional approach to feature-space pattern recognition is the statistical approach, where the boundaries between the regions representing pattern classes in feature space are found by statistical inference based on a design set of sample patterns of known class membership [8]. Feature-space methods are useful in situations where the distinction between different pattern classes is readily expressible in terms of numerical measurements of this kind. The traditional goal of feature extraction is to characterize the object to be recognized by measurements whose values are very similar for objects in the same category, and very different for objects in different categories. This leads to the idea of seeking distinguishing features that are invariant to irrelevant transformations of the input. The task of the classifier component proper of a full system is to use the feature vector provided by the feature extractor to assign the object to a category [9]. Image classification is implemented by computing the similarity score between a target discriminating feature vector and a query discriminating feature vector [10].

B. Chi Square Test

Chi-square is a non-parametric test of statistical significance for analysis. Any appropriately performed test of statistical significance lets you know the degree of confidence you can have in accepting or rejecting a hypothesis. Typically, the hypothesis tested with Chi Square is whether or not two different samples (of people, texts, whatever) are different enough in some characteristic or aspect of their behavior that we can generalize from our samples that the population from which our samples are drawn are also different in the behavior or characteristics.

On the basis of hypothesis assumed about the population, we find the expected frequencies Ei (I =1,2,…,n), corresponding to the observed frequencies Oi(i=1,2,…,n) such that $\sum$ Ei = $\sum$ Oi. It is known that

$$\chi^2 = \sum_{i=1}^{n} \frac{(Oi-Ei)^2}{Ei}$$

follows approximately a $\chi 2$ - distribution with degrees of freedom equal to the number of independent frequencies. To test the goodness of fit, we have to determine how far the difference between Oi and Ei can be attributed to fluctuations of sampling and when we can assert that the differences are large enough to conclude that the sample is not a simple sample from the hypothetical population[11][12].

C. Entropy

The entropy is equivalent (i.e., monotonically functionally related) to the average minimal probability of decision error and is related to randomness extraction. For a given fuzzy sketch construction, the objective is then to derive a lower bound on the min entropy of the biometric template when conditioned on a given sketch, which itself yields an upper bound on the decrease in the security level measured as the min-entropy loss, which is defined as the difference between the unconditional and conditional min entropies [13] Shannon gave a precise mathematical definition of the average amount of information conveyed per source symbol, which is termed as Entropy [14].

Consider two random variables and having some joint probability distribution over a finite set. The unconditional uncertainty of can be measured by different entropies, the most famous of which is the Shannon entropy. Some of them have been given practical interpretations, e.g., the Shannon entropy can be interpreted in terms of coding and the min entropy in terms of decision making and classification [15]

Entropy is a statistical measure that summarizes randomness. Given a discrete random variable, its entropy is defined by

$$H(X) = -E_x[\log P(X)]$$

$$= -\sum_{xi \in \Omega x} P(X = xi) \log(X = xi) \qquad \ldots(1)$$

Where Ωx is the sample space and xi is the member of it. P(X=xi) represents the probability when X takes on the value xi. We can see in (1) that the more random a variable is, the more entropy it will have.

D. Artificial Neural Network

In recent years, there has been an increase in the use of evolutionary approaches in the training of artificial neural networks (ANNs). While evolutionary techniques for neural networks have shown to provide superior performance over conventional training approaches, the simultaneous optimization of network performance and architecture will almost always result in a slow training process due to the added algorithmic complexity [16].

1) Feed Forward Network

Feed forward networks may have a single layer of weights where the inputs are directly connected to the output, or multiple layers with intervening sets of hidden units. Neural networks use hidden units to create internal representations of the input patterns [17].

A Feed forward artificial neural network consists of layers of processing units, each layer feeding input to the next layer in a Feed forward manner through a set of connection weights or strengths. The weights are adjusted using the back propagation learning law. The patterns have to be applied for several training cycles to obtain the output error to an acceptable low value.

The back propagation learning involves propagation of the error backwards from the input training pattern, is determined by computing the outputs of units for each hidden layer in the forward pass of the input data. The error in the output is propagated backwards only to determine the weight updates [18]. FFNN is a multilayer Neural Network, which uses back propagation for learning.

As in most ANN applications, the number of nodes in the hidden layer has a direct effect on the quality of the solution. ANNs are first trained with a relatively small value for hidden nodes, which is later increased if the error is not reduced to acceptable levels. Large values for hidden nodes are avoided since they significantly increase computation time [19].

The Back propagation neural network is also called as generalized delta rule. The application of generalized delta rule at any iterative step involves two basic phases. In the first phase, a training vector is presented to the network and is allowed to propagate through the layers to compute output for each node. The output of the nodes in the output layers is then compared against their desired responses to generate error term. The second phase involves a backward pass through a network during which the appropriate error signal is passed to each node and the corresponding weight changes are made. Common practice is to track network error, as well as errors associated with individual patterns. In a successful training session, the network error decreases with the number of iterations and the procedure converges to a stable set of weights that exhibit only small fluctuations with additional training. The approach followed to establish whether a pattern has been classified correctly during training is to determine whether the response of the node in the output layer associated with the pattern class from which the pattern was obtained is high, while all the other nodes have outputs that are low [20].

Backpropogation is one of the supervised learning neural networks. Supervised learning is the process of providing the network with a series of sample inputs and comparing the output with the expected responses. The learning continues until the network is able to provide the expected response. The learning is considered complete when the neural network reaches a user defined performance level. This level signifies that the network has achieved the desired accuracy as it produces the required outputs for a given sequence of inputs [21].

2) Self Organizing Map

The self-organizing map, developed by Kohonen, groups the input data into cluster which are, commonly used for unsupervised training. In case of unsupervised learning, the target output is not known [17].

In a self-organizing map, the neurons are placed at the nodes of a lattice that is usually one or two dimensional. Higher dimensional maps are also possible but not as common. The neurons become selectively tuned to various input patterns or classes of input patterns in the course of a competitive learning process. The locations of the neurons so tuned (i.e., the wining neurons) become ordered with respect to each other in such a way that a meaningful coordinate system for different input features is created over the lattice. A self-organizing map is therefore characterized by the formation of a topographic map of the input patterns in which the spatial locations of the neurons in the lattice are indicative of intrinsic statistical features contained in the input patterns, hence the name "self-organizing map"[22]. The algorithm of self-organizing map is given below:

Algorithm SelfOrganize;
- Select network topology;
- Initialize weights randomly; and select $D(0)>0$;
- While computational bounds are not exceeded, do
  1. Select an input sample i_l;
  2. Find the output node j^* with minimum $\sum_{k=1}^{n} (i_{l,k}(t)-w_{j,k}(t))^2$;
  3. Update weights to all nodes within a topological distance of $D(t)$ from j^*, using
     $w_j(t+1)= w_j(t) +\eta(t)(i_l(t)-w_j(t))$,
     where $0< \eta(t)\leq \eta(t-1)\leq 1$;
  4. Increment t;

  End while.

Figure 1. Algorithm of Self Organizing Map

III. Experimental Results And Discussion

In order to assess the efficiency of proposed methodology which is discussed above, we performed experiments over Face94 and ORL dataset using FFNN and SOM neural network as a classifier.

A. Face94 Dataset

Face94 dataset consist of 20 female and 113 male face images having 20 distinct subject containing variations in illumination and facial expression. From these dataset we have selected 20 individuals consisting of males as well as females [23].

Face94 dataset used in our experiments includes 250 face images corresponding to 20 different subjects. For each individual we have selected 15 images for training and 5 images for testing.

Figure 2. Some Face Images from FACE94 Database

B. ORL

The Olivetti Research Lab (ORL) Database [4] of face images provided by the AT&T Laboratories from Cambridge University has been used for the experiment. It was collected between 1992 and 1994. It contains slight variations in illumination, facial expression (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). It is of 400 images, corresponding to 40 subjects (namely, 10 images for each class). Each image has the size of 112 x 92 pixels with 256 gray levels. Some face images from the ORL database are shown in figure3

For both database, we selected 50 images for testing genuine as well imposter faces. To extract the facial region, the images are normalized. All images are gray-scale images.

Figure 3. Some Face images from ORL Database

C. Steps used in Face Recognition

- Read input image, convert it into gray scale image then resize it to 200x180 pixels.
- Divide image into 4x4 blocks of 50x45 pixels.
- Obtain hybrid features from face by combining values of Chi Square test and Entropy together.
- Classify the images by Feed forward neural network and Self organizing map neural network.
- Analyse the performance by computing FAR and FRR.

D. Performance Evaluation

The accuracy of biometric-like identity authentication is due to the genuine and imposter distribution of matching. The overall accuracy can be illustrated by False Reject Rate (FRR) and False Accept Rate (FAR) at all thresholds. When the parameter changes, FAR and FRR may yield the same value, which is called Equal Error Rate (EER). It is a very important indicator to evaluate the accuracy of the biometric system, as well as binding of biometric and user data [25].

A typical biometric verification system commits two types of errors: false match and false non-match. Note that these two types of errors are also often denoted as false acceptance and false rejection; a distinction has to be made between positive and negative recognition; in positive recognition systems (e.g., an access control system) a false match determines the false acceptance of an impostor, whereas a false non-match causes the false rejection of a genuine user. On the other hand, in a negative recognition application (e.g., preventing users from obtaining welfare benefits under false identities), a false match results in rejecting a genuine request, whereas a false non-match results in falsely accepting an impostor attempt.

The notation "false match/false non-match" is not application dependent and therefore, in principle, is preferable to "false acceptance/false rejection." However, the use of false acceptance rate (FAR) and false rejection rate (FRR) is more popular and largely used in the commercial environment [26].

Traditional methods of evaluation focus on collective error statistics such as EERs and ROC curves. These statistics are useful for evaluating systems as a whole. Equal-Error Rate (EER) denotes the error rate at the threshold t for which false match rate and false non-match rate are identical: FAR(t) = FRR(t) [27].

FAR and FRR values for all persons with different threshold values. The FRR and FAR for number of participants (N) are calculated as specified in Eq. (2) and in equation Eq. (3) [28]:

$$FRR = \frac{1}{N}\sum_{n=1}^{N} FRR(n) \qquad \ldots (2)$$

$$FAR = \frac{1}{N}\sum_{n=1}^{N} FAR(n) \qquad \ldots (3)$$

When the experiment was carried out on ORL database 96% result is obtained with FFNN. In case of FACE94 database, the result obtained with SOM is 94%. Table1 and Table2 give the performance of hybrid feature extraction technique for FFNN and SOM respectively.

In addition to this experimentation was also carried out to recognize impostor faces. Graph1 and Graph2 illustrate the result of genuine and impostor face recognition.

Conclusion

This paper investigates the feasibility and effectiveness of face recognition with Chi square test and Entropy. Face recognition based on Chi square test and Entropy is performed by supervised and unsupervised network. Experimental results on Face94 and ORL database demonstrate that the proposed methodology outperforms in recognition.

TABLE I. PERFORMANCE OF FACE RECOGNITION FOR CHI SQUARE TEST+ENTROPY AND FFNN

Face database	**No. of test faces recognized**	**Rate of recognition**	**FRR**	**No. of impostor faces recognized**	**Rate of recognition**	**FAR**
FACE 94	46	92	0.08	39	78	0.22
ORL	48	96	0.04	26	52	0.48

TABLE II. PERFORMANCE OF FACE RECOGNITION FOR CHI SQUARE TEST+ ENTROPY AND SOM

Face database	**No. of test faces recognized**	**Rate of recognition**	**FRR**	**No. of test impostor faces recognized**	**Rate of recognition**	**FAR**
FACE 94	47	94	0.06	35	70	0.3
ORL	40	80	0.2	26	52	0.48

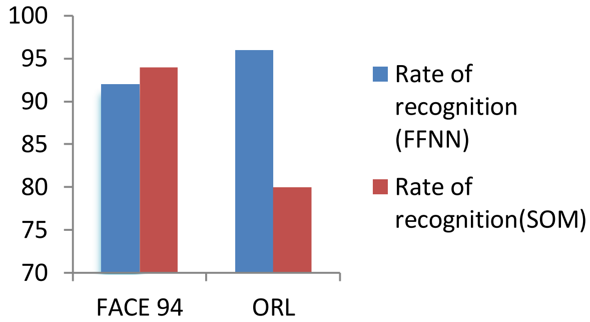


Figure 4. Graph1: Performance of Genuine faces using Chi Square+Entropy for FFNN and SOM

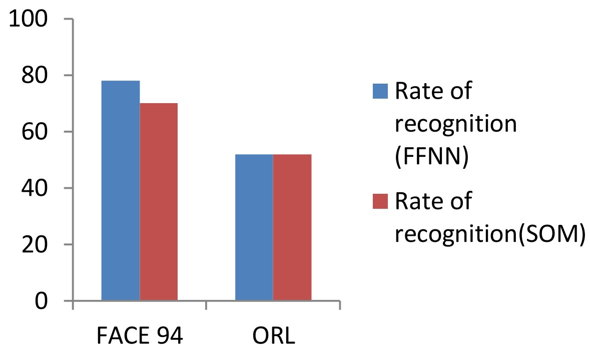


Figure 5. Graph2: Performance of Imposter faces using Chi Square+Entropy for FFNN and SOM

REFERENCES

[1] Yu Su, Shiguang Shan, Xilin Chen, and Wen Gao, "Hierarchical Ensemble of Global and local Classifiers for Face Recognition", IEEE Transactions On Image Processing,Vol. 18, No. 8, August 2009, pp 1885 – 1896.

[2] Chengjun Liu, Harry Wechsler," Independent Componenet Analysis of Gabor Features for Face Recognition", IEEE Transactions on Neural Networks, Vol. 14, July 2003, pp 919-928.

[3] Juwei Lu, Konstantinos N. Plataniotis, Anastasios N. Venetsanopoulos: Face Recognition Using Kernel Direct Discriminant Analysis Algorithms, IEEE Transactions On Neural Networks, Vol. 14, No. 1, January 2003, pp 117-126.

[4] S.N. Kakarwal, S.D. Sapkal, P.J.Ahire, Dr. D.S. Bormane: Analysis of Facial Image Classification using Discrete Wavelet Transform, Proc. International Conference ICSCI, 2007,pp 700-705.

[5] S.N. Kakarwal, Mahananda Malkauthekar, Shubhangi Sapkal, Dr. Ratnadeep Deshmukh: Face Recognition using Fourier Descriptor and FFNN, Proc. IEEE International Advance Computing Conference, 2009, pp 2740-2742.

[6] S.N. Kakarwal, Dr. R.R. Deshmukh: Wavelet Transform based Feature Extraction for Face Recognition, pp 100-104, IJCSA Issue-I June 2010, ISSN 0974-0767

[7] Bai-Ling Zhang, Haihong Zhang and Shuzhi Sam Je: Face Recognition by Applying Subband Representation and Kernel Associative Memory, IEEE Transaction on Neural Networks, Vol. 15, Jan 2004, pp 166-177.

[8] Daisheng Luo: Pattern Recognition and Image Processing (Horwood Publishing Limited 1998), pp 2-3

[9] Richard O. Duda, Peter E. Hart, David G. Stork: Pattern Classification (John Wiley 2001), pp 11-12

[10] Chengjun Liu: Learning the Uncorrelated, Independent, and Discriminating Color Spaces for Face Recognition, IEEE Transactions On Information Forensics and Security, Vol. 3, No. 2, June 2008, pp 213-222.

[11] T. Veerarajan, "Probability, Statistics and Random Processes", TMH, 2003,pp. 311-312.

[12] Chuang, K.-S.; Huang, H.K., Comparison of Chi- Square and Join-Count Methods for Evaluating Digital Image Data, IEEE Transaction on Medical Imaging, Vol. 11, No. 1, March 1992, 33.

[13] Su Hongtao, David Dagan Feng, Zhao Rong-chun, Wang Xiu-ying, "Face Recognition Method Using Mutual Information and Hybrid Feature", 0-7695- 1957-1/03 © 2003 IEEE.

[14] Richard Wells, Applied Coding and Information Theory for engineers, Pearson Education, pp. 10

[15] Jovan Dj. Golic´ and Madalina Baltatu, "Entropy Analysis and New Constructions of Biometric Key Generation Systems", IEEE Transactions On Information Theory, Vol. 54, No. 5, May 2008, pp.2026-2040

[16] Chi-Keong Goh, Eu-Jin Teoh, and Kay Chen Tan, Hybrid Multiobjective Evolutionary Design for Artificial Neural Networks, IEEE Transactions on Neural Networks, Vol. 19, no. 9, September 2008

[17] S.N. Sivanandanam, S. Sumathi, S. N. Deepa, Introduction to Neural Networks using MATLAB 6.0, TMH, pp 20

[18] B. Yegnanarayana, Artificial Neural Networks, PHI, pp. 117-120.

[19] Kevin Stanley McFall and James Robert Mahan, Artificial Neural Network Method for Solution of Boundary Value Problems with Exact Satisfaction of Arbitrary Boundary Conditions, IEEE Transactions on Neural Networks, Vol. 20, No. 8, pp. 1221-1233, August 2009

[20] Rafel C. Gonzalez, Richard E. Woods, Digital Image Processing, pp. 896

[21] S. Jayaraman, E. Esakkirajan, T. Veerakumar, Digital Image Processing, pp. 425

[22] Simon Haykin, Neural Networks, LPE, pp. 465-466

[23] Dr Libor Spacek Computer Vision Science Research Projects, Face94Dataset http://dces.essex.ac.uk/mv/allfaces /faces94.zip

[24] The ORL Database of Faces, Available: http://www.uk.research.att.com/facedatabase.html

[25] Stan Z. Li, Anil K. Jain, Encyclopedia of Biometrics, Springer, pp 75

[26] Davide Maltoni, Dario Maio, Anil K. Jain, Salil Prabhakar: Handbook of Fingerprint Recognition (Springer),pp 3

[27] Neil Yager and Ted Dunstone: The Biometric Menagerie, IEEE Transactions on Pattern Analysis and Machine Inte-ligence, VOL. 32, No. 2, FEBRUARY 2010, pp 220-230

[28] Website:http://www.bromba.com/faq/biofaqe

Improving Seek Time for Column Store Using MMH Algorithm

Tejaswini Apte
Sinhgad Institute of Business Administration and Research
Kondhwa(BK),Pune-411048

Dr. Maya Ingle
Devi Ahilya VishwaVidyalay
Indore

Dr. A.K.Goyal
Devi Ahilya VishwaVidyalay
Indore

***Abstract*—Hash based search has, proven excellence on large data warehouses stored in column store. Data distribution has significant impact on hash based search. To reduce impact of data distribution, we have proposed Memory Managed Hash (MMH) algorithm that uses shift XOR group for Queries and Transactions in column store. Our experiments show that MMH improves read and write throughput by 22% for TPC-H distribution.**

Keywords- Load; Selectivity; Seek; TPC-H; Algorithms; Hash.

I. INTRODUCTION

Searching in Column Store (CS) is greatly influenced by the address lookup process. Hashing algorithms have been widely adopted to provide fast address look-up process [2, 3, 8]. Bob Jenkins' hashing algorithm processes the key twelve octets at a time; the post processing step is slightly more complex because of handling of partial final block [14] in CS. However, it is possible to improve the throughput rate for fast address lookup in CS.

For various data warehouse applications, address lookup performs major role in performance measurement. The related and existing techniques of hashing and lookup are discussed in Section 2. Hash scan participates in performance of CS; Section 3 summarizes the hash scan for simple and complex queries. The proposed algorithm is an improved version of Jenkins' algorithm named as MMH. The informal and formal description of algorithm is discussed in Section 4. Case study was presented to show the effectiveness of our algorithm MMH with the help of implementation details in Section 5. Result analysis of MMH over Jenkins' is discussed in Section 6. Finally, we conclude with future work in Section 7.

I.RELATED WORK

Hashing has been used most successfully to avoid block conflicts in interleaved parallel memory systems used in multiprocessors and vector processors. Linear skewing functions, computes the block number using integer arithmetic [2, 3]. Stride patterns are mapped conflict-free when the stride and the number of memory blocks are relative primes [4].

To minimize the latency in computing per-block address, fragmentation was introduced in the Burroughs Scientific Processor, however it wastes 1/17th of the memory [5]. Fragmentation and complex block number computations are not necessary to obtain conflict free access to stride patterns. It has been observed that some particular types of XOR-based hash functions that are based on the division of binary polynomials, can simultaneously map a large number of stride-based patterns conflict-free [6]. XOR-based interleaving functions mainly focused on constructing a conflict-free hash function for several patterns complete with success [15, 8]. Bob Jenkins' hash produces uniformly distributed values for the hash tables [14]. However, literature reveals that there is a scope to improve the seek time of Jenkins algorithm for Column Store.

II. COLUMN STORE HASH SCAN

This section describes Hash Scan for simple and complex queries both for column store.

A. Hash Scan for Simple Queries

The complexity of hash scan is highly influenced by the size of data warehouse. Hash function may use partial or entire record as key to generate hash value. The parameters for hash based search are selectivity and cardinality for the given query. For shift XOR, with uniform distribution, if the key is having n values, probability density function (pdf) is:

Selectivity = n/ (number of distinct values)

Pdf (n)=1/ selectivity

B. Hash Scan for Complex Queries

Assume the given relation has multiple attributes stored in CS architecture. Let AK is the length of attribute, LID is the length of the tuple identifier or primary key and MROW is the matched row of second segment.

The number of seeks for given query is expressed as:

- Number of seeks required to retrieve tuples from the scanned segment.

  Number of seeks = ((numberOfRows) *AK + LID)* blocksize
- Number of seeks required to retrieve the remainder of the original tuples for those transactions which require it.

Number of seeks = ((numberOfRows) *AK + LID)* blocksize + ((MROW)*AK+LID)*blocksize

III. Proposed Algorithm - MMH

MMH algorithm is designed and tested with varying selectivity and cardinality of TPC-H distribution. The performance improvements could be demonstrated by executing following query on TPC-H schema with & without MMH algorithm.

```
select
        s_acctbal,
        s_name,
        n_name,
        p_partkey,
        p_mfgr,
        s_address,
        s_phone,
        s_comment
from
        part,
        supplier,
        partsupp,
        nation,
        region
where
        p_partkey = ps_partkey
        and s_suppkey = ps_suppkey
        and p_size = 29
        and p_type like '% BURNISHED TIN'
        and s_nationkey = n_nationkey
        and n_regionkey = r_regionkey
        and r_name = 'MIDDLE EAST'
        and ps_supplycost = (
                select
                        min(ps_supplycost)
                from
                        partsupp,
                        supplier,
                        nation,
                        region
                where
                        p_partkey = ps_partkey
                        and s_suppkey = ps_suppkey
                        and s_nationkey = n_nationkey
                        and n_regionkey = r_regionkey
                        and r_name = 'MIDDLE EAST'
        )
order by
        s_acctbal desc,
        n_name,
        s_name,
        p_partkey;
```

A. Informal Description

The proposed algorithm MMH is broadly designed with four functions:

- query(TPC-H-Q) Input parameter is a TPC-H query and return a valid sql query as output. This function is necessary to provide query for generation of hash value to improve search time.
- strHash(q) Input parameter is a valid sql query, this function uses CSXOR function to change the query to appropriate hash value. Primitive operations on database points to BUN heap, contains the atomic values inside the two columns. Fixed-sized atoms, reside directly in the BUN heap.
- HEAPalloc(d, size, 1) Input parameters are the memory heap and size. This function carries out checks for allocation of memory.
- CSXOR (h,s) Input parameters are memory heap and query. Execution generates hash value and is placed in passed heap.

B. Formal Description - MMH

/* Memory Managed Hash (*MMH*) Algorithm stores hash values in memory location for TPC-H schema query processing */

```
/* Main program begins */
main()
{
  s=query(TPC-H);
  strHash (char *s)
  {
     /* Declaration of variables */
    Heap d, size = 1<<10 * sizeof(stridx);
    /* Checking  allocation size */
     if  HEAPalloc(d, size, 1) >= 0)
     {
       d->free = 1<<10* sizeof(stridx);
     /* Declare and initialize Binary UNits      (BUN)*/
       BUN res=1<<10-1;
    /* Call a function with string s and store in BUN */
       res=CSXOR(d,s);
        /* stores hash values  in heap d with its base
         value, base position and free space*/
     memset(res->base, 0, res->free);
         }
  }
}
end.
/* End of main */
CSXOR(Heap h, const  char v)
/* This function performs string search with shift XOR operation;
with input parameters h as memory space and v as constant string; It
outputs generated hash values */
begin
{
 /* Declaration of variables */
 stridx_t *ref, *next;
/*Extend memory allocation by allocating more binary units BUN */
EXLEN=BUN size+1<<3-1;
size_t exlen=h->hashash?EXLEN:0;
/* Initializing binary units*/
 BUN off;
 off=1<<10-1;
/* Shift XOR operation for generating hash values and to search
            string */
   hkeyvalues=0;
  for i = h to v do
    {
      hkeyvalues ^=  (hkeyvalues << 10) +
                              (hkeyvalues  >> 7) +v;
       h.base=hkeyvalues;
     }
```

```
  /* Searching for string in  heap */
 for ref = h.base+off  to h.length do
 {
    ref=next;
    next=(h.base+ref);
   if STRCMP(v, (str) (next + 1) + exlen) == 0)
   return  ((sizeof(stridx_t) + *ref + exlen) >>3);
  }
 }
end.
```

IV. Case Study

MMH is designed from the shift-XOR class of hashing function. To support the hypothesis, we experimentally evaluate the MMH on real data sets i.e. TPC-H schema. In our experiments, we have focused on certain table sizes and load factors, to allow comparisons with original algorithm. We first investigated average search lengths for successful and unsuccessful search. The MMH results are compared to Jenkins' algorithm (Table 1 and Table 2). As can be seen, proposed algorithm performs better for TPC-H schema.

TABLE I. Response of Execution of Queries

TPC-H Queries	**Jenkins' time (in ms)**	**MMH time (in ms)**
2	196.959	92.73
3	1100	900
4	611.088	541.846
5	929.93	900
6	601.881	540.692
7	1000	900
8	227.257	196.401
9	855.851	672.403
10	4200	3000
12	850.772	727.156
16	423.415	305.074
17	396.955	258.854
19	3400	2500
21	1000	772

TABLE II. Comparison of Transaction Load Time

Relation	**Jenkins' (in ms)**	**MMH (in ms)**
Region	179.696	140
Nation	136.036	102.321
Supplier	272.617	202.886
Customer	1600	1000
Part	1600	1200
PartSupp	25300	18200
Orders	30000	17700
LineItem	140000	100000

II.Result Analysis

The proposed algorithm performs uniformly and efficiently independently of data size. From experiments with large sets of keys we have observed that with poorly chosen hashing function, performance can deteriorate markedly as the number of keys increases (Figure 1). Experimental results for the expected length of the load search time (LST) values vary significantly between runs. We chose a random set of TPC-H schema keys, the distribution of LST values is even narrower. MMH improvement to average LST is 30% on Red Hat Linux 2.4 GHz Intel processor and 1GB of RAM. (Figure 2). To our knowledge these are the first experiments testing these predictions.

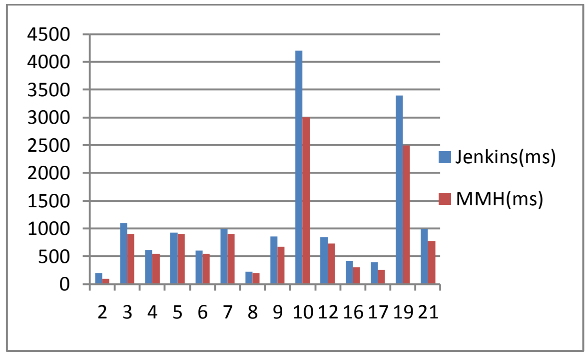


Figure 1. Result Analysis for Transaction Query Time

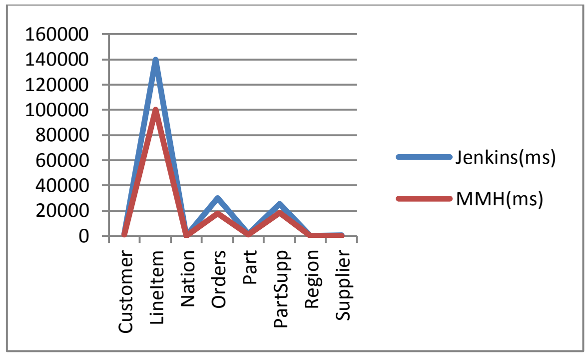


Figure 2. Figure 2: Result Analysis for Transaction Load Time

V. Conclusion and Future Work

The proposed algorithm is a generic search algorithm for CS data storage. The algorithm is designed specifically for use in query intensive environment. A key design principle of MMH to improve the throughput by minimizing the disk seeks. To achieve we used the hash function of shift-XOR class. We experimentally demonstrated gain in performance by MMH. The continued evolution of hard disk technology should make such performance advantages clearer in the future. The most obvious avenue for future work is an extension of MMH algorithm for multiple instances of CS. The most significant question that must be addressed when extending the MMH to a multi-instance environment is handling synchronization for various disks seeks.

References

[1] H. Vandierendonck and K. De Bosschere, "Randomized Caches for Power-Efficiency," IEICE Trans. Electronics, vol. E86-C, no. 10, pp. 2137-2144, 2003.

[2] D.J. Kuck, "ILLIAC IV Software and Application Programming," IEEE Trans. Computers, vol. 17, no. 8, pp. 758-770, Aug. 1968.

[3] G.S. Sohi, "High-Bandwidth Interleaved Memories for Vector Processors—A Simulation Study," IEEE Trans. Computers, vol. 42, no. 1, pp. 34-44, Jan. 1993.

[4] D.J. Kuck and R.A. Stokes, "The Burroughs Scientific Processor (BSP)," IEEE Trans. Computers, vol. 31, no. 5, pp. 363-376, May 1982.

[5] D.H. Lawrie and C.R. Vora, "The Prime Memory System for Array Access," IEEE Trans. Computers, vol. 31, no. 5, pp. 435-442, May 1982.

[6] B.R. Rau, “Pseudo-Randomly Interleaved Memory,” Proc. 18th Ann. Int’l Symp. Computer Architecture, pp. 74-83, May 1991.

[7] R. Raghavan and J.P. Hayes, “On Randomly Interleaved Memories,” SC90: Proc. Supercomputing ’90, pp. 49-58, Nov. 1990.

[8] J.M. Frailong, W. Jalby, and J. Lenfant, “XOR-Schemes: A Flexible Data Organization in Parallel Memories,” Proc. 1985 Int’l Conf. Parallel Processing, pp. 276-283, Aug. 1985.

[9] N. Topham and A. Gonza´lez, “Randomized Cache Placement for Eliminating Conflicts,” IEEE Trans. Computers, vol. 48, no. 2, pp. 185-192, Feb. 1999.

[10] Jenkins Bob "A hash function for hash Table lookup".

[11] A. Gonza´ lez, M. Valero, N. Topham, and J.M. Parcerisa, “Eliminating Cache Conflict Misses through XOR-Based Placement Functions,” Proc. 1997 Int’l Conf. Supercomputing, pp. 76-83, July 1997.

[12] Expected worst-case performance of hash files. Copmuter journal, Volume 25,Number 3, pages 347-352,1982

[13] VY. Lum. General performance analysis of key-to address transformations method using an abstract file concept. communication of the ACM, volume 16, Number10, pages 603-612,1973

A Discrete Event Barber Shop Simulation

Md. Masud Rana, Md. Jobayer Hossain
Dept. of Electronics & Communication Engineering,
Khulna University of Engineering & Technology
Khulna—9203, Bangladesh

Ratan Kumar Ghosh, Md. Mostafizur Rahman
Dept. of Electronics & Communication Engineering,
Khulna University of Engineering & Technology
Khulna—9203, Bangladesh

***Abstract*—A simulation based project is designed which can be practically implemented in a workspace (in this case, a barber shop). The design algorithm provides the user different time varying features such as number of people arrival, number of people being served, number of people waiting at the queue etc depending upon input criteria and system capability. The project is actually helpful for a modern barber shop in which a close inspection is needed. The owner or the manager of the shop can easily have a thorough but precise idea about the performance of the shop. At the same time he/she can easily decide to modify the arrangement and capability of the shop.**

Keywords- Normal distribution; Simulation.

I. INTRODUCTION

Simulation [10] is nowadays an accepted means to imitate the operation of real world systems. It has emerged as one of the dominant management science methods for the study and analysis of various types of systems such as communication system, traffic management system, hospital or a market management system etc. Due to complexity, stochastic relations and other situations lying in the system, it is not always possible to represent a real world system in a model form. One may think of an analytical model for such a system. But it essentially requires so many sampling assumptions to make the solution satisfactory for implementation. In such a situation the only alternative lying with sufficient reliabilities to the decision maker is simulation [1].

It is noticed that a graphical representation of output is more feasible for human brain to understand. Keeping in mind that thing, we designed the system so that it can provide different performances of the shop graphically. At first the algorithm asks for the day input. Then it takes input time i.e. the time at which the shop is open. After that it enters into corresponding time slot and performs logical operations and arithmetic calculations. The logical comparison serves the system features. For specific it performs the comparison between the no. of people arrival and the no. of server activated. From this comparison it determines the number of people rejected at any instant, number of people waiting at the queue, number of people being served and efficiency of the system.

In this paper the simulation model we used is presented in section II. It assumes general simulation process and features of the model. Section III shows the flow chart and the simulated result respectively. In section IV, we concluded about the design, its importance at practical case, limitations and also the possible improvements to the simulation model that are expected to be done in future.

II. SIMULATION MODEL

A simulation model [1] often takes the form of a collection of assumptions about the operation of a system. These assumptions are generally expressed as a mathematical or logical relation among the objects of interest in the system. Then the model is executed over time to obtain representative samples of performance measurements. This can be easily done using a computer [4, 7]. To have the best approximation of the mean of the performance measurements, we averaged the sample results. However, other factors such as the starting conditions of the simulation, the length of the period being simulated, and the accuracy of the model depend on how good our final estimate will be. The general simulation process [11] has the following steps:

- STATEMENT OF OBJECTIVES.
- MODEL DEVELOPMENT
- DATA COLLECTION AND PREPARATION
- SOLVING THE MODEL USING COMPUTER PROGRAM
- VERIFICATION
- VALIDATION
- DOCUMENTATION

In case of our design we at first developed the model to fulfill the objectives. To implement the design, we took computer program approach. We adopted MATLAB software to program.

In this simulation process the principle behind the methods is to develop a computer based analytical model. The model predicts the behavior of the system. Then, the model is evaluated, and therefore the behavior is predicted several times. Each evaluation (or called simulation cycle) is based on some randomly selected conditions for the input parameters of the system. We can use several analytical tools to ensure the random selection of the random parameters according to their individual probability distributions for each evaluation. Thus a number of predictions of the behavior are obtained.

At first the design algorithm takes day input. As our program architecture, the input may be of two types: 0 for Friday (off day in Bangladesh) and 1 for other days. The program pointer enters the corresponding day block and asks for time input. This is the time at which the shop is opened. Depending upon the time slot selected the program performs logical operations. Actually the program compares between number of server required and number of server available. Depending upon the comparison result the program calculates number of people arrival, number of people before giving service, number of people taking service, number of people waiting at the queue after servicing and efficiency of the system at any instant.

To determine number of people arriving at any instant we can adopt probabilistic method. There are different types of distribution functions [5] which can be used to fulfill our purpose. We may use one of the following distribution functions:

- Triangular Distribution function
- Uniform Density Function
- Cumulative Function
- Poisson Distribution Function
- Exponential Distribution Function
- Normal Distribution Function

For simplicity and compatibility with the system we used Normal distribution to construct the program code.

A. Normal Distribution Function

Normal distribution [3], also called Gaussian distribution is the most common distribution function for independent, randomly generated variables. It is often used in statistical reports, survey analysis, quality control resource allocation and other presentation schemes. It generally takes the form of a bell-shaped curve.

The graph of the normal distribution [16] is characterized by two parameters: mean and standard deviation. Mean or average of the normal distribution is the maximum value of the graph. The graph is always symmetric about the point of mean value. The rest one thing, standard deviation determines the amount of dispersion away from the mean. A small standard deviation compared with the mean produces a steep graph, whereas a large standard deviation produces a flat graph. The normal distribution can be expressed by the normal density function [6, 15].

$$p(x) = [\exp\text{-}(x - \mu)^2/2\sigma^2]/[\sigma\sqrt{2\pi}.]. \quad (1)$$

In the above equation, exponential function e is the constant 2.718281828…, μ is the mean, and σ is the standard deviation of the distribution. The probability of a random variable falling within any given range is proportional to the area enclosed by the graph and x-axis within that range. The denominator ($\sigma\sqrt{2\pi}$) is known as the normalizing coefficient. It causes the total area enclosed by the graph to be exactly equal to unity. So, probabilities can be obtained directly from the corresponding area—i.e., an area of 0.5 corresponds to a probability of 0.5.

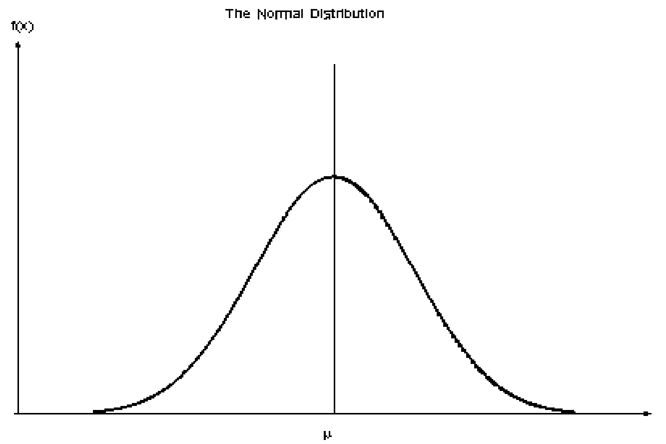


Figure 1. Example of a Graphical representation[17] of normal distribution

Working mathematically with normal distribution is easier than other distributions. It also provides satisfactory estimation performances. It is the main reason behind why normal distribution is used widely in many practical cases. But whether we can choose normal distribution or not is determined by the size of a sample N. If the sampling distribution corresponds to large N then we can approximate it by the normal distribution even if the population distribution itself is not normal.

B. Generation of Random Probability Value

Random numbers are always in the form of real values. If we normalize these values dividing by the largest possible value, results are the real values in the range [0, 1]. These numbers follow uniform distribution on the range [0, 1]. A complete set of random numbers should also satisfy the condition of non-correlation for the purpose of simulation use. The common types of random number generation [9] methods are:

- Mechanical
- Tabulated
- Computer based

We adopted the third one that runs recursive functions upon a seed.

C. Generation of Random Variables

Inverse Transformation Method [12]:

Inverse transformation method is a familiar random number generation method. The computer based recursive functions of the project use this method. The steps in inverse transform method are:

- A random number R is first generated in the range [0, 1]
- The value of a generated continuous random variable X, is determined as follows

$X=F^{-1}_X(R)$ = the inverse of the Normal distribution function of the random variable X evaluated at u.

Recall the normal probability density function

$$p(x) = (exp-(x-\mu)^2/2\sigma^2)/\sigma\sqrt{2\pi} \qquad (2)$$

In the above equation, σ = Standard deviation, μ = mean. For a single variable, $\sigma = x-\mu$. So putting this in (1) we get $x=\mu+(2*pi*r)^{-1}$. Here x is the random variable.

III. Simulated Outcome

The program executes according to the flowchart of figure 2. It asks for day and time input. Then as per the simulation model it performs necessary comparison, selection, estimation and calculations. Then it demonstrates different desired performance characteristics in graphical manner. Figure 3 shows the number of people arrival at any instant. This number is equal to the quantity that is found from normal distribution.

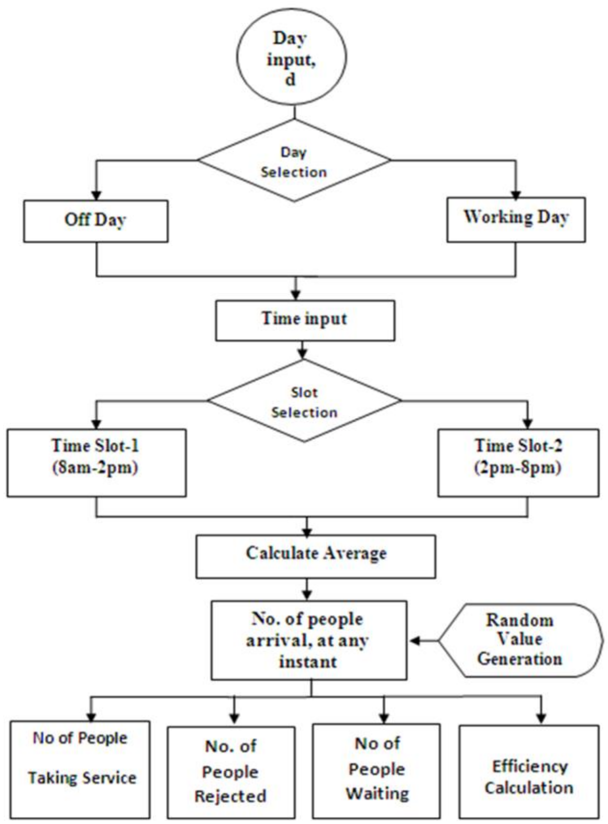


Figure 2. Flow Chart of Discrete Event Barber Shop Simulation

Figure 4 shows total number of people that is to be served. This is equal to the quantity that is found from normal distribution plus the previous people waiting for service. First we took the day input. According to the input day we divide the day into two slots. If input is =0 the algorithm realizes it as Friday. If input is =1 then algorithm realizes it as other days If input is =1 then algorithm realizes it as other days. The program pointer enters the corresponding day block and asks for time input. This is the time at which the shop is opened.

Depending upon the time slot selected the program performs logical operations specified for corresponding time slot. Actually, the program compares between number of server required and number of server available. Depending upon the comparison result the program calculates number of people arrival, number of people before giving service, number of people taking service, number of people waiting at the queue after servicing and efficiency of the system at any instant.

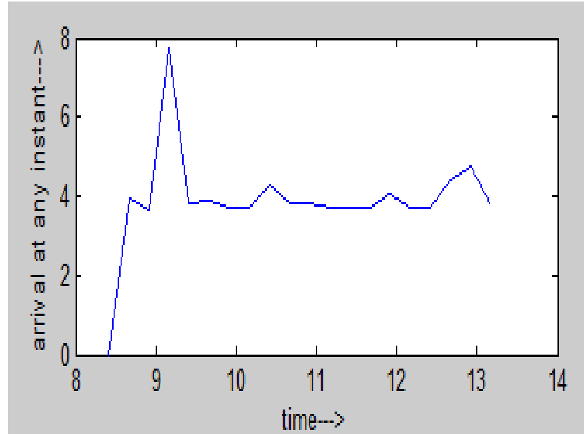


Figure 3. Depiction of no. of people arriving at any instant.

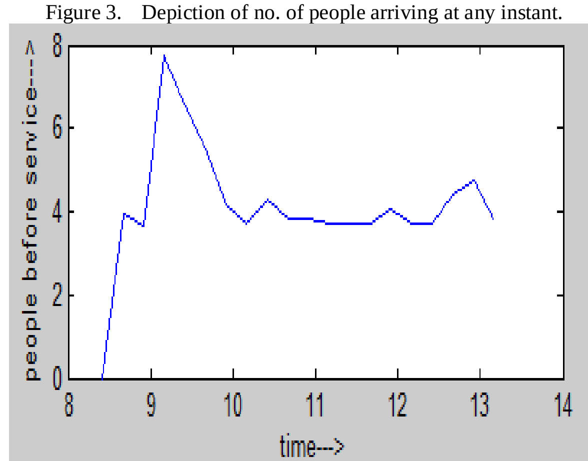


Figure 4. Depiction of no. of people at the queue before service

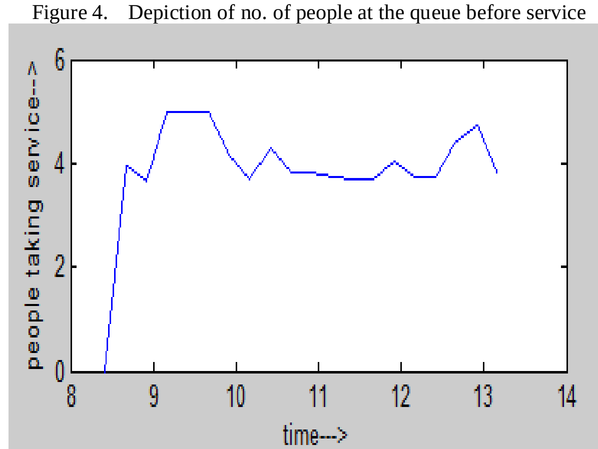


Figure 5. Demonstration of no. of people taking service

Fig 5 represents number of people that is being served. If total number of people is to be served is less than the capability of system then number of people is being served is equal to people arrival. If number of people to be served is greater than capability then the excess people will be rejected. This is shown in figure 6. Figure 7 shows the efficiency of the system.

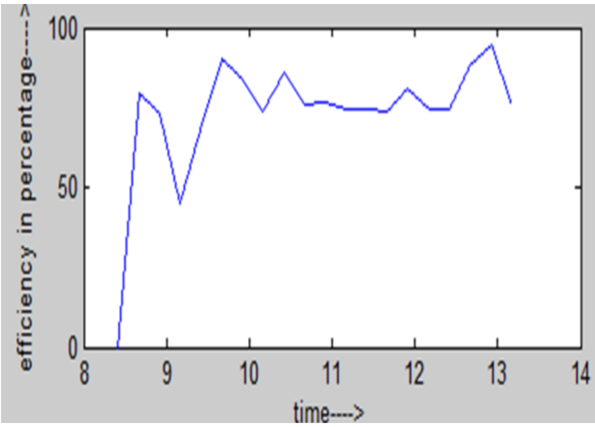


Figure 6. Demonstration of no. of people at the queue after service

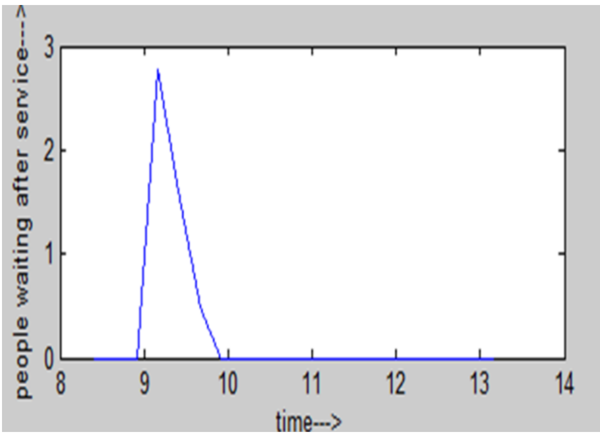


Figure 7. Shows the efficiency of the system at any instant

IV. CONCLUSION

The project we performed is actually helpful for a modern barber shop in which a close inspection is needed. The user can have a precise idea about the performance of the system. The logical development of this project can be easily understood.

The calculation of random variable (number of people arrival in this case) is done using normal distribution function. The average is specified in each slot programmed. Then comparing with the capability (no of server), simulation of the performance is done.

Although this simulation logic is not complex, it is very useful in practical case. The simulation plays vital role in a modern barber shop. Observing the graphical outcomes the owner or the manager of the shop can easily decide number of server at any instant and use the existing servers efficiently. He or she can also have highest efficiency observing the response of the system.

Again we assigned the average of the people arrival, at each time slot manually in the program code. The average can be calculated from the inspection of performances at several instants randomly. The average may be arithmetic mean, geometrical mean. This will make the simulation model more realistic and smarter.

REFERENCES

[1] Dr. Ibrahim. Assakkaf , Monte carlo simulation, fall-2003, Chapter 11a

[2] Deborah Rumsey Ph. D, “Probability for dummies

[3] Bikram Sharda, Scott J. Bury, “A discrete-event simulation model for reliability modeling of a chemical plant”

[4] Marco Semini, Hakon Fauske, Jan Ola Strandhagen, “Application of discrete-event simulation to support manufacturing logistic decision decision-making: A survey”

[5] Schaum’s outline of “Theory and problems of statistics”, international edition 2003-2004

[6] John Bird , “Engineering Mathmatics”, 4th edition

[7] Deborah Rumsey Ph. D , Probability for dummies, page 167

[8] “Henk Tijms , “Understanding Probability, Chance Rules in Everyday life”, 2nd edition

[9] Averill M. Law, W. David Kelton, “Simulation modeling and analysis”, third Edition

[10] G. Gordon, “System Simulation”, second edition

[11] Tranter, Shanmugam, Rappaport, “principles of Communication systems simulation with applications”

[12] Scheldon Ross, “A first Course in Probability”, 5th Edition ,page 455

[13] David Temperley,” Music and probability” copyright: 2007 Massachusetts Institute of Technology

[14] Henk Tijms , “Understanding Probability, Chance Rules in Everyday life”, 2nd edition

[15] www.mathworks.com

[16] www.britannica.com/EBchecked/topic/418227/normal-distribution

[17] http://www.google.com/imgres?imgurl=http://itl.nist.gov/div898

AUTHORS PROFILE

Md. Masud Rana is currently the Assistant Professor at Department of Electronics & Communication Engineering, Khulna University of Engineering & Technology, Bangladesh. He has authored and co-authored several journal/conference papers. His current research interests include broadband digital communications systems, channel estimation and tracking, interference prediction and cancellation, equalization, digital signal processing in communications, OFDM, cooperative communications as well as single and multicarrier transmission.

Dr. Md. Mostafizur Rahman was born in Satkhira, Bangladesh. He received his B.Sc. Engg. and M. Sc. Engg. Degrees in Electrical and Electronic Engineering from Bangladesh Institute of Technology, Rajshahi and Khulna University of Engineering and Technology, Bangladesh in 1992 and 2005 respectively. He took his Ph.D. degree Graduate School and Resource Science, Akita University, Japan. He worked in Bangladesh Civil Service (BCS) as radio engineer from 1995 to 2001.He is currently the head of Electronics and Communication Engineering department, Khulna University of Engineering and Technology, Bangladesh He is a member of IEEE and fellow IEB, Bangladesh

Md. Jobayer Hossain was born in Rajarhat, Kurigram, Bangladesh. He is currently a student of Electronics & Communication Engineering department in Khulna University of Engineering and Technology, Khulna-9203, Bangladesh

Now he is in 4th year and going to receive his B.Sc. degree within May 2012. He is interested in digital electronics, microelectronics, chip design, wireless and mobile communications. He is currently working on reconfigurable communication system on chip design

Ratan Kumar Ghosh is a student of Khulna University of Engineering and Technology (KUET), Khulna, Bangladesh, in the department of electronics & communication engineering. Currently, he is in 4th year and about to receive his B.Sc. degree (May 2012).

Speaker Identification using Frequency Dsitribution in the Transform Domain

Dr. H B Kekre
Senior Professor, Computer Dept.,
MPSTME, NMIMS University,
Mumbai, India.

Vaishali Kulkarni
Associate Professor, Electronics and Telecommunication,
MPSTME, NMIMS University,
Mumbai, India.

***Abstract*— In this paper, we propose Speaker Identification using the frequency distribution of various transforms like DFT (Discrete Fourier Transform), DCT (Discrete Cosine Transform), DST (Discrete Sine Transform), Hartley, Walsh, Haar and Kekre transforms. The speech signal spoken by a particular speaker is converted into frequency domain by applying the different transform techniques. The distribution in the transform domain is utilized to extract the feature vectors in the training and the matching phases. The results obtained by using all the seven transform techniques have been analyzed and compared. It can be seen that DFT, DCT, DST and Hartley transform give comparatively similar results (Above 96%). The results obtained by using Haar and Kekre transform are very poor. The best results are obtained by using DFT (97.19% for a feature vector of size 40).**

Keywords-Speaker Identification; DFT; DCT; DST; Hartley; Haar; Walsh; Kekre's Transform.

I. Introduction

Recently a lot of work is being carried out in the field of biometrics. There are several categories of biometrics like fingerprint, iris, face, palm, signature voice etc. Voice as a biometric has certain advantages over other biometrics like: it is easy to implement, no special hardware is required, user acceptability is more, and remote login is possible [1]. In spite of these advantages it has not been implemented to a very large extent because of the problems like security, changes in human voice etc. Human beings are able to recognize a person by hearing his voice. This process is called Speaker Identification. Speaker Identification falls under the broad category of Speaker Recognition [2 – 4], which covers Identification as well as Verification.

Speaker Identification (also known as closed set identification) is a 1: N matching process where the identity of a person must be determined from a set of known speakers [4 - 6]. Speaker Verification (also known as open set identification) serves to establish whether the speaker is who he claims to be [7]. Speaker Identification can be further classified into text-dependent and text-independent systems. In a text dependent system, the system knows what utterances to expect from the speaker. However, in a text-independent system, no assumptions about the text can be made, and the system must be more flexible than a text dependent system. Speaker Recognition systems have been developed for a wide range of applications like control access to restricted services, for example, for giving commands to computer, phone access to banking, database services, shopping or voice mail, and access to secure equipment [8 - 11]. Speaker Identification encompasses two main aspects: feature extraction and feature matching. Traditional methods of speaker recognition use MFCC (Mel Frequency Cepstral Coefficients) [13 – 16], LPC (Linear Predictive Coding) [12] for feature extraction. Feature matching has been done using Vector Quantization [17 – 21], HMM (Hidden Markov Model) [21 – 22], GMM (Gaussian Mixture Model) [23].

We have proposed Speaker Identification using row mean of DFT, DCT, DST and Walsh Transforms on the speech signal [24 – 25].We have proposed speaker recognition using the concept of row mean of the transform techniques on the spectrogram of the speech signal [26]. We have also proposed speaker identification using power distribution in the frequency domain [27 - 28].

In this paper we have extended the technique of power distribution of the frequency domain to four more transforms i.e. Hartley, Walsh, Haar and Kekre Transform. Here we have used the power distribution in the frequency domain to extract the features for the reference as well as test speech samples. The feature matching has been done using Euclidean distance. The various transform techniques have been explained in section II. In Section III, the feature vector extraction is explained. Results are discussed in section IV and conclusion ion section V.within parentheses, following the example.

II. Transform Techniques

The Transform when applied on a speech signal converts the converts it from time domain to frequency domain. In this paper seven different Transform techniques have been used. Let y(t) be the speech signal in the time domain and y0, y1, y2, yN-1 be the samples of y(t) in the time domain. The Discrete Fourier Transform of this signal is given by (1). The DFT is implemented using Fast Fourier Transform (FFT).

$$Y_k = \sum_{n=0}^{N-1} Y_n e^{-j2\pi kn/N} \qquad (1)$$

Where $y_n = y(n\Delta t)$ is the sampled value of continuous signal y(t); k= 0, 1, 2…, N-1.Δt is the sampling interval.

The discrete cosine transform which is closely related to the DFT has been used in compression because of its capability of reconstruction with a few coefficients.

The DCT of the signal y(t) can be given by (2) and w_k as given by (3).

$$Y_k = w_k \sum_{n=1}^{N} y_n \cos \frac{\pi(2n-1)(k-1)}{2N} \tag{2}$$

$$w_k = 1/\sqrt{N} \qquad \text{For } k=1 \tag{3}$$

$$= \sqrt{\frac{2}{N}} \qquad 2 \le k \le N$$

A discrete sine transform (DST) expresses a sequence of finitely many data points in terms of a sum of sine functions. The DST of the signal y(t) can be given by (4).

$$Y_k = \sum_{n=1}^{N} y(n) \sin(\pi \frac{kn}{N+1}) \tag{4}$$

The Walsh transform or Walsh–Hadamard transform is a non-sinusoidal, orthogonal transformation technique that decomposes a signal into a set of basis functions. These basis functions are Walsh functions, which are rectangular or square waves with values of +1 or –1.

The Walsh–Hadamard transform is used in a number of applications, such as image processing, speech processing, filtering, and power spectrum analysis. Like the FFT, the Walsh–Hadamard transform has a fast version, the fast Walsh–Hadamard transform (`fwht`). Compared to the FFT, the FWHT requires less storage space and is faster to calculate because it uses only real additions and subtractions, while the FFT requires complex values. The FWHT is able to represent signals with sharp discontinuities more accurately using fewer coefficients than the FFT. FWHT is a divide and conquer algorithm that recursively breaks down a WHT of size N into two smaller WHTs of size N / 2. This implementation follows the recursion of the definition 2N Hadamard 2N× matrix H_N *as given by (5).*

$$H_N = \frac{1}{\sqrt{2}} \begin{bmatrix} H_{N-1} & H_{N-1} \\ H_{N-1} & -H_{N-1} \end{bmatrix} \tag{5}$$

A discrete Hartley transform (DHT) is a real transform similar to the discrete Fourier transform (DFT). If the speech signal is represented by y(t) then the DHT is given by (6).

$$Y_k = \sum_{n=0}^{N-1} y_n [\cos(\frac{2\pi}{N} nk) + \sin(\frac{2\pi}{N} nk)] \tag{6}$$

The Haar transform is derived from the Haar matrix. The Haar transform is separable and can be expressed in matrix form as shown in (7).

$$[F] = [H].[f].[H]^T \tag{7}$$

Where [f] is an N×1 signal, [H] is an N×N Haar transform matrix and [F] is an N×1 transformed signal. The transformation H contains sampled version of the Haar basis function $h_k(t)$ which are defined over the continuous closed interval t Є [0, 1].

The Haar basis functions are

- When k=0, the Haar function is defined as a constant as in (8).

$$h_0(n) = 1/\sqrt{N} \tag{8}$$

- When k>0, the Haar function is defined as in (9).

$$h_k(n) = \begin{cases} 2^{p/2} & (q-1)/2^p \le t \le (q-0.5)/2^p \\ -2^{p/2} & (q-0.5)/2^p \le t \le q/2^p \\ 0 & \text{Otherwise} \end{cases} \tag{9}$$

Where $0 \le p < \log 2N$ and $1 \le q \le 2p$

For example, when N=4, we have H_4 as given by (10).

$$H_4 = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \end{bmatrix} \tag{10}$$

Kekre Transform matrix can be of any size N x N, which need not have to be in powers of 2 (as is the case with most of other transforms including Haar Transform). All upper diagonal and diagonal values of Kekre transform matrix are one, while the lower diagonal part except the values just below diagonal are zero. Generalized N×N Kekre Transform Matrix can be given as in (11). The formula for generating the term K_{xy} of Kekre transform matrix is given by (12).

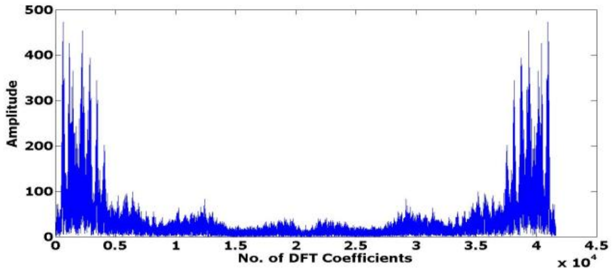
$$K_{N\times N} = \begin{bmatrix} 1 & 1 & 1 & .. & 1 & 1 \\ -N+1 & 1 & 1 & .. & 1 & 1 \\ 0 & -N+2 & 1 & .. & 1 & 1 \\ : & : & : & : & : & : \\ 0 & 0 & 0 & 0 & -N+(N-1) & 1 \end{bmatrix} \tag{11}$$

$$K_{xy} = \begin{cases} 1 & ; x \le y \\ -N+(x-1) & ; x = y+1 \\ 0 & ; x > y+1 \end{cases} \tag{12}$$

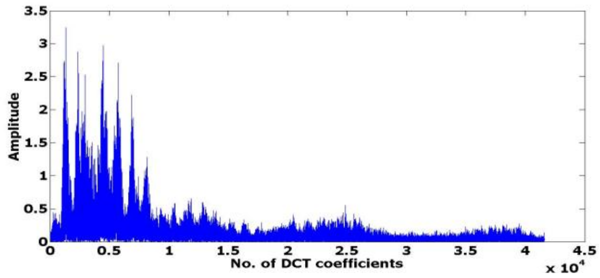
III. FEATURE EXTRACTION

The feature vector extraction process is described as below.

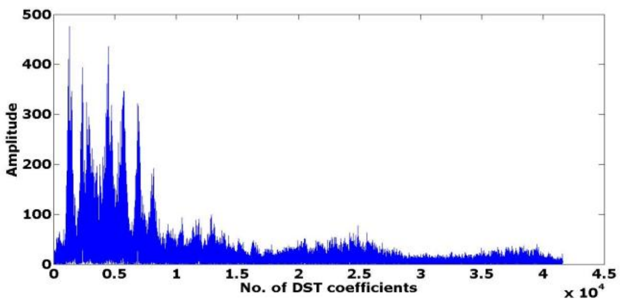
1. The speech signal was converted into frequency domain by applying the transform techniques described in section II, for three different lengths of speech signal. (8.192 sec, 4.096 sec and 2.048 sec) as it gives 2^{16}, 2^{15} and 2^{14} samples at 8 KHz sampling rate.
2. The magnitude of the signal in the transform domain was considered for feature extraction. Figure 1 shows the magnitude plot of the various transforms for the speech signal of length 8.192 sec.



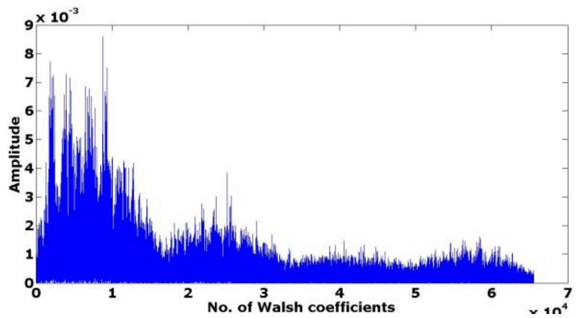
(A)



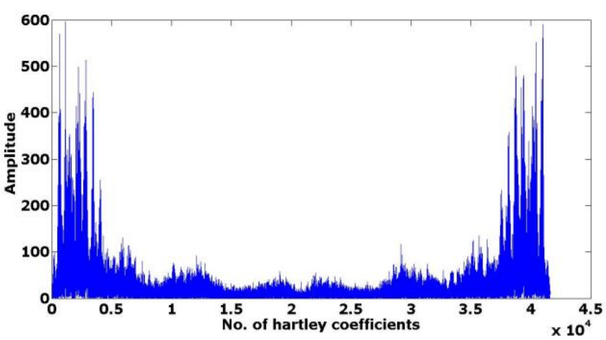
(B)



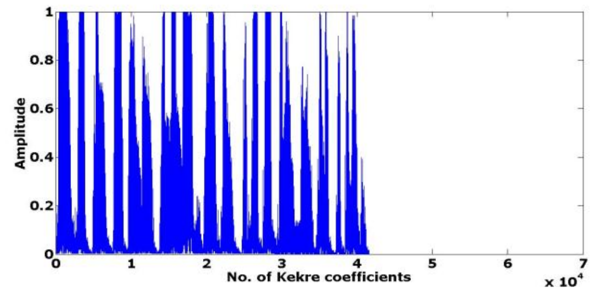
(C)



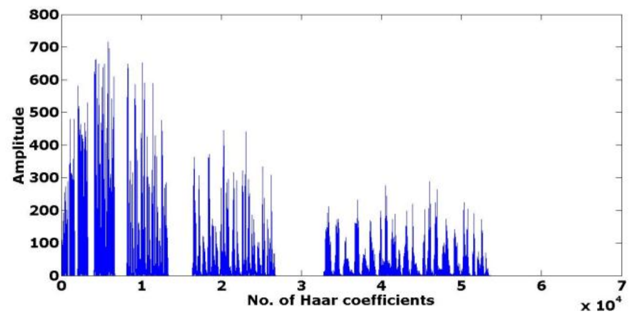
(D)



(E)



(F)



(G)

Figure 1. Frequency Spectrum of the different transforms. (A)FFT, (B) DCT (C) DST (D) Walsh (E) Hartley (F) Kekre (G) Haar

3. This was then divided into various groups and the sum of the magnitude for each group forms the feature vector.

IV. EXPERIMENTAL RESULTS

The speech samples used in this work are recorded using Sound Forge 4.5. The sampling frequency is 8000 Hz (8 bit, mono PCM samples). Table I shows the database description. The samples are collected from speakers of different age group ranging from 12 to 75 years. Five iterations of four different sentences of varying lengths are recorded from each of the speakers. Twenty samples per speaker are taken. For text dependent identification, four iterations of a particular sentence are kept in the database and the remaining one iteration is used for testing. These speech signals have an amplitude range of '-1' to '+1'.

TABLE I. DATABASE DESCRIPTION

Parameter	Sample characteristics
Language	English
No. of Speakers	107
Speech type	Read speech, microphone recorded
Recording conditions	Normal
Sampling frequency	8000 Hz
Resolution	8 bps

The simulation was done using MATLAB 7.7.0. For DFT, the FFT algorithm was used to calculate the transform coefficients. For DCT, DST and Walsh, the in-built functions in MATLAB were used. To calculate the Hartley Transform coefficients, first the FFT of the real part of speech signal was calculated and then the imaginary part of the complex transform was subtracted from its real part. This is shown in by (13).

$$Y1 = fft(y)$$
$$Y2 = real(Y1) - imaginary(Y1) \quad (13)$$

For calculating the Kekre Transform, the difficulty was to generate the Transform matrix of the order of 65536×65536, 32768×32768 and 16384×16384 which gave 'out of memory' error.

Instead of computing the transform matrix, the coefficients were calculated as given in (14).

$$S_1 = \sum_{n=0}^{N-1} y_n \qquad ; k = 0$$
$$S_k = S_1 - \sum_{n=0}^{N-2} y_n - (n-k) y_{N-1} \qquad ; 0<k\leq N-1 \quad (14)$$

For calculating the Haar Transform coefficients also, the same order of Transform matrix was required. Again here also, the problem was solved by directly calculating the coefficients using the butterfly diagram approach. Thus after transforming the signal into transform domain, the magnitude plot was generated as shown in figure 1. As can be seen from the magnitude plots, the energy concentration is in the lower order coefficients. This concept was utilized and the frequency spectrum was divided into groups and the sum of the magnitude for each group formed the feature vector. The feature vectors of all the reference speech samples were calculated for the different transforms and stored in the database in the training phase. In the matching phase, the test sample that is to be identified is taken and similarly processed as in the training phase to form the feature vector. The stored feature vector which gives the minimum Euclidean distance with the input sample feature vector is declared as the speaker identified. The accuracy of the identification system is calculated as given by (15).

$$Accuracy(\%) = \frac{no\_of\_samples\_identified}{Total\_no.\_of\_samples\_tested} \times 100 \quad (15)$$

The sentences in the database are of varying sizes. We have performed the simulations for three different lengths of the sentences. In the first case we considered only the first 2.048 sec (16384 samples) of the sentence for each speaker in the training as well as in the testing phase. Figure 2 shows the accuracy obtained for different Transforms for the speech signal of length 2.048 sec (16384 samples). We have begun by taking the entire spectrum as one group and then taking the sum of the magnitude as the feature vector. In this case there is only one element in the feature vector. As can be seen the accuracy is very less for all the transforms. For FFT we get an accuracy of around 6.54%. As we divide the spectrum into more number of groups and then take the sum of each group as the element of the feature vector, the accuracy goes on increasing. For FFT, the accuracy is 93.45% for a feature vector of size 56. Above a feature vector of size 56, the accuracy decreases and we an accuracy of 92.52% for a feature vector of size 88. DCT and DST also show a similar trend, with a maximum accuracy of 89.71% for a feature vector of size 40. With Walsh transform though the trend is similar, the maximum accuracy is only 79.43% for a feature vector of size 80. Hartley transform shows a behavior similar to FFT and the maximum accuracy is 93.45% for a feature vector of size 56. As can be seen from the magnitude spectrum also, the energy compaction in case of Kekre transform and Haar transform is less than other transforms. This explains the lower performance

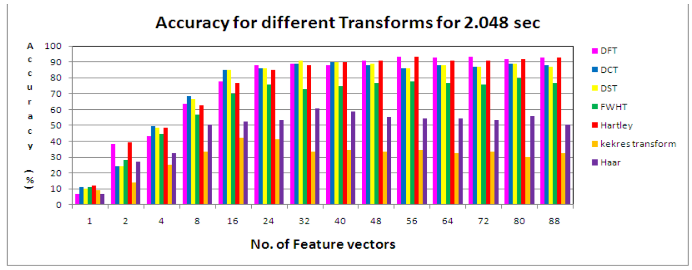


Figure 1. Accuracy for different Transforms for 2.048 sec

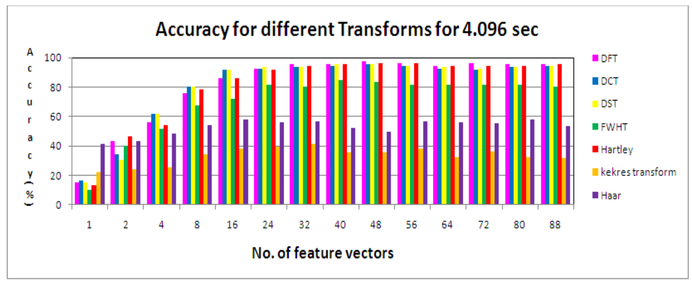


Figure 2. Accuracy for different Transforms for 4.096 sec

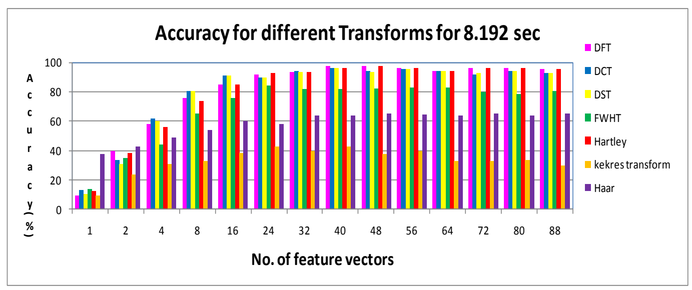


Figure 3. Accuracy of different Transforms for 8.192 sec

for both the transforms, Kekre transform 41.12% and Haar transform 60.74%. For the second set of simulations, the first 4.096 sec of the sentence spoken by each speaker was considered in the training as well as in the testing phase. Figure 3 shows the results obtained for this set of experiments. As can be seen from figure 3, the overall trend shown by each transform is the same as in figure 2. But here the effect of the increase in length of the speech signal considered is that the accuracy increases. With FFT, the maximum accuracy 97.19% for a feature vector of size 48. For DCT and DST, the maximum accuracy is 95.32% for a feature vector of size 48. With Walsh transform, the maximum accuracy is now around 85%. Hartley transform gives a maximum accuracy of 96.26% for a feature vector of size 48. There is no significant improvement as far as the Kekre transform and Haar transform are considered. Overall there is a gain in accuracy by increasing the length of the speech signal under consideration. Figure 4 shows the results obtained by increasing the length of the speech signal to 8.192 sec (64536 samples). If the length of the speech signal is smaller than 8.192 sec, then it is padded with zeros to make them all of equal length. As can be seen from the results, there is not much gain over that obtained by considering 4.096 sec. the maximum accuracy is still 97.19% for FFT with feature vector of size 40 now. The trend shown by all the transforms remains the same.

The overall results indicate that the accuracy increases with the increase in the size of feature vector up to a certain point and then it decreases. FFT, DCT, DST and Hartley transforms give very good results. Walsh gives comparatively lower results. Haar and Kekre transform give lesser accuracy compared to all other transforms. This technique of using the magnitude spectrum is very simple to implement and gives comparable results with the traditional techniques used for speaker identification. For the present study we have not used any preprocessing techniques for the speech signal. The database is collected using different brands of locally available microphones under normal conditions. This shows that the results obtained are independent of the recording instrument specifications.

V. Conclusion and Future Scope

In this paper we have shown a comparative performance of speaker identification by using seven different transform techniques. The approach used in this work is entirely different from the studies which have been done in this area. Here we are simply using the distribution in the magnitude spectrum for feature vector extraction. Also for feature matching we are using minimum Euclidean distance as a measure. This makes the system very easy to implement. The maximum accuracy is 97.19% with FFT for a feature vector of size 48. The present study is ongoing and we are trying to analyze the transform domain still further, as it has proved to be a promising way for feature vector extraction. Different algorithms for extracting the feature vector using transforms are being developed.

References

[1] Lisa Myers, An Exploration of Voice Biometrics, GSEC Practical Assignment version 1.4b Option 1, 2004

[2] Lawrence Rabiner, Biing-Hwang Juang and B.Yegnanarayana, “Fundamental of Speech Recognition”, Prentice-Hall, Englewood Cliffs, 2009.

[3] S Furui, “50 years of progress in speech and speaker recognition research”, ECTI Transactions on Computer and Information Technology, Vol. 1, No.2, November 2005.

[4] D. A. Reynolds, “An overview of automatic speaker recognition technology”, Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP’02), 2002, pp. IV-4072–IV-4075.

[5] Joseph P. Campbell, Jr., Senior Member, IEEE, “Speaker Recognition: A Tutorial”, Proceedings of the IEEE, vol. 85, no. 9, pp. 1437-1462, September 1997.

[6] S. Furui. Recent advances in speaker recognition. AVBPA97, pp 237--251, 1997

[7] F. Bimbot, J.-F. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-García, D.Petrovska-Delacrétaz, and D. A. Reynolds, “A tutorial on text-independent speaker verification,” EURASIP J. Appl. Signal Process., vol. 2004, no. 1, pp. 430–451, 2004.

[8] D. A. Reynolds, “Experimental evaluation of features for robust speaker identification,” IEEE Trans. Speech Audio Process., vol. 2, no. 4, pp. 639–643, Oct. 1994.

[9] Tomi Kinnunen, Evgeny Karpov, and Pasi Fr¨anti, “Real-time Speaker Identification”, ICSLP2004.

[10] Marco Grimaldi and Fred Cummins, “Speaker Identification using Instantaneous Frequencies”, IEEE Transactions on Audio, Speech, and Language Processing, vol., 16, no. 6, August 2008.

[11] Zhong-Xuan, Yuan & Bo-Ling, Xu & Chong-Zhi, Yu. (1999). “Binary Quantization of Feature vectors for robust text-independent Speaker Identification” in IEEE Transactions on Speech and Audio Processing Vol. 7, No. 1, January 1999. IEEE, New York, NY, U.S.

[12] J.Tierney,”A study of LPC Analysis ofspeech in additive noise”, IEEE Trans. Acoust., Speech Signal Processing, vol. ASSP-28, pp 389-397, Aug 1980.

[13] Sandipan Chakroborty and Goutam Saha, Improved Text-Independent Speaker Identification using Fused MFCC & IMFCC Feature Sets based on Gaussian Filter., International Journal of Signal Processing 5, Winter 2009.

[14] Speaker recognition using MFCC by S. Khan, Mohd Rafibul lslam, M. Faizul, D. Doll, presented in IJCSES (International Journal of Computer Science and Engineering System) 2(1): 2008.

[15] Speaker identification using MFCC coefficients –Mohd Rasheedur Hassan, Mustafa Zamil, Mohd Bolam Khabsani, Mohd Saifur Rehman, 3rd international conference on electrical and computer engineering (ICECE), (2004).

[16] Molau, S, Pitz, M, Schluter, R, and Ney, H., Computing Mel-frequency coefficients on Power Spectrum, Proceedings of IEEE ICASSP-2001, 1: 73-76 (2001).

[17] C.D. Bei and R.M. Gray.An improvement of the minimum distortion encoding algorithm for vector quantization, IEEE Transactions on Communications, October (1998).

[18] F. Soong, A. Rosenberg, L. Rabiner, and B-H. Juang, “A Vector Quantization Approach to Speaker Recognition,” In International Conference on Acoustics, Speech, and Signal Processing in Florida, IEEE, pp. 387-390,1985.

[19] F. K. Soong, A. E. Rosenberg, L. R. Rabiner, and B. H. Juang, “A Vector Quantization Approach to Speaker Recognition.” AT&T Technical Journal, Vol. 66, No. 2, pp. 14-26, 1987.

[20] Burton D.K. “Text-dependent Speaker verification using VQ source coding”, IEEE Transactions on Acoustics, Speech and Signal processing, vol. ASSP 35 No. 2 February 1987, pp 133-143.

[21] T Matsui and S Furui, “Comparison of Text Independent speaker recognition methods using VQ-distortion and discrete/continuous HMMs”, in Proc. IEEE ICASSP, Mar. 1992, pp. II. 157 – II.164.

[22] N. Z. Tishby, “On the Application of Mixture AR Hidden Markov Models to Text Independent Speaker Recognition,” IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 39, No. 3, pp. 563 – 570, 1991

[23] D A Reynolds, “A Gaussian mixture modeling approach to text independent speaker identification, PhD Thesis, Georgia Inst. of Technology, Sept. 1992.

[24] Dr. H B Kekre, Vaishali Kulkarni “Comparative Analysis of Speaker Identification using row mean of DFT, DCT, DST and Walsh Transforms”, International Journal of Computer Science and Information Security, Vol. 9, No.1, January 2011.

[25] Dr. H B Kekre, Vaishali Kulkarni, Sunil Venkatraman, Anshu Priya, Sujatha Narashiman, “Speaker Identification using Row Mean of DCT and Walsh Hadamard Transform”, International Journal on Computer Science and Engineering, Vol. 3, No.1, March 2011.

[26] Dr. H B Kekre, Vaishali Kulkarni, “Speaker Identification using Row Mean of Haar and Kekre’s Transform on Spectrograms of Different Frame Sizes”, (IJACSA) International Journal of Advanced Computer Science and Applications, Special Issue on Artificial Intelligence.

[27] Dr. H B Kekre, Vaishali Kulkarni,”Speaker Identification using Power Distribution in Frequency Spectrum”, Technopath, Journal of Science, Engineering & Technology Management, Vol. 02, No.1, January 2010.

[28] Dr. H B Kekre, Vaishali Kulkarni, “Speaker Identification by using Power Distribution in Frequency Spectrum”, ThinkQuest - 2010 International Conference on Contours of Computing Technology”, BGIT, Mumbai,13th -14th March 2010.

AUTHORS PROFILE

Dr. H. B. Kekre has received B.E. (Hons.) in Telecomm. Engineering, from Jabalpur University in 1958, M.Tech (Industrial Electronics) from IIT Bombay in 1960, M.S.Engg. (Electrical Engg.) from University of Ottawa in 1965 and Ph.D. (System Identification) from IIT Bombay in 1970. He has worked Over 35 years as Faculty of Electrical Engineering and then HOD Computer Science and Engg. at IIT Bombay. For last 13 years worked as a Professor in Department of Computer Engg. at Thadomal Shahani Engineering College, Mumbai. He is currently Senior Professor working with Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS University, Vile Parle (w), Mumbai, INDIA. He has guided 17 Ph.D.s, 150 M.E./M.Tech Projects and several B.E./B.Tech Projects. His areas of interest are Digital Signal processing, Image Processing and Computer Networks. He has more than 450 papers in National / International Conferences / Journals to his credit. Recently twelve students working under his guidance have received best paper awards. Recently five research scholars have received Ph. D. degree from NMIMS University Currently he is guiding eight Ph.D. students. He is member of ISTE and IETE.

Vaishali Kulkarni. Author has received B.E in Electronics Engg from Mumbai University in 1997, M.E (Electronics and Telecom) from Mumbai University in 2006. Presently she is pursuing Ph. D from NMIMS University. She has a teaching experience of around 10 years. She is Associate Professor in telecom Department in MPSTME, NMIMS University. Her areas of interest include networking, Signal processing, Speech processing: Speech and Speaker Recognition. She has 17 papers in National / International Conferences / Journals to her credit.

New design of Robotics Remote lab

Mohammad R. Kadhum
Art, Sciences and Technology University, Lebanon

Seifedine Kadry
Art, Sciences and Technology University, Lebanon

***Abstract*—The Robotic Remote Laboratory (RRL) controls the Robot labs via the Internet and applies the Robot experiment in easy and advanced way. If we want to enhance the RRL system, we must study requirements of the Robot experiment deeply. One of key requirements of the Robot experiment is the Control algorithm, which includes all important activities to affect the Robot; one of them relates the path or obstacle. Our goal is to produce a new design of the RRL that includes a new treatment to the Control algorithm which depends on isolating one of the Control algorithm's activities that relates the paths in a separated algorithm, i.e., design the (Path planning algorithm) is independent of the original Control algorithm. This aim can be achieved by depending on the light to produce the Light obstacle. To apply the Light obstacle, we need to hardware (Light control server and Light arms) and software (path planning algorithm).The NXT 2.0 Robot will sense the Light obstacle depending on its Light sensor. The new design has two servers: one for the path (Light control server) and other for the other activities of the Control algorithm (Robot control server).The website of the new design includes three main parts (Lab Reservation, Open Lab, Download Simulation).We proposed a set of scenarios for organizing the reservation of the Remote Lab. Additionally, we developed an appropriate software to simulate the Robot and to practice it before using the Remote lab.**

Keywords-Robotic Remote Laboratory; Robot experiment; Light obstacle; Control algorithm; NXT 2.0Robot; Robot lab; Path planning algorithm.

I. INTRODUCTION

Since, the Control algorithm which includes all important activities to effect the Robot is one of key requirements of the Robot experiment, researchers should focus their efforts on developing new ways in design the Control algorithm to produce a new design of the RRL is more safe, fast and robust, so, one of most important questions will be asked is: How can we design the Control algorithm (if we decide to replace or treat physical obstacles)?

This question is addressed without answer in [3] and this is the main goal of this article. Our proposed solution to this problem is by producing a new design of the RRL includes a new treatment to the Control algorithm depends on isolating one of the Control algorithm's activities that relates the paths or obstacles in a separated algorithm, i.e., design a (Path planning algorithm) is independent of the original Control algorithm. The Path planning algorithm will be responsible for planning the path only.

The proposed solution will depend on the light to produce the path or obstacle, i.e., produce a Light obstacle can be sensed depending on the Light sensor of the NXT 2.0 Robot. Another question must be asked here: How can we create the Light obstacle? The answer is by using a special hardware such as (Light control server and Light-arms) and special software such as a Path planning algorithm. The Light control server is responsible of control the Light-arms for lighting on a special area of Remote lab depending on orders of the Path planning algorithm to produce the Light obstacle, as we can see in Fig. 1.

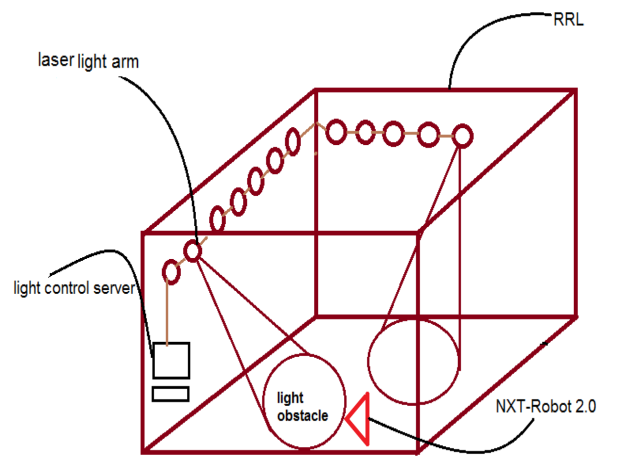


Figure 1. Generating the Light obstacle

II. BACKGROUND

In 1991, the first proposal for the Robotic Remote lab was presented. In 1992, testing first project of the Robot lab over a WAN [1]. In 1994, the first successful implementation of Robots lab via the Internet was developed by Goldberg at the University of South California. In 1995, McKee and Barson enhanced Remote lab by allowing Robot and its sensory devices in the lab to be controlled remotely.

In 1996, the Rls were being updated by increasing the interactivity between users and RL. In 1997, updating automated measurement system of the RLs, to allow multiple users. In 1998, start with use a self contained camera system that can be controlled via the Internet. In 1999, Virtual laboratory was produced by the National University of Singapore (NUS). In 2002, start with use the RL in universities, such as University of South Australia that depends on the RL in the lectures of it. In 2003, Automatic Control Tele-lab ACT system was produced at University of Siena. In 2009, the RLs became sharable among many users to treat the lack of modern labs.

III. Robotic Remote Laboratory

Robotic Remote laboratory (RRL) means control the Robot labs via the Internet, i.e., the Robot experiment which is to be run locally but directed remotely. A user can be any computer connected to the Internet using a web browser and has the ability for monitoring and control the labs. This technique resulted from merging two spread fields (remote operations and the Robot labs), this merge, led to revolution in computer science and communication field, and find out a strong relationship between them. Remote laboratory makes the Robot experiment available 24 a day-7 days a week for any authorized user. As well as, it provides a short cut access to the Robot experiment, i.e., each student or user can access the lab from any point that has a connection to the Internet. When we talk about Remote laboratory system, we must refer to its topology that consists of the Client side (users) and the Server side which has more than server (Robot control server, Camera server) those deal directly with web server via the Internet. Many applications can appear with the Robotic Remote Laboratory, such as, Tele-teaching, Tele-maintenance, Tele-experiments, and Tele-production. To declare Remote laboratory in a best manner, we must distinguish between it and (Virtual laboratory), where, a user can interact with physical experiments instead of dealing with a graphical interface designed in software to simulate the reality, as well as, design and implementation of Remote lab are more complex and cost than Virtual lab.

IV. Need For Robotic Remote Lab

The first purpose of using the Remote Lego laboratory is to allow students to compete remotely. The second is to treat lack of the modern laboratory in scientific institutes, by sharing the Robot labs among them. The third is to make Robot lab available 24 a day -7 days a week, for authorized users with no time limit. The forth is to enable users to apply their Robot experiment in easy and advanced way.

V. The Proposed Design

In this section, we will propose a new design of the Robotic Remote Lab assist in improving the interactivity between users and the RRL and provide a large space of flexibility especially in the operation which relates constructing the remote path or (Obstacle). Whenever user wants to test his/her Robot experiment on the Remote Lab, he/she must request the website of the RRL that will be responsible for offering the interaction between remote user and the RRL. The Internet is a most suitable transfer for applying that purpose, the following figure will show its basic design.

In our proposed design, the website of the RRL will include three main parts:

A. *Lab Reservation:* Which is responsible of organizing the access to the RRL considering it as a critical recourse must be synchronized.

B. *Open Lab:* Which allow access to users to the RRL and this will not be discussed here.

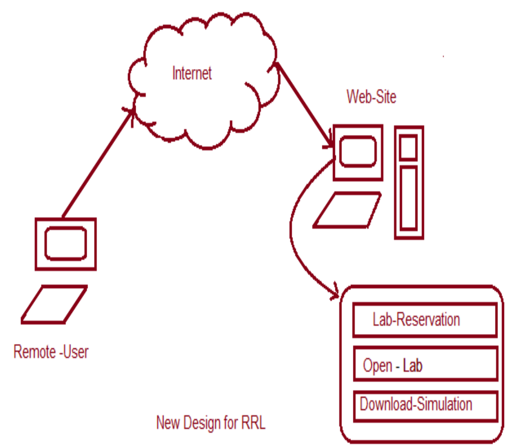


Figure 2. New design of the RRL

C. *Download Simulation:* An application program which will simulate the reality of operations, we tend to apply it because the high cost of the Robots and the long distance of the Robotic Remote Lab.

If we consider our lab and software include 2 Robots both want to use same lab for competition, then, they shall either choose the download simulation option or choose the Open Lab option which. Additionally, we can add an option to our software for generating the drawn path, and then to apply it physically to the lab by depending on the light techniques, i.e., we will provide the Remote lab with a further hardware for applying this purpose such as the Light control server and the Light-arms. As well as that, we will use a special input at user side such as the Mouse or the Touch Screen to apply the remote path. Whenever a user draws the path by the Mouse, the effect will be transferred to the Light control server of the Remote lab to generate the Light obstacles. The very important question must be asked here is: How can we organize the reservation of the Remote Lab, if we take in consideration, the Remote Lab is a critical area must be synchronized? To answer that, we suggest the following scenarios for dealing one or more users:

1) *for single user option:* A user has to create a new account and specify the date/duration of usage. After, the system will generate a pin and set it to that user, this pin will be valid only for the specified period. After the user log in, no other user can log in to the lab. The very important question that must be asked here is: How can we behave if two users want to access the same lab at the same time exactly? The answer is depending of the precedence for each user, i.e., a user which wasn't accessing that lab since last 24 hour will have a high primacy for access it. The lab "Open" option will be available after the specified period.

2) *for two users option:* A Coach (3rd person) should log in, create a new competition session and specify the date/duration.

The system will generate 3 different pins, one for the Coach (Coach Privilege) and two pins (user privilege) for two users.

During the Robot experiment, the Coach can start a competition session, watch it online and declare a winner of the match. Any user wins in 4 matches will have the ability to be the Coach of the next match. If a user lost more than 5 matches, the system will recognize him/her and prevent him/her from the competion for the next 24 hour. Further, the system will prepare a suitable plan for training the losers.

3) *for three user's option:* First user should have a (high priority), second user should have a (middle priority) and third user should have a (low priority), each user has to create a new account for usage. The lab system will generate 3 different pins, (high privilege) for the first, (middle privilege) for the second and (low privilege) for the third, the lab system will prefer the first more than the others, i.e., the first can specify a (t) duration while the second and the third can specify a (t/2) and a (t/3) duration for usage. The pins will be valid only for the specified period. The lab will determine a number of users (only three) for each period, so, after the three users log in, no other user can log in to the lab, the lab "Open" option will be available after the specified period.

4) *For four users or more option:* First user should be a master and the others should be slaves, in this case, the master user can consider as a central controller for the slave users, i.e., if the master user follow a special path or perform any operation, the slave users must achieve that also. Besides, the master can specify the date/duration and determine a number of the slaves (at least three). We note here, in spite of applying the same operation by the slave Robots, the amount of speed for each slave Robot will depend on a user of it. The lab system will generate four or more different pins, one for the master user (master privilege) and three or more pins (slave privilege) for the slave users. We note, after the four users log in, no other user can log in to the lab; the lab "Open" option will be available after the specified period.

5) *For five users option:* The lab system will determine a number of users (only five) for each duration, deal with them at the same priority, no one better than one (round table), in period way, each user will have a (time slice) for operation, i.e., share the duration time of the lab among users. All users shall be given the same date/duration for usage and each user has to create a new account. The system will generate five different pins, one pin for each user, this pin will be valid only for the specified period. After the five users log in, no other user can log in. The lab "Open" option will be available after the specified period.

6) *for six users or more option:* Parents (early two users) should login, specify the date/duration and determine a number of children (at least four). If one of parents want to log in alone, the system will refuse him/her and request two parents for trust, i.e., single user is not accepted, just for families, as we can see in fig. 3. After the children log in, no other user can log in to the lab. The system will generate six or more different pins, two pins for the parents (parent privilege) and four or more pins (child privilege) for the children. These pins will be valid only for the specified period. The lab "Open" option will be available after the specified period. After the suggested scenarios, we can construct an obvious idea about the nature of the new design. The application program which will simulate the reality of operations will be used for practicing Robots before usage the Remote lab. We design and implement it because the high cost of the Robots and the long distance of the Robotic Remote Lab. the next section will produce appropriate discussion for third option of the new design which is Download Simulation.

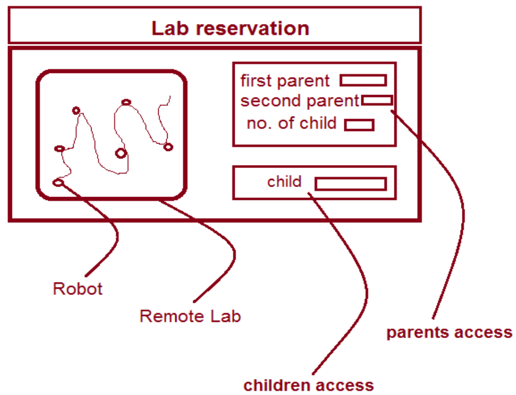


Figure 3. The Sixth scenario of lab reservation

VI. Simulation And Analysis

The goal of this section is to design and implement the third option of our new design which is Download-simulation that aims to simulate the operations of the Robotic Remote Lab and provide a simulation environment can be used to practice users before usage the remote lab. We do this simulation, because the high cost of the Robots and the long distance of the Robotic Remote Lab. Before starting design and implement this application, we will take two case studies for the system, the first before applying the new design and the other after applying it to declare the new design in best manner as follows:

In the first case, a user designs the Control algorithm that contains all activities of the Robot as (path planning, speed determination,… etc.), then, a user sends the designed algorithm to the Robot control server to start the interaction between a user and the Robot. We see, all the activities of the Robot will be combined in one algorithm and send to one server for treating, as in fig. 4, if the server fails, the interaction will fail (no connection). As well as, the Robot control server must have a large memory to contain the algorithm that includes all activities. The (defined, predefined) path will be sent through the (designed, predesigned) Control algorithm to the Robot control server.

In the second case, a user designs path of the Robot alone in a special algorithm (Path planning algorithm) and the other activities alone in the (Control algorithm), then, the path algorithm will be sent to the Light control server to apply it, and the algorithm of the other activates to the Robot control server to perform it. If the Robot control server fails, the interaction will not fail, i.e., the system will continue generating the path but it will not have any updating about the other activities such a speed limitation and system tracking. The Control algorithm will be divided into two parts, the first is responsible of the path and the second is responsible of the other activities such as (speed, start, track,….etc.) as we see in fig. 5, each part has it's server. The system will become robust, since activities will be distributed between two servers instead of one. We see also, two operations shall be applied at the same time instead of one led to fast system.

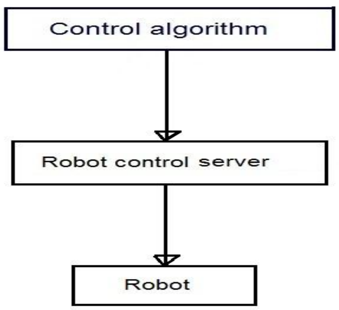


Figure 4. case study1

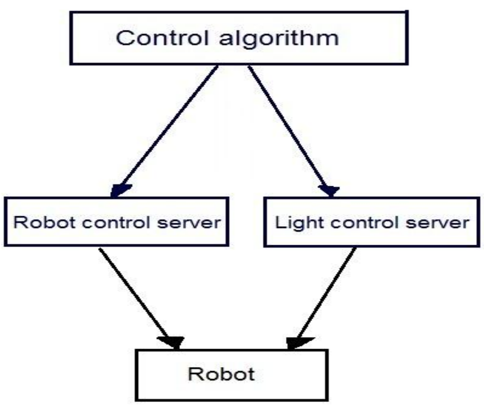


Figure 5. case study2

At the first of the simulating program, we will start with predefine all important variables which will be used later, we see, most of them defied as a (Public) to make them usable by any part or object in the program considering them as a (global variables). We see also, a previous linking with the important libraries such as the sound-library (SpeechLib) and the graph-library (Graphics) which will be used as an important part in shaping colored lines, arcs and zigzag paths. Further, we determine type of color (red, blue), dimensions of arrays that uses for keeping track of the paths and set of variables which will be used in different parts in the program, as we see in the following orders: PrivategAsGraphics Dimp1AsNewPen(Color.Red)

Dim mrk

Dim spAsSpeechLib.ISpeechAudioPublic

path1(200, 2) As Integer Public i2 As Integer

Besides, we use the (Visual basic 2008) language as a programming tool offers an appropriate environment for simulation, the DELL-laptop-Studio 32- OS Intel(R) Core (TM) 2 Due CPU T5800 2.00GHZ 2.00GHZ (RAM) 4.00 GB and Windows 7 Ultimate.

The Download Simulation will be separated into two main parts:

A. Control Algorithm

B. Robot's behavior

A. Control Algorithm

The web site of our new design must provide the tools which are responsible of simulation the Control algorithm in easy and clear way. The behavior of the Robot will be changed according to change in the Control algorithm. The important activates that affect the Robot's behavior are the path, the speed, the track and the start/stop operation.

1) Path Construction

This effect is responsible of simulation the path planning activity of the Control algorithm, offering the manner will be used for constructing the colored path. The Mouse is used to enter the path, so, the very important question here, how can we take the coordinates of the Mouse pointer during the movement of it? The answer will be via the programming sentence:

(Private Sub Form1_Click(By Val sender As Object, By Val e As MouseEventArgs)Handles Me. Click). Depending on the previous sentence, any click on the form will led to keep the coordinates of the Mouse pointer in the system variables (e.x,e.y), then, they shall pass to the function (draw) which is responsible of drawing any required shape or path, if we consider, the color of the shape is determined previously. Other important question must ask here, how can the Robot sense the path and recognize the color of it ? The answer is by making the Robot follow a special frequency of color more than others, i.e., first Robot will follow the red color and recognize it only, while, the second will follow other color such the blue, as we see in fig. 6.

The following code will perform this operation:

```
PrivateSubdraw(ByVal xpo As Integer,ByVal ypo As Integer)
           If cl1 = 1 Then g.DrawLine(p1,xp1,yp1,xpo,ypo)
              xp1=xpo
               yp1=ypo
          EndIf
        If cl2 =2 Then g.DrawLine(p2,xp2,yp2,xpo,ypo)
             xp2=xp
            oyp2=ypo
         EndIf
  g.Flush()
EndSub.
```

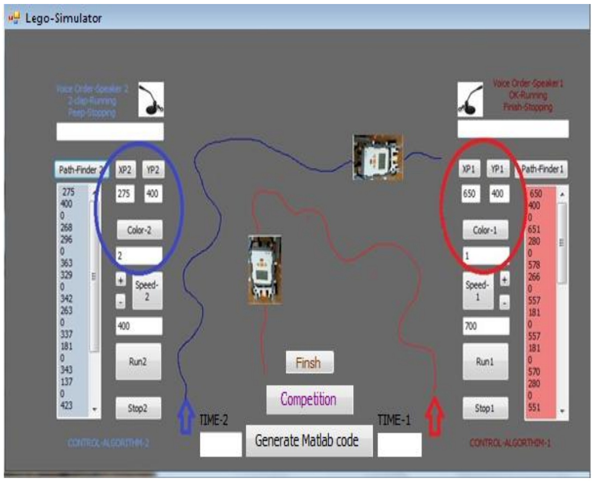


Figure 6. Path Construction by Mouse pointer

2) Speed Limitation

This effect is responsible of simulation the speed activity of the Control algorithm, offering tools for increase or decrease steps of the Robot. Using it, we can simulate the competion between two Robots. The time of each Robot will be computed during the match, as in fig. 7.

The following code will perform this operation:

```
Private Sub Button8_Click(ByVal sender As
System.Object)
   If    TextBox2.Text    =    1000    Then
MsgBox("You are accessed the limtation of
speed(High Speed!!!)")
            Timer1.Enabled = False
        ElseIf TextBox2.Text = 0 Then
MsgBox("You areaccessed the limtation of
speed(Low Speed!!!)")
            Timer1.Enabled = False  Else
Timer1.Interval=1000-Val(TextBox2.Text)
        End If
End Sub
```

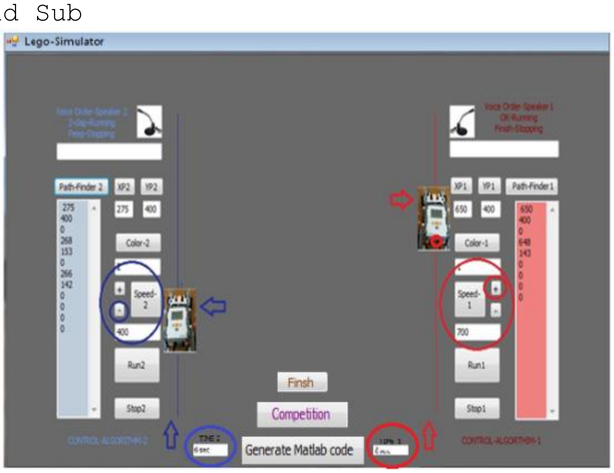


Figure 7. Using Speed limtations in competion

3) Run & Stop Operations

This effect is responsible of simulation the Start and Stop activity of the Control algorithm, offering two orders, the first can animate the Robot and the second can stop it.

There are two ways to apply these orders:

- Mechanical order: we can run the Robot by click on (Run) button, as we see in fig. 8, the following code will do this operation:

```
Private Sub Button5_Click(ByVal sender As
System.Object,ByValeAsSystem.EventArgs)
HandlesButton5.Click
     If cl1=1 Then
          Timer1.Enabled=True
     EndIf
End Sub.
```

And we can stop it by clicking on (**Stop**) button, as we see in fig. 8.

The following code will do this operation:

```
Private Sub Button4_Click(ByVal sender As
System.Object,          ByVal          e
AsSystem.EventArgs)HandlesButton4.Click
Timer1.Enabled = False
EndSub.
```

- Vocal order : in this case, we can run the Robot by giving a (voice order) to the Robot which can sense sound, for example, if we say the word (ok), the Robot will run and if we say the word (finsh) it will stop, as we see in fig. 8.

The following code will perform this operation:

```
PrivateSubTextBox3_TextChanged(ByValsende
rAsSystem.Object,By
Val   e   As   System.EventArgs)   Handles
TextBox3.TextChanged
    If TextBox3.Text <> "" Then
       IfTextBox3.Text="OK"Then
           If Timer1.Enabled =True Then
mrk=CreateObject("sapi.spvoice")mrk.speak
              ("Iam  run  now  why  you  are
repeated")
          Else Timer1.Start()
mrk=CreateObject("sapi.spvoice")mrk.speak
("thank  you  for  run  me  with  order"  &
TextBox3.Text )
      End If
ElseIf
 TextBox3.Text = "Finish" Then
      If Timer1.Enabled = True Then
          Timer1.Stop()
mrk=CreateObject("sapi.spvoice")mrk.speak
("why  you  are  stopped  me  with  order"&
TextBox3.Text )
 Else
mrk = CreateObject("sapi.spvoice")
                    mrk.speak("Iam     not
running to stop me now ")
                End If
                End If
        End If
    End Sub
```

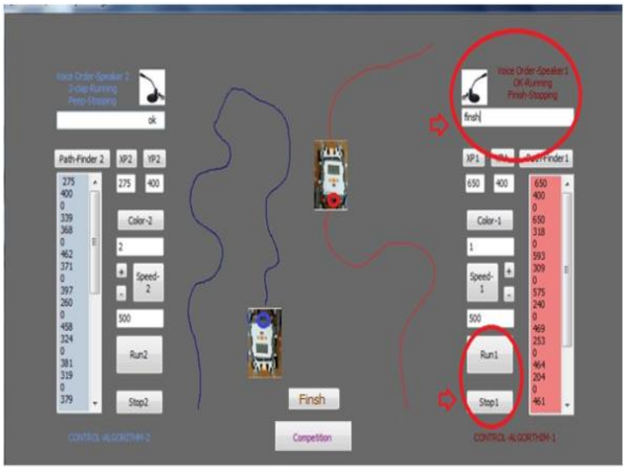


Figure 8. Using vocal and mechanical order to start and stop

4) Route Derivation:

This effect is responsible of simulation the (Keep track) activity of the Control algorithm, producing a final report of points which the Robot pass them through the travel of it, as we see in fig. 9.We can do that, by click on the (Path-Finder) button. The following code will perform this operation:

```
Private Sub chkpath1()
 n1 = n1 + 1
 path1(n1, 0) = xg
 path1(n1, 1) = yg
 End Sub
Private Sub prinpath1()
        TextBox7.Text = " "
        For i = 0 To n1 + 1
            For j = 0 To 2
                TextBox7.Text =
TextBox7.Text & path1(i, j) & vbCrLf
            Next
        Next
    End Sub
```

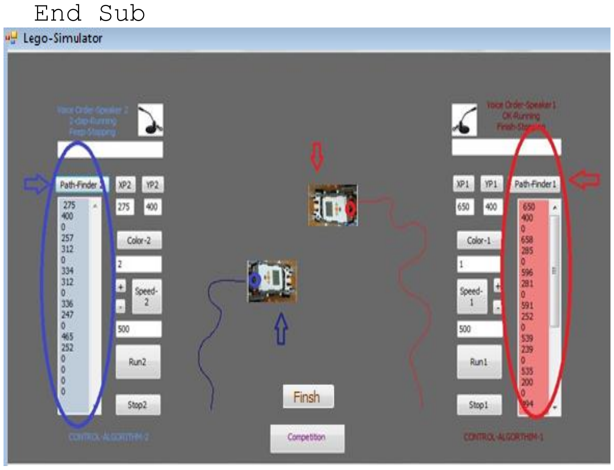


Figure 9. Final report of Route Derivation

B. Robot's behavior

In this part we will simulate the relation between the Robot and the Control algorithm, so, we will see every effect in the Control algorithm will affect the Robot's behavior, i.e., draw the path, increase or decrease the speed, keep the track and start the operation, shall led to produce the final behavior of the Robot, such as select the correct path, rotate in a true angle and move to one of the four sides, as we see in fig. 10.

The following code will perform this operation:

```
  Private Sub Timer1_Tick(ByVal sender As
System.Object,        ByVal       e       As
System.EventArgs) Handles Timer1.Tick
      If i2 Mod 2 = 0 Then

PictureBox1.Load("D:\games1\1.gif")
        Else

PictureBox1.Load("D:\games1\2.gif")
        End If
        i2 = i2 + 1
My.Computer.Audio.Play(My.Resources.ss1,
AudioPlayMode.Background)
        If     PictureBox1.Location.Y     >
path1(n1, 1) Then
       PictureBox1.Top   =   PictureBox1.Top
- 10
        Else
            If   PictureBox1.Location.X   >
path1(n1, 0) Then
pictureBox1.Image.RotateFlip(RotateFlipTy
pe.Rotate90FlipX)
          PictureBox1.Left               =
PictureBox1.Left - 5
            Else
      If     PictureBox1.Location.X     <
path1(n1, 0) Then
PictureBox1.Image.RotateFlip(RotateFlipTy
pe.Rotate90FlipY)
           PictureBox1.Left=
ictureBox1.Left + 5
                Else
PictureBox1.Image.RotateFlip(RotateFlipTy
pe.Rotate180FlipX)
                    PictureBox1.Top      =
PictureBox1.Top + 10
              End If
           End If
        End If
    End Sub
```

VII. CONCLUSION

After analyzing the results of the two case studies, we can conclude that applying the new design of the RRL will led to isolation the path planning activity of the Control algorithm in independent algorithm (path planning algorithm) depending on Light techniques, i.e., we provide another server for the RRL (Light control server) responsible of producing the path or the Light obstacle only, So, the new system will have two servers, one for applying the path(Light control server) and other for applying the other activates of the Control algorithm(Robot control server) .

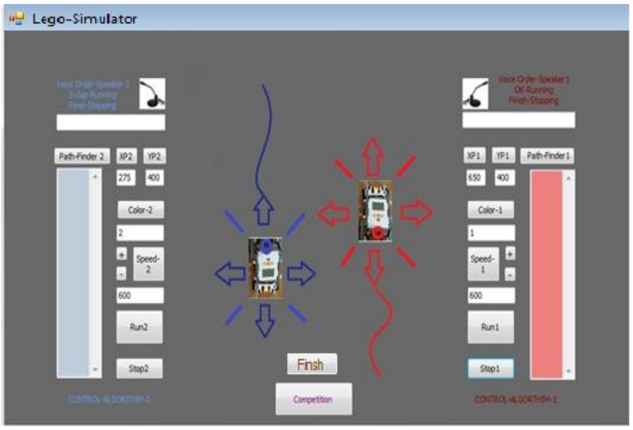


Figure 10. Robot's behavior

The system will become stronger, faster and more reliable, since, the work is divided between two servers instead of one. As well as, the new design will open the door to huge developments in the RRL system depend on partition and distribution the other activities of the Control algorithm among many servers, each activity has its server.

VIII. FUTURE WORK

The future developments regard the possibility of use a Virtual driver for the Robot experiment, determine the speed of the Robot according to position of the Light obstacle, if, it is far, the Virtual driver will increase the speed (speeder), else it will decrease the speed to average, this feature will allow users to design and test speed change algorithm.

REFERENCES

[1] J. Machotka, A. Nafalski & Z. Nedić. The history of developments of Remote experiments. World Conference on Technology and Engineering Education. 2011

[2] F. MOUTAOUAKKIL, M. EL BAKKALI and H. MEDROMI. New Approach of TeleRobotic over The Internet. Proceedings of the World Congress on Engineering and Computer Science. 2010

[3] Fabio Carusi, Marc0 Casini, Domenico hattichizzo, Antonio Vicino. Distance Learning in Robotics and Automation by Remote Control of Lego Mobile Robots. International Conference on Robotics & Automation. 2004

[4] Carlos A. Jara, Francisco A. Candelas, Fernando Torres. Virtual and Remote Laboratory for Robotics E-Learning. European Symposium on Computer Aided Process Engineering. 2008

[5] [5] Christopher Beland, Wesley Chan,D. Mindell. LEGO Mindstorms. Structure of Engineering Revolutions. 2000

[6] Marco Casini, Andrea Garulli, Antonio Giannitrapani, Antonio Vicino. A LEGO Mindstorms multi-Robot setup in the Automatic Control Telelab. Dipartimento di Ingegneria dell'Informazione, Italy. 2011

[7] Marco Casini, Domenico Prattichizzo, Antonio Vicino. REMOTE CONTROL OF A LEGO MOBILE ROBOT THROUGH THE WEB. DII - Universit`a di Siena via Roma 56, 53100 Siena, Italy. 2009

[8] David Be, Michel García, Cinhtia González, Carlos Miranda. Wireless Control LEGO NXT robot using voice commands. International Journal on Computer Science and Engineering.2011.

Performance Evaluation of Adaptive Virtual Machine Load Balancing Algorithm

Prof.Meenakshi Sharma
CSE Department SSCET Badhani
Pathankot, India

Pankaj Sharma
CSE Department SSCET Badhani
Pathankot, India

***Abstract*— The conception of Cloud computing has not only reshaped the field of distributed systems but also extend businesses potential. Load balancing is a core and challenging issue in Cloud Computing. How to use Cloud computing resources efficiently and gain the maximum profits with efficient load balancing algorithm is one of the Cloud computing service providers' ultimate goals. In this paper firstly an analysis of different Virtual machine(VM) load balancing algorithms was done, a new VM load balancing algorithm has been proposed and implemented in Virtual Machine environment of cloud computing in order to achieve better response time and cost.**

Keywords-Virtual machine; load balancing; cloudsim.

I. INTRODUCTION

Cloud computing is a fast growing area in computing research and industry today. It has the potential to make the new idea of 'computing as a utility' in the near future. The Internet is often represented as a cloud and the term "cloud computing" arises from that analogy. Cloud computing is the dynamic provisioning of IT capabilities (hardware, software, or services) from third parties over a network [7]. It is generally supposed that there are three basic types of cloud computing: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) [1].

In IaaS grids or clusters, virtualized servers, memory, networks, storage and systems software are delivered as a service. Perhaps the best known example is Amazon's Elastic Compute Cloud (EC2) and Simple Storage Service (S3), IaaS Provide access to computational resources, i.e. CPUs. And also Provide (managed and scalable) resources as services to the user [7]. PaaS typically makes use of dedicated APIs to control the behavior of a server hosting engine which executes and replicates the execution according to user requests .E.g Force.com, Google App Engine. Software as a Service (SaaS) Standard application software functionality is offered within a cloud. Examples: Google Docs, SAP Business by design.Load balancing is one of prerequisites to utilize the full resources of parallel and distributed systems. Load balancing mechanisms can be broadly categorized as centralized or decentralized, dynamic or static, and periodic or non-periodic. Physical resources can be split into a number of logical slices called Virtual Machines (VMs).

All VM load balancing methods are designed to determine which Virtual Machine assigned to the next cloudlet [11]. This document introduce a new VM load balancing algorithm and compare the performance of this algorithms with the already existing algorithms like throttled and active monitoring VM load balancer [11]. Section III introduce the problem formulation, section IV include the purpose algorithm of the problem and result in section V

II. EXISTING VM LOAD BALANCER

Virtual machine enables the abstraction of an Operating System and Application running on it from the hardware. The interior hardware infrastructure services interrelated to the Clouds is modelled in the simulator by a Datacenter element for handling service requests. These requests are application elements sandboxed within VMs, which need to be allocated a share of processing power on Datacenter's host components. DataCenter object manages the data center management activities such as VM creation and destruction and does the routing of user requests received from User Bases via the Internet to the VMs. The Data Center Controller [11] uses a VmLoadBalancer to determine which VM should be assigned to the next request for processing. Most common Vmloadbalancer are throttled and active monitoring load balancing algorithms.

A. Throttled load balancer

It maintain a record of the state of each virtual machine (busy/ideal), if a request arrive concerning the allocation of virtual machine, throttled load balancer send the ID of ideal virtual machine to the data center controller and data center controller allocates the ideal virtual machine.

B. Active monitoring load balancer

Active VM Load Balancer maintains information about each VMs and the number of requests currently allocated to which VM. When a request to allocate a new VM arrives, it identifies the least loaded VM. If there are more than one, the first identified is selected. ActiveVmLoadBalancer returns the VM id to the Data Center Controller; the data Center Controller sends the request to the VM identified by that id.DataCenterController notifies the ActiveVmLoadBalancer of the new allocation.

III. PROBLEM FORMULATION

In this paper a study of various virtual machine load balancing algorithms in cloud computing environment is done. The algorithms are round robin, throttled load balancer and active monitoring load balancer. A new algorithm has been proposed after modifying the throttled load balancing algorithm in Virtual Machine environment of cloud computing in order to achieve better response time, processing time and cost.

IV. PROPOSED VM LOAD BALANCING ALGORITHM

The Proposed VM Load balancing algorithm is divided into three phases.

The first phase is the initialization phase, where in the expected response time of each VM has been found.

Second Phase finds the efficient VM (VM having less response time), Last Phase returns the ID of efficient VM to datacenter controller.

- Efficient algorithms find expected response time of each Virtual machine.

// expected response time find with the help of resource info program

- When a request to allocate a new VM from the Datacenter Controller arrives, Algorithms find the most efficient VM (efficient VM having least loaded, minimum expected response time) for allocation.
- Proposed algorithms return the id of the efficient VM to the Datacenter Controller.
- Datacenter Controller notifies the new allocation
- Proposed algorithm updates the allocation table increasing the allocations count for That VM.
- When the VM finishes processing the request and the DataCenerController receives the Response. Datacenter controller notifies the efficient algorithm for the VM de-allocation.
- Start from step 2

The proposed algorithm finds the expected Response Time of each Virtual Machine because each virtual machine is of heterogeneous platform, the expected response time of each virtual machine can be found with the help of the following formula:

$$\text{Response Time} = Fin_t - Arr_t + TDelay \quad (1)$$

Where Arr_t is the arrival time of user request and Fin_t is the finish time of user request and the transmission delay can be determined using the following formula:

$$TDelay = Tlatency + Ttransfer \quad (2)$$

Where TDelay is the transmission delay, T latency is the network latency and T transfer is the time taken to transfer the size of single request from source location to destination.

$$Ttransfer = D / Bwperuser \quad (3)$$

$$Bwperuser = Bwtotal / Nr \quad (4)$$

Where Bw_{total} is the total available bandwidth and N_r is the number of user requests currently in transmission. The Internet Characteristics also keeps track of the number of user requests in-flight between two regions for the value of N_r.

V. EXPERIMENTAL RESULT

Proposed algorithm is implemented with the help of simulation packages like CloudSim and cloudSim based tool [11]. Java language is used for implementing VM load balancing algorithm.

We Assume that the application has been deployed in one data center having 50 virtual machines (with 1024Mb of memory in each VM running on physical processors capable of speeds of 100 MIPS) where the parameter values are as under:

TABLE I. PARAMETER VALUE

Parameter	***value***
Data Center OS	Window 7
VM Memory	1024 mb
Data Center Architecture	X86
Service Broker Policy	Optimize Response Time
VM Bandwidth	1000

Followings are the experimental results based on Efficient VM Load Balancing Algorithm:

TABLE II. RESULT DETAIL

Overall Avg Response Time With Efficient VM Load Balancing Algorithms			
Overall Response Time	***Avg(ms)***	***Min(ms)***	***Max(ms)***
	171.43	35.06	618.14
Cost with Efficient Load Balancing Algorithm			
Cost	***VM Cost $***	***Data Transfer Cost $***	***Total Cost$***
	240.11	1.94	242.05

TABLE III. COMPARISON OF AVG RESPONSE TIME OF VM LOAD BALANCING ALGORITHMS.

Response Time(ms)	**Throttled (ms)**	**Active Monitoring (ms)**	**Efficient (ms)**
	263.14	264.02	171.43

Fig.1 shows the graphical representation of average Response time of VM load balancing algorithms. In our experiments, Average response Time of three VM load balancing algorithms was not same.

This experiment notifies that if we select an efficient virtual machine then it affects the overall performance of the cloud Environment. Fig 1 represent the average response time of each VM load balancing algorithm.

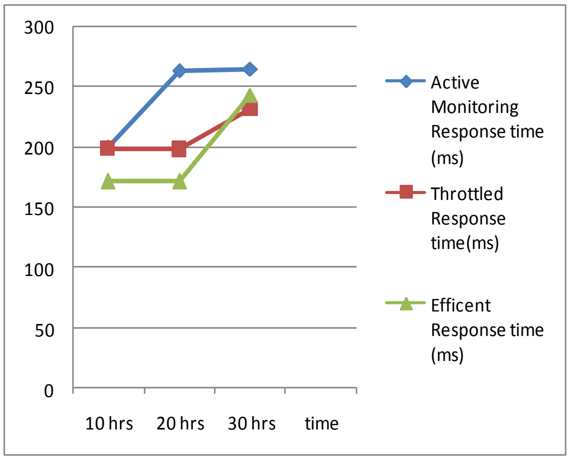


Figure 1. Comparison of Avg Response Time of VM Load balancing Algorithms.

VI. CONCLUSION

In this paper, a new VM load balancing algorithm is proposed which is implemented in CloudSim, an abstract cloud computing environment using java language. Proposed algorithm finds the expected response time of each resource (VM) and sends the ID of virtual machine having minimum response time to the data center controller for allocation to the new request. According to this experiment, we conclude that if we select a efficient virtual machine then it effects the overall performance of the cloud Environment and also decreases the average response time.

REFERENCES

[1] CLOUD COMPUTING MADE EASY by Cary Landis and Dan Blacharski,version 0.3

[2] Ioannis Psoroulas, Ioannis Anagnostopoulos, Vassili Loumos and Eleftherios Kayafas. "A Study of the Parameters Concerning Load Balancing Algorithms" IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.4, April 2007,Page No 202-214 .

[3] Sandeep Sharma, Sarabjit Singh, and Meenakshi Sharma "Performance Analysis of Load Balancing Algorithms" World Academy of Science, Engineering and Technology 38 ,2008 page no 269- 272.

[4] Luqun Li ," An Optimistic Differentiated Service Job Scheduling System for Cloud Computing Service Users and Providers" Third International Conference on Multimedia and Ubiquitous Engineering,IEEE Explore , 4-6 June 2009, Page No 295-299.

[5] Rodrigo N. Calheiros, Rajiv Ranjan, César A. F. De Rose, and Rajkumar Buyya, "Modeling and Simulation of Scalable Cloud Computing Environment and CloudSim Toolkit: Challenges and opportunities".IEEE Explore International Conference on high performance Computing And simulation, 21-24 June 2009 Page No 1-11.

[6] QI CAO , ZHI-BO WEI ,WEN-MAO GONG." An Optimized Algorithm for Task Scheduling Based On Activity Based Costing in Cloud Computing". International Conference on Bioinformatics and Biomedical Engineering,IEEE Explore 14 July 2009 Page No 1-3.

[7] Bhasker Prasad Rimal, Eummi Choi, Lan Lump "A Taxonomy and Survey of Cloud Computing System" 5th International Joint Conference on INC,IMS and IDC,IEEE Explore 25-27 Aug 2009,Page No 44-51.

[8] Yi Zhao, Wenlong Huang," Adaptive Distributed Load Balancing Algorithm based on Live Migration of Virtual Machines in Cloud" IEEE 5th International Joint Conference on INC,IMS and IDC, 25-27 Aug 2009,Page No 170-176

[9] Huai Zhang, Shufen Zhang, Xuebin Chen, Xiuzhen Huo," Cloud Computing Research and Development Trend" Second International Conference on Future Networks, 22-24 Jan 2010, Page No 93-97.

[10] Martin Randles, David Lamb, A. Taleb-Bendiab, A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing, IEEE 24th International Conference on Advanced Information Networking and Applications Workshops 20-23 April 2010, Page No 551-556.

[11] Bhathiya Wickremasinghe, Rodrigo N. Calheiros, and Rajkumar Buyya. "CloudAnalyst: A CloudSim-based Visual Modeller for Analysing Cloud Computing Environments and Applications",20-23 April 2010 Page No 446-452.

[12] Zehua Zhang and Xuejie Zhang," A Load Balancing Mechanism Based on Ant Colony and Complex Network Theory in Open Cloud Computing Federation" 2010 2nd International Conference on Industrial Mechatronics and Automation. 30-31 May 2010 Page No 240-243.

[13] Thomas A. Hen zinger Anmol V. Singh Vasu Singh Thomas Wies Damien Zufferey, FlexPRICE:Flexible Provisioning of Resources in a Cloud Environment, IEEE 3rd International Conference on Cloud Computing 5-10 july 2010.Page No 83-90.

[14] Shu-Ching Wang, Kuo-Qin Yan *(Corresponding author), Wen-Pin Liao and Shun-Sheng Wang "Towards a Load Balancing in a Three-level Cloud Computing Network"IEEE 3rd International Conference on Computer Science & Information Technology 9-11 July 2010, Page No 108-113.

[15] Hai Zhong1, 2, Kun Tao1, Xuejie Zhang," An Approach to Optimized Resource Scheduling Algorithm for Open-source Cloud Systems" The Fifth Annual ChinaGrid Conference 16-18 July 2010, Page No 124-129

[16] Yoshitomo MURATA, Ryusuke EGAWA, Manabu HIGASHIDA, Hiroaki KOBAYASHI," A History-Based Job Scheduling Mechanism for the Vector Computing Cloud" 2010 10th Annual International Symposium on Applications and the Internet.19-23 July 2010.Page No 125-128

[17] Andrew J. Younge, Gregor von Laszewski, Lizhe Wang, Sonia Lopez-Alarcon, Warren Carithers," Efficient Resource Management for Cloud Computing Environments" International Conference on Green Computing.15-18 Aug 2010, Page No 357-364

[18] LIU Jia, HUANG Ting-Lei," Dynamic Route Scheduling for Optimization of Cloud Database. International Conference on Intelligent Computing and Integrated System, 22-24 Oct 2010,Page No 680-682

[19] Jinhua Hu, Jianhua Gu, Guofei Sun,Tianhai Zhao," A Scheduling Strategy on Load Balancing of Virtual Machine Resources In Computing Environment" 3rd International Symposium on Parallel Architecture, Algorithms and Programming IEEE Explore 18-20 Dec 2010,page no. 89-96.

Polylogarithmic Gap between Meshes with Reconfigurable Row/Column Buses and Meshes with Statically Partitioned Buses

Susumu Matsumae
Dept. of Information Science
Saga University, Japan

Abstract—**This paper studies the difference in computational power between the mesh-connected parallel computers equipped with dynamically reconfigurable bus systems and those with static ones. The mesh with separable buses (MSB) is the mesh-connected parallel computer with dynamically reconfigurable row/column buses. The broadcast buses of the MSB can be dynamically sectioned into smaller bus segments by program control. We show that the MSB of size $n \times n$ can work with $O(\log^2 n)$ step even if its dynamic reconfigurable function is disabled. Here, we assume the word-model broadcast buses, and use the relation between the word-model bus and the bit-model bus.**

Keywords- ***mesh-connected parallel computer; dynamically reconfigurable bus; statically partitioned bus; simulation algorithm.***

I. Introduction

The mesh-connected parallel computers equipped with dynamically reconfigurable bus systems gained much attention due to their strong computational powers [3, 11, 12, 13, 14]. The dynamic reconfigurable function enables the models to make efficient use of broadcast buses, and to solve many important, fundamental problems efficiently, mostly in a constant or polylogarithmic time [13]. Such reconfigurability, however, makes the bus systems complex and causes negative effects on the communication latency of global buses [2]. Hence, it is practically important to study the trade-off between such points quantitatively.

In this paper, we investigate the impact of reconfigurable capability on the computational power of mesh-connected computers with global buses. Here, we deal with the *meshes with separable buses* (MSB) [3, 12] and a variant of the meshes with partitioned buses called the *meshes with multiple partitioned buses* (MMPB) [4]. The MSB and the MMPB are the mesh-connected computers enhanced by the addition of broadcast buses along every row and column.

The broadcast buses of the MSB, called *separable buses*, can be dynamically sectioned into smaller bus segments by program control, while those of the MMPB, called *partitioned buses*, are statically partitioned in advance and cannot be dynamically reconfigurable. In the MSB model, each row/column has only one separable bus, while in the MMPB model, each row/column has L partitioned buses ($L \geq 1$). By comparing the relative power between these models, we clarify the difference in computational power between the parallel models equipped with reconfigurable bus systems and those with static ones. In this paper, we assume that the size of MSB and that of MMPB are of $n \times n$. The case of different sizes was investigated in [8].

Here, we study how much slowdown is necessary when we deprive the MSB of its reconfigurable function. In [5, 6], we have shown that the MSB of size $n \times n$ can be simulated time-optimally in $O(n^{1/(2L+1)})$ steps using the MMPB of size $n \times n$, where L is constant and the global buses are of word-model, i.e., the bus-width is the same as the number of bits in one word. From this result, it is natural to think that the slowdown may be at least of polynomial time. However, here we show that we can suppress the slowdown to polylogarithmic time, by making use of the relation between the word-model bus and the bit-model bus.

In this paper, we show that the $n \times n$ MSB can work with $O(\log^2 n)$ step slowdown even if its reconfigurable function is disabled. Here, we assume that the broadcast buses are of word-model, and use the relation between the word-model bus and the bit-model bus. As a corollary, since we have shown that the MSB of size $n \times n$ can simulate the reconfigurable mesh [1, 11, 14] (or PARBS, the processor array with reconfigurable bus systems) of size $n \times n$ in $O(\log^2 n)$ steps [10], we can say that the reconfigurable mesh of size $n \times n$ can also work with $O(\log^4 n)$ step slowdown even if its reconfigurable function is unused. In [7], we have proposed more efficient algorithm, which exploits the pipeline technique heavily. Although the algorithm presented here is slower than the one in [7] by the factor of log *n*, the key ideas and explanations are much simpler than those in [7].

This paper is organized as follows: Section II describes the MSB and the MMPB models, and briefly explains how to solve the simulation problem of the MSB by using the MMPB. Section III shows that the $n \times n$ MSB can work with $O(\log^2 n)$ step slowdown even if its reconfigurable function is disabled. Lastly, Section IV offers concluding remarks.

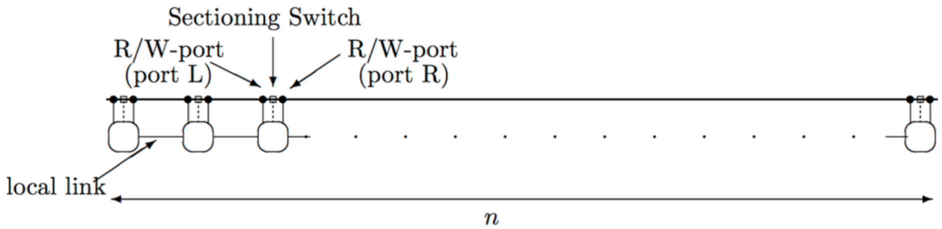


Figure 1. A separable bus along a row of the n × n MSB. Each PE has access to the bus via the two read/write-ports beside the sectioning switch.

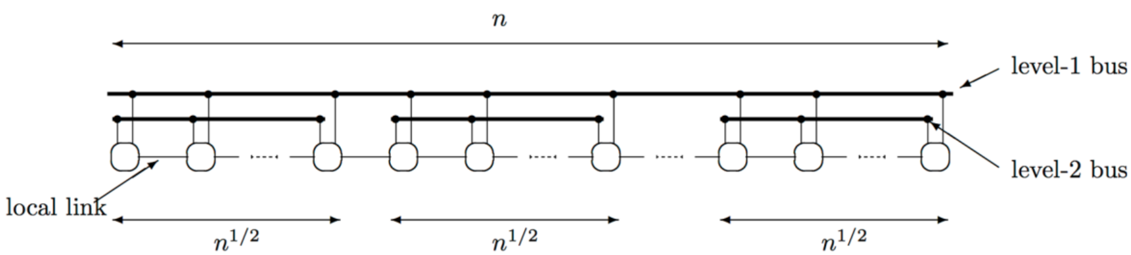


Figure 2. Partitioned buses along a row of the n × n MMPB. Here, L = 2, ℓ_1 =n, and $\ell_2 = n^{1/2}$.

II. Preliminaries

A. Models

An $n \times n$ mesh consists of n identical processors or processing elements (PEs) arranged in a two-dimensional grid with n rows and n columns. We assume that all the meshes are synchronous. The PE located at the grid point (i, j), denoted as PE[i,j], is connected via bi-directional unit time communication links to those PEs at $(i \pm 1, j)$ and $(i, j \pm 1)$, provided they exist $(0 \le i, j < n)$. PE[0,0] is located in the top-left corner of the mesh. Each PE[i, j] is assumed to know its coordinates (i, j).

An $n \times n$ mesh with separable buses (MSB) and an $n \times n$ mesh with multiple partitioned buses (MMPB) are the $n \times n$ meshes enhanced with the addition of broadcast buses along every row and column. The broadcast buses of the MSB, called separable buses, can be dynamically sectioned through the PE-controlled switches during the execution of programs, while those of the MMPB are statically partitioned in advance by a fixed length. In the MSB model, each row/column has only one separable bus (Fig. 1), while in the MMPB model each row/column has L partitioned buses (Fig. 2). The MSB is essentially the same model as the horizontal-vertical reconfigurable mesh (HV-RM) described in [1, 13]. Those L partitioned buses of the MMPB are indexed as level-1, level-2, ..., level-L, respectively. We assume that the partitioned buses of the MMPB are equally partitioned by the same length if they belong to the same level. For each level-k, the value ℓ_k denotes the length of a bus segment of the partitioned bus in level-k. Without loss of generality, we assume $\ell_k \ge \ell_{k+1}$.

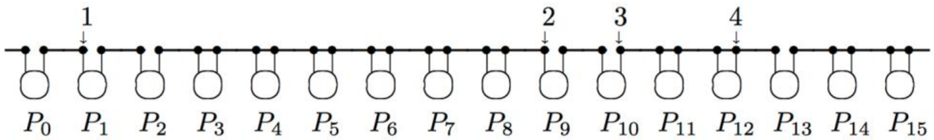
We assume that the word size of processor is $\lceil \log n \rceil$ for a mesh of size $n \times n$. As for the bus-width, we consider two types of bus-models: word-model and bit-model [13]. In the word-model, a broadcast bus consists of $\lceil \log n \rceil$ wires and conveys one word of data in one step; in the bit-model, a broadcast bus consists of a single wire and conveys only one bit of data in a step. We call the MSB (resp. MMPB) with word-model global bus by the word-model MSB (resp. MMPB). The bit-model MSB and MMPB are termed similarly. (Strictly speaking, the bit-model defined in [13] assumes that both the processor word-size and bus-width are constant. Here, we assume that only bus-width is constant, and that processor word-size is of $\lceil \log n \rceil$ for the mesh of size $n \times n$.)

A single time step of the MSB and the MMPB is composed of the following three sub-steps:

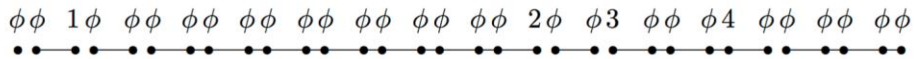
1) *Local communication sub-step:*
Every PE communicates with its adjacent PEs via local links.
2) *Broadcast sub-step:*
Every PE changes its switch configurations by local decision (this operation is only for the MSB). Then, along each broadcast bus segment, several of the PEs connected to the bus send data to the bus, and several of the PEs on the bus receive the data transmitted on the bus.
3) *Compute sub-step:*
Every PE executes some local computation.

Here, we assume that a PE writes to only one bus at a time in the MMPB model. The bus accessing capability is similar to that of the Common-CRCW PRAM model. If there is a write-conflict on a bus, the PEs on the bus receive a special value ⊥ (i.e., PEs can detect whether there is a write-conflict on a bus or not). If there is no data transmitted on a bus, the PEs on the bus receive a special value ϕ (i.e., PEs can know whether there is data transmitted on a bus or not).

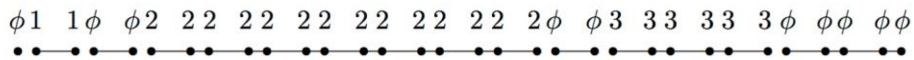
B. Port-Connectivity-Graph



(a) Broadcasts to be simulated. The processors P_1, P_9, P_{10}, and P_{12} respectively send data.



(b) The corresponding port-connectivity graph with initial labels.



(c) After connected-component labeling, vertices in each component C is labeled by the smallest initial label of all the vertices in C, with regarding ϕ as the greatest element.

Figure 3. Broadcasts on a separable bus along a row of the 16×16 MSB are simulated by connected-component labeling of the port-connectivity graph.

To simulate operations of the MSB by using the MMPB, we focus on how to mimic the broadcast sub-step of the MSB by using the MMPB, because the local communication and the compute sub-steps of the MSB can be easily simulated in a constant number of steps by the MMPB.

The simulation of broadcast sub-step can be achieved by connected-component labeling (CC-labeling) of a *port-connectivity graph* (pc-graph). See Fig. 3 (a) and (b) for an example. Vertices of the pc-graph correspond to read/write-ports of PEs, and edges stand for the port-to-port connections. Each vertex is initially labeled by the value that is sent through the corresponding port by the PE at the broadcast sub-step. If there is no data sent through the port, the vertex is labeled byϕ. The CC-labeling is done in such a way that vertices in each component C is labeled by the smallest initial label of all the vertices in C, with regarding ϕ as the greatest value (Fig. 3 (c)). These labels are called *component labels*. Obviously, the broadcast sub-step of the MSB can be simulated in $O(T)$ steps on the MMPB if the CC-labeling of the corresponding pc-graph can be solved in $O(T)$ steps by the MMPB. If there occurs collision on a bus segment, at least one of the senders can detect it by comparing its sending data with the component label obtained by the CC-labeling algorithm (e.g., P_{12} in Fig. 3 senses the collision). Then, by distributing such collision information using the CC-labeling algorithm, PEs can resolve the collision.

In Section III, we solve the CC-labeling problem by the divide-and-conquer strategy composed of the following three phases:

Phase 1: *{local labeling}*

Divide the pc-graph into sub-graphs, and label vertices locally within each sub-graph. These labels are called *local component labels*. In each sub-graph, also check whether the two vertices located at the boundary of the sub-graph are connected to each other or not (this connectivity information is used in Phase 2).

Phase 2: *{global labeling of boundary vertices}*

Label those vertices located at the boundary of each sub-graph with component labels.

Phase 3: *{local labeling for adjustment}*

Update vertex labels with component labels within each of the sub-graph for the consistency with Phase 2.

In the next section, we implement the above algorithm on the MMPB model.

III. SIMULATION ALGORITHM

In this section, we show that the $n \times n$ MSB can work with $O(\log^2 n)$ step slowdown even if its reconfigurable function is disabled. For clarity, let MMPB$^{\langle L\rangle}$ denote the MMPB that has L partitioned buses per row/column.

First, we prove that any step of the word-model MSB of size $n \times n$ can be simulated by the word-model MMPB$^{\langle L\rangle}$ of size $n \times n$ in $O\left(Ln^{1/(2L+1)}\right)$ steps even when L is non-constant. Next, we show that any step of the word-model MSB of size $n \times n$ can work with $O(\log^2 n)$ step slowdown even if we deprive the MSB of its reconfigurability, by considering the relation between the word-model bus and the bit-model one.

A. Simulation of the word-model MSB by the word-model MMPB

In [5, 6], we have proved the following lemma, assuming that *L* is a fixed constant.

Lemma 1 [5, 6] *Any step of the word-model MSB of size $n \times n$ can be simulated by the word-model MMPB<L> of size $n \times n$ in $O\left(\sqrt{\ell_L} + \sum_{j=1}^{L-1} \sqrt{\ell_j/\ell_{j+1}} + n/\ell_1\right)$ steps where L is a fixed constant.* ■

In this section, we show that we obtain the almost same result as Lemma 1, even if we assume that *L* is non-constant.

In what follows, we mainly focus on how to simulate the broadcasts along a row of the simulated MSB by using the corresponding row of the simulating MMPB. The simulation for columns can be achieved similarly.

To begin with, we introduce two fundamental results.

Lemma 2 [9] *The broadcasts taken on the separable bus in the row i of the word-model MSB of size $n \times n$ can be simulated in $O\left(\sqrt{\ell_1} + n/\ell_1\right)$ steps by the word-model MMPB<1> of size $n \times n$* .
■

Corollary 1 [9] *The broadcasts taken on the separable bus in the row i of the word-model MSB of size $n \times n$ can be simulated in $O\left(\sqrt{n}\right)$ steps by the word-model MMPB<1> of size $n \times n$ when $\ell_1 = n$* .
■

Then, we can prove the following lemma even if *L* is non-constant.

Lemma 3 *The broadcasts taken on the separable bus in the row i of the word-model MSB of size $n \times n$ can be simulated in the row i of the word-model MMPB<L> of size $n \times n$ in $O\left(\sqrt{\ell_L} + \sum_{j=1}^{L-1} \sqrt{\ell_j/\ell_{j+1}} + n/\ell_1 + L\right)$ steps.*

Proof: Let define $T_k(n)$, *U*(*n*), and *V*(*n*) as follows:

$T_k(n)$: the time cost for simulating the broadcasts taken along the separable bus in row *i* of the word-model MSB of size $n \times n$ using row *i* of the word-model MMPB<k>of size $n \times n$.

U(*n*): the time cost for simulating the broadcasts taken along the separable bus in row *i* of the word-model MSB of size $n \times n$ by using row *i* of the word-model MMPB<1>of size $n \times n$.

V(*n*): the time cost for simulating the broadcasts taken along the row *i* of the word-model MSB of size $n \times n$ by using row *i* of the word-model MMPB<1> of size $n \times n$ when $\ell_1 = n$.

From Lemma 2 and Corollary 1, there exist some constants c_1 and c_2 such that the following two inequalities hold:

$$U(n) \leq c_1 \left(\sqrt{\ell_1} + \frac{n}{\ell_1}\right), \tag{1}$$

$$V(n) \leq c_2\left(\sqrt{n}\right). \tag{2}$$

In what follows, we prove that the following equation holds for some constant *c*. The proof is done by mathematical induction on *k* ($k \geq 1$).

$$T_k(n) \leq c\left(2\sqrt{\ell_k} + 2\sum_{j=1}^{k-1} \sqrt{\frac{\ell_j}{\ell_{j+1}}} + \frac{n}{\ell_1} + k\right) \tag{3}$$

Here, without loss of generality, we assume $c \geq c_1, c_2$.

For the base case where $k = 1$, from Eq. (1) and $c \geq c_1$, we have

$$T_1(n) = U(n) \leq c_1\left(\sqrt{\ell_1} + \frac{n}{\ell_1}\right) \leq c\left(2\sqrt{\ell_1} + \frac{n}{\ell_1} + 1\right)$$

and thus Eq. (3) holds.

For the inductive case where $k > 1$, we prove Eq. (3), assuming that the following inductive hypothesis holds.

$$T_{k-1}(n) \leq c\left(2\sqrt{\ell_{k-1}} + 2\sum_{j=1}^{k-2} \sqrt{\frac{\ell_j}{\ell_{j+1}}} + \frac{n}{\ell_1} + k - 1\right) \tag{4}$$

Let P_j and P'_j respectively denote PE[i, j] of the $n \times n$ MSB and PE[i, j] of the $n \times n$ MMPB<k> ($0 \leq j < n$). Now, we explain how to implement the algorithm defined in Section II B. We divide the pc-graph corresponding to the broadcasts on the row separable bus into n/ℓ_k disjoint sub-graphs $G_0, G_1, \ldots, G_{(n/\ell_1)-1}$ of width ℓ_k. Here, we say that a sub-graph of pc-graph is of width *w* if it contains 2*w* vertices corresponding to the read/write-ports of *w* consecutive PEs. The CC-labeling of such defined pc-graph is carried out on the MMPB<k> as follows. We divide the row of the simulating MMPB<k> into n/ℓ_k disjoint blocks $B_0, B_1, \ldots, B_{(n/\ell_1)-1}$ in a way that each B_p consists of P'_j ($p\ell_k \leqq j < (p+1)\ell_k$). Note that each sub-graph G_p is processed by block B_p alone. Then, for each block B_p, since the PEs in B_p and a bus segment of the level-*k* partitioned bus can be seen as a linear processor array of ℓ_k PEs with a single broadcast bus of length ℓ_k, Phase 1 can be executed in $V(\ell_k)$ steps. As for Phase 2, the number of active PEs is $2n/\ell_k$, and each of those PEs can communicate in a constant time with next such PEs via either a local communication link or a bus segment of the level-*k* partitioned bus. Hence, by conveying the information of boundary vertices of each G_p to the leftmost PE in B_p, and letting the information be processed by the leftmost PE in B_p alone, Phase 2 is essentially the same problem as simulating the broadcast operation of the $1 \times n/\ell_k$ MSB using the $1 \times n/\ell_k$ MMPB<k-1> where each level-*j* partitioned buses are segmented by the length $\ell'_j = \ell_j/\ell_k$ ($1 \leq j < k$). Here, It should be noted that $\ell_l \geq \ell_{l+1}$ holds for each l ($1 \leq l < k$). The operations required for such adjustment (data transmission to/from the leftmost PE of each B_p, etc.) can be completed in a constant number of steps, and let c_3 be the time cost for them. Without violating the argument here, we assume that $c \geq c_3$ holds. (We can chose the constant *c* appropriately in advance so that $c \geqq c_3$ holds.) Then, from Eq. (4), Phase 2 can be completed in

$$T_{k-1}\left(\frac{n}{\ell_k}\right) + c_3$$

where level-*j* bus is segmented by $\ell'_j = \ell_j/\ell_k$

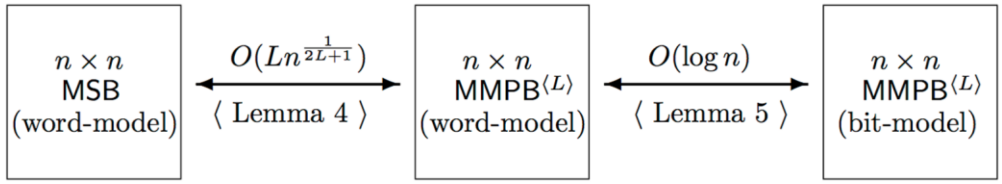


Figure 4. The simulation costs among the MSB and MMPBs.

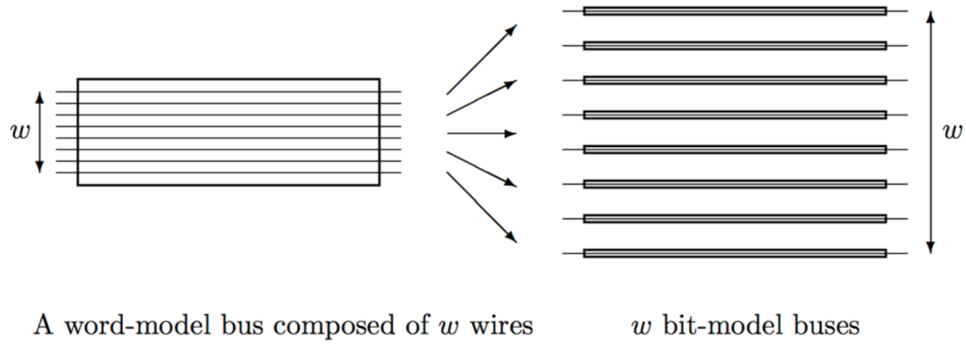


Figure 5. A word-model bus is decomposed to w bit-model buses where w is the word-length of processor.

$$\leq \langle \text{ Eq.}(4), c \geq c_3 \rangle$$

$$c\left(2\sqrt{\ell'_{k-1}} + 2\sum_{j=1}^{k-2}\sqrt{\frac{\ell'_j}{\ell'_{j+1}} + \frac{n}{\ell_k}/\ell'_1} + k\right)$$

$$=$$

$$c\left(2\sqrt{\frac{\ell_{k-1}}{\ell_k}} + 2\sum_{j=1}^{k-2}\sqrt{\frac{\ell_j}{\ell_k}\Big/\frac{\ell_{j+1}}{\ell_k} + \frac{n}{\ell_k}\Big/\frac{\ell_1}{\ell_k}} + k\right)$$

$$=$$

$$c\left(2\sum_{j=1}^{k-1}\sqrt{\frac{\ell_j}{\ell_{j+1}} + \frac{n}{\ell_1}} + k\right)$$

steps. Phase 3 can be done in $V(\ell_k)$ steps similarly to Phase 1. As a whole, the algorithm can be executed in

$$2V(\ell_k) + c\left(2\textstyle\sum_{j=1}^{k-1}\sqrt{\frac{\ell_j}{\ell_{j+1}} + \frac{n}{\ell_1}} + k\right)$$

$$\leq \langle \text{ Eq.}(2) \rangle$$

$$2c_2\sqrt{\ell_k} + c\left(2\sum_{j=1}^{k-1}\sqrt{\frac{\ell_j}{\ell_{j+1}} + \frac{n}{\ell_1}} + k\right)$$

$$\leq \langle\ c \geq c_2 \rangle$$

$$c\left(2\sqrt{\ell_k} + 2\sum_{j=1}^{k-1}\sqrt{\frac{\ell_j}{\ell_{j+1}} + \frac{n}{\ell_1}} + k\right)$$

steps, and thus Eq. (3) holds for $k > 1$ as well. The conclusion follows. ■

The local communication and the compute sub-steps of the MSB can be easily simulated in a constant number of steps by the MMPB. Hence, by letting each $\ell_j = \Theta(n^{\alpha_j})$ where $\alpha_j = 2(L - j + 1)/(2L + 1)$, we have the following lemma from Lemma 3.

Lemma 4 *Any step of the word-model MSB of size $n \times n$ can be simulated by the word-model MMPB$^{<L>}$ of size $n \times n$ in $O(Ln^{1/(2L+1)})$ steps.* ■

B. Simulation of the word-model MSB by the bit-model MMPB

In this section, we show that any step of the word-model MSB of size $n \times n$ can work with $O(\log^2 n)$ step slowdown even if we deprive the MSB of its reconfigurable function, by considering the relation between the word-model bus and the bit-model one.

First, we prove the following lemma.

Lemma 5 *Any step of the word-model MMPB$^{<L>}$ of size $n \times n$ can be simulated by the bit-model MMPB$^{<L>}$ of size $n \times n$ in $O(\log n)$ steps.*

Proof: The $\lceil \log n \rceil$ bits of one word data can be conveyed sequentially in $\lceil \log n \rceil$ steps, one bit per step, in the bit-model MMPB$^{<L>}$. ■

We illustrate the results of Lemma 4 and 5 in Fig. 4. Obviously, Fig. 4 implies the following lemma:

Lemma 6 *Any step of the word-model MSB of size $n \times n$ can be simulated by the bit-model MMPB$^{<L>}$ of size $n \times n$ in $O(Ln^{1/(2L+1)} \log n)$ steps.* ■

By letting $L = \log n$, we obtain the following corollary.

Corollary 2 *Any step of the word-model MSB of size $n \times n$ can be simulated by the bit-model MMPB$^{<\log n>}$ of size $n \times n$ in $O(log^2 n)$ steps.*

Proof: By letting $L = \log n$, we can calculate the time-complexity as follows:

$$O\left(Ln^{1/(2L+1)}\log n\right)$$

$$= \langle\ L = \log n\ \rangle$$

$$O\left(n^{1/(2\log n+1)}\log^2 n\right)$$

$$= \langle\ n^{1/(2\log n+1)} \leq c \quad \text{for some constant } c\ \rangle$$

$$O(\log^2 n)$$

Thus, the conclusion follows. ■

Since the word-model MSB of size $n \times n$ has $\log n$ wires for each row/column, we can view a word-model bus as $\log n$ bit-model buses (Fig. 5). Hence, without increasing any circuit-complexity, we obtain the bit-model MMPB$^{<\log n>}$ of size $n \times n$ from the word-model MSB of size $n \times n$. With this observation and Corollary 2, we can state the main theorem of this paper as follows:

Theorem 1 *Any step of the word-model MSB of size $n \times n$ can work with $O(\log^2 n)$ step slowdown even if its reconfigurable capability is unused.* ■

IV. CONCLUDING REMARKS

In this paper, we showed that the word-model MSB of size $n \times n$ can work with $O(\log^2 n)$ step slowdown even if its reconfigurable capability is unused. We obtain the result from these two facts: 1) every global bus of the word-model MSB of size $n \times n$ consists of $\lceil \log n \rceil$ wires, and 2) we can obtain the bit-model MMPB of size $n \times n$ with $L = \lceil \log n \rceil$ from the word-model MSB of size $n \times n$ without increasing circuit-complexity. In [7], we have proposed more efficeint algorithm that exploits the pipeline technique heavily. Although the simulation algorithm presented here is slower than the one in [7] by the factor of $\log n$, the key ideas and explanations are much simpler than those in [7].

From a practical viewpoint, we expect that the communication latency of the broadcast buses of the MMPB is much smaller than that of the MSB. Each broadcast bus of the MSB of size $n \times n$ can form the broadcast bus whose length is n, and such a bus contains $O(n)$ sectioning switch elements in it. As for the MMPB of size $n \times n$, though the bus length is also at most n, but no switch element is inserted to the bus because it has no sectioning switch. Hence, compared to the MSB, the MMPB model has an advantage that each broadcast bus has smaller propagation delay introduced by the switch elements inserted into the bus (i.e., device propagation delay), and thus our simulation algorithm is practically useful when the mesh size becomes so large that we cannot neglect the delay. In future work, we will study the effectiveness of our simulation algorithm, by taking into account the propagation delay.

REFERENCES

[1] Y. Ben-Asher, D. Gordon, A. Schuster, "Efficient self-simulation algorithms for reconfigurable arrays," J. of Parallel and Distributed Computing 30 (1), 1−22, 1995

[2] T. Maeba, M. Sugaya, S. Tatsumi, K. Abe, "An influence of propagation delays on the computing performance in a processor array with separable buses," IEICE Trans. A J78-A (4), 523−526, 1995

[3] T. Maeba, S. Tatsumi, M. Sugaya, "Algorithms for finding maximum and selecting median on a processor array with separable global buses," IEICE Trans. A J72-A (6), 950−958, 1989

[4] S. Matsumae, "Simulation of meshes with separable buses by meshes with multiple partitioned buses," Proc. of the 17th International Parallel and Distributed Processing Symposium (IPDPS2003), IEEE CS press, 2003.

[5] S. Matsumae, "Optimal simulation of meshes with dynamically separable buses by meshes with statically partitioned buses," Proc. of the 2004 International Symposium on Parallel Architectures, algorithms and Networks (I-SPAN'04), IEEE CS Press, 2004.

[6] S. Matsumae, "Tight bounds on the simulation of meshes with dynamically reconfigurable row/column buses by meshes with statically par-titioned buses," J. of Parallel and Distributed Computing 66 (10), 1338−1346, 2006

[7] S. Matsumae, "Impact of reconfigurable function on meshes with row/column buses," International Journal of Networking and Computing 1 (1), 36−48, 2011

[8] S. Matsumae, "Effective Partitioning of Static Global Buses for Small Processor Arrays," Journal of Information Processing Systems 7 (1), 2011

[9] S. Matsumae, N. Tokura, "Simulating a mesh with separable buses," Transactions of Information Processing Society of Japan 40 (10), 3706−3714, 1999

[10] S. Matsumae, N. Tokura, "Simulation algorithms among enhanced mesh models," IEICE Transactions on Information and Systems E82−D (10), 1324−1337, 1999

[11] R. Miller, V. K. Prasanna-Kumar, D. Reisis, Q. F. Stout, "Meshes with reconfigurable buses," Proc. of the 5th MIT Conference on Advanced Research in VLSI, Boston, 1988.

[12] M. J. Serrano, B. Parhami, "Optimal architectures andalgorithms for mesh-connected parallel computers with separable row/column buses," IEEE Trans. Parallel and Distributed Systems 4 (10), 1073−1080, 1993

[13] R. Vaidyanathan, J. L. Trahan, Dynamic Reconfiguration, Kluwer Academic/Plenum Publishers, 2004.

[14] B. Wang, G. Chen, "Constant time algorithms for the transitive closure and some related graph problems on processor arrays with reconfigurable bus systems," IEEE Trans. Parallel and Distributed Systems 1 (4), 500−507, 1990

Post-Editing Error Correction Algorithm For Speech Recognition using Bing Spelling Suggestion

Youssef Bassil, Mohammad Alwani
LACSC – Lebanese Association for Computational Sciences
Beirut, Lebanon

***Abstract*—ASR short for Automatic Speech Recognition is the process of converting a spoken speech into text that can be manipulated by a computer. Although ASR has several applications, it is still erroneous and imprecise especially if used in a harsh surrounding wherein the input speech is of low quality. This paper proposes a post-editing ASR error correction method and algorithm based on Bing's online spelling suggestion. In this approach, the ASR recognized output text is spell-checked using Bing's spelling suggestion technology to detect and correct misrecognized words. More specifically, the proposed algorithm breaks down the ASR output text into several word-tokens that are submitted as search queries to Bing search engine. A returned spelling suggestion implies that a query is misspelled; and thus it is replaced by the suggested correction; otherwise, no correction is performed and the algorithm continues with the next token until all tokens get validated. Experiments carried out on various speeches in different languages indicated a successful decrease in the number of ASR errors and an improvement in the overall error correction rate. Future research can improve upon the proposed algorithm so much so that it can be parallelized to take advantage of multiprocessor computers.**

Keywords- Speech Recognition; Error Correction; Bing Spelling Suggestion.

I. INTRODUCTION

With the ever increasing number of computer-based applications, modern digital computers are no more solely used for crunching numbers and performing high-speed mathematical computations. Instead, they are currently being used for a wider spectrum of tasks including gaming, sound editing, text editing, aided-design, industrial control, medical diagnosis, communication, and information sharing over the World Wide Web. As a matter of fact, computer scientists and researchers from all over the globe have been rigorously carrying out novel innovations and developing groundbreaking solutions to automate every area of life. Automatic Speech Recognition (ASR) is one of the most evolving computing fields that has already been exhaustively employed for an assortment of applications including but not limited to automated telephone services (ATS), voice user interface (VUI), voice-driven industrial control systems (ICS), speech-driven home automation systems (Domotics), speech dictation systems, and automatic speech-to-text systems (STT). Lately, ASR has been of great attraction to computer researchers, manufacturers, and consumers [1]. At heart, the task of ASR is to transform an acoustic waveform into a string of words that can be manipulated by a computing machine [2]. It is thereby abridging the complexity of man-machine interface (MMI) by replacing conventional input devices with an easier, faster, more efficient, and more natural method, allowing users to seamlessly operate, control, and manage computer systems [3]. However, ASR systems are still error-prone and inaccurate especially if they are deployed in an inadequate environment [4]. Generally, ASR errors are manifested by lexical misspellings and linguistic mistakes in the recognized output text, and are primarily caused by the excessive noise in the surroundings, the quality of the speech, the dialect and the utterance of the discourse, and the vocabulary size of the ASR system [4], [5].

In an attempt to reduce the number of errors generated by ASR systems and improve their accuracy, several error-correction techniques were envisioned, some of which are manual as they post-edit the recognized output transcript to correct misspellings; while, others are enhanced acoustic mathematical models aimed at improving the interpretation of the input waveform to prevent errors at early stages [6]. Despite all these endeavors, ASR errors are still at their peak as the mainstream error-correction algorithms are still far from perfect and word errors in speech recognition are always the rule, rather than the exception.

This paper proposes an automatic post-editing context-based real-word error correction approach based on Bing web search engine's spelling suggestion technology [7], to detect and correct linguistic and lexical errors generated by ASR recognition systems. Post-editing (i.e. post-processing) implies that detecting and correcting errors are done after the input wave has been transformed into text. Algorithmically, after the speech has been recognized and converted into text, a list of word-tokens t are generated from the text and then sent successively to Bing search engine as search queries. If Bing returns an alternative spelling suggestion for ti in the form of "Including results for ci", where ci is the suggested correction for ti, then ti is said to contain some misspelled words and ci is its predictable substitute correction; otherwise no correction is needed for ti and the next token is processed. Ultimately, when all tokens get validated, all the initial correct tokens { t1…tn }, in addition to the corrected ones { c1…cm } are concatenated together, yielding to a new transcript with fewer misspellings.

II. ERRORS IN ASR SYSTEMS

Despite the latest developments in ASR systems, they still exhibit misspellings and linguistic errors in the output text. An evaluation conducted at IBM [8] to measure the number of errors generated by speech recognition dictation systems that were operated by IBM employees, showed that these systems were committing an average of 105 errors per minute, most of which can only be corrected by manually post-editing the text after the end of the speaking. In effect, these errors are caused by two factors: external and internal factors.

A. External Factors

The noise in the environment is one of the most key external factors that determine the error rate in ASR systems. If the recognition process is to occur in a quiet location rather than in a noisy open place, superior and accurate results can be attained. It is worth noting that in addition to the raw noise in the setting, the quality of the input devices has a collateral influence in raising the SNR (Signal-to-Noise) ratio. For this reason, high-quality expensive microphones and audio systems are often used to subtly filter the background noise and eliminate the Hiss effect in the input signal which eventually helps exalting the overall precision of the ASR system.

Another weighty factor that needs to be considered is the type of speech being recognized; it is either isolated-word speech or continuous speech. In effect, the recognition of isolated-word speech such as control, telephony, and voice user interface systems is far much easier than the recognition of continuous speech such as dictation or translation systems due to the abundance of pauses in the discourse which makes it less complicated to process and less resource demanding.

Last but not least, the dialect and the speech utterance have a weighty effect on ASR errors. In fact, the dialect of a language varies from epoch to epoch, from country to country, from region to region, and from speaker to speaker. Basically, the accent of non-native speakers makes it harder for ASR systems to recognize and interpret the speech. It has been reported that the error rate is four times higher for non-native speakers than for native speakers [9]. Moreover, the way words are uttered has a direct impact on the recognition system as a whole, for instance, people with a quivering and wavering voice such as children and handicapped may create some hitches during the recognition process.

B. Internal Factors

The Internal factors that are responsible for the emergence of ASR errors typically arose from within the components of ASR systems. Inherently, an ASR system is composed of an acoustic model (AM) based on a phonetic lexicon, and a language model (LM) based on an n-gram lexicon [10], [11], [12].

The acoustic model (AM) which computes the likelihood of the observed input phoneme given linguistic units (phones) is based on a lexicon or a dictionary of words with their corresponding phones and pronunciations. These phones are used to recognize the spoken words. Consequently, a deficiency in the dictionary to cover all possible pronunciations would prevent the system from correctly identifying the words in a speech. This situation is often referred to as out-of-vocabulary (OOV) which usually occurs when an ASR system is unable to match a spoken word with any of the entries in its phonetic lexicon [10].

The language model (LM) which approximates how likely a given word is next to occur in a particular text, depends on a probabilistic n-gram model trained on specific corpus of text to predict the next word in a sequence of spoken words. Since it is practically impossible to find a corpus containing all valid words of a language, mismatches and ambiguities would befall during speech recognition, leading subsequently to an increase in the ASR error rate.

As a result and since an ASR system is exclusively based on two types of lexicons; one phonetic of static pronunciations and one probabilistic of n-gram collocations, the larger the vocabulary these lexicons have, the more accurate and the least erroneous the recognition process is considered to be.

III. RELATED WORK

Different error correction techniques exist, whose purpose is to detect and correct misspelled words generated by ASR systems. Broadly, they can be broken down into several categories: Manual error correction, error correction based on alternative hypothesis, error correction based on pattern learning, and post-editing error correction.

In manual error correction, a staff of people is hired to review the output transcript generated by the ASR system and correct the misspelled words manually by hand. This is to some extent considered laborious, time consuming, and error-prone as the human eye may miss some errors.

Another category is the alternative hypothesis error correction in which an error is replaced by an alternative word-correction called hypothesis. The chief drawback of this method is that the hypothesis is usually derived from a lexicon of words; and hence it is susceptible to high out-of-vocabulary rate. In that context, Setlur, Sukkar, and Jacob [13] proposed an algorithm that treats each utterance of the spoken word as hypothesis and assigns it a confidence score during the recognition process. The hypothesis that bypasses a specific threshold is to be selected as the correct output word. The experiments showed that the error rate was reduced by a factor of 0.13%.

Likewise, Zhou, Meng, and Lo [14] proposed another algorithm to detect and correct misspellings in ASR systems. In this approach, twenty alternative words are generated for every single word and treated as utterance hypotheses. Then, a linear scoring system is used to score every utterance with certain mutual information, calculated from a training corpus. This score represents the number of occurrence of this specific utterance in the corpus. After that, utterances are ranked according to their scores; the one with the highest score is chosen to substitute the detected error. Experiments conducted, indicated a decrease in the error rate by a factor of 0.8%.

Pattern learning error correction is yet another type of error correction techniques in which error detection is done through finding patterns that are considered erroneous. The system is first trained using a set of error words belonging to a specific domain. Subsequently, the system builds up detection rules that can pinpoint errors once they occur. At recognition time, the ASR system can detect linguistic errors by validating the output text against its predefined rules.

The drawback of this approach is that it is domain specific; and thus, the number of words that can be recognized by the system is minimal. In this perspective, Mangu and Padmanabhan [15] proposed a transformation-based learning algorithm for ASR error correction. The algorithm exploits confusion network to learn error patterns while the system is offline. At run-time, these learned rules assist in selecting an alternative correction to replace the detected error. Similarly, Kaki, Sumita, and Iida, [16] proposed an error correction algorithm based on pattern learning to detect misspellings and on similarity string algorithm to correct misspellings. In this technique, the output recognized transcript is searched for potential misspelled words. Once an error pattern is detected, the similarity string algorithm is applied to suggest a correction for the error word. Experiments were executed on a Japanese speech and the results indicated an overall 8.5% reduction in ASR errors. In a parallel effort, statistical-based pattern learning techniques were also developed. Jung, Jeong, and Lee [17] employed the noisy channel model to detect error patterns in the output text. Unlike other pattern learning techniques which exploit word tokens, this approach applies pattern learning on smaller units, namely individual characters. The global outcome was a 40% improvement in the error correction rate. Furthermore, Sarma and Palmer [18] proposed a method for detecting errors based on statistical co-occurrence of words in the output transcript. The idea revolves around contextual information which states that a word usually appears in a text with some highly co-occurred words. As a result, if an error occurs within a specific set of words, the correction can be statistically deduced from the co-occurred words that often appear in the same set.

The final type of error correction is post-editing. In this approach, an extra layer is appended to the ASR system with the intention of detecting and correcting misspellings in the final output text after recognition of the speech is completed. The advantage of this technique is that it is loosely coupled with the inner signal and recognition algorithms of the ASR system; and thus, it is easy to be implemented and integrated into an existing ASR system while taking advantage of other error correction explorations done in sister fields such as OCR, NLP, and machine translation.

As an initial attempt, Ringger and Allen [19] proposed a post-processor model for discovering statistical error patterns and correct errors. The post-processor was trained on data from a specific domain to spell-check articles belonging to the same domain. The actual design is composed of a channel model to detect errors generated during the speech recognition phase, and a language model to provide spelling suggestions for those detected errors. As outcome, around 20% improvement in the error correction rate was achieved. On the other hand, Ringger and Allen [20] proposed a post-editing model named SPEECHPP to correct word errors generated by ASR systems. The model uses a noisy channel to detect and correct errors, in addition to the Viterbi search algorithm to implement the language model. Another attempt was presented by Brandow and Strzalkowski [21] in which, text generated from the ASR system is collected and aligned with the correct transcription of the same text. In a training process, a set of correction rules are generated from these transcription texts and validated against a generic corpus; rules that are void or invalid are discarded. The system loops for several iterations until all rules get verified. Finally, a post-editing stage is employed which harnesses these rules to detect and correct misspelled words in the ASR generated transcript.

IV. PROPOSED SOLUTION

This paper proposes a new post-editing approach and algorithm for ASR error correction based on Bing's spelling suggestion technology [7]. The idea hinges around using Bing's enormous indexed data to detect and correct real-word errors that appear in the ASR recognized output text. In other words, error correction is applied to spell check the final text that resulted from the transformation of the input wave into text; and hence is referred to as post-editing error correction. The algorithm starts first by chopping the ASR output text into several word tokens. Then, each token is sent to Bing's web search engine as a search query. If this query contains a misspelled word, Bing suggests a spelling correction for it, and consequently, the algorithm replaces it with the spelling suggestion.

A. Bing's Spelling Suggestion

Bing's spelling suggestion technology can suggest alternative corrections for the often made typos, misspellings, and keyboarding errors. At the core, Bing has a colossal database of billions of online web pages containing trillions of term collections and n-gram words that can be used as groundwork for all kinds of linguistic applications such as machine translation, speech recognition, spell checking, as well as other types of text processing problems. Fundamentally, Bing's spelling suggestion algorithm is based on the probabilistic n-gram model originally proposed by Markov [22] for predicting the next word in a particular sequence of words. In brief, an n-gram is simply a collocation of words that is n words long.

For instance, "The boy" is a 2-gram phrase also referred to as bigram, "The boy scout" is a 3-gram phrase also referred to as trigram, "The boy is driving his car" is a 6-gram phrase, and so forth. The Bing's algorithm automatically examines every single word in the search query for any possible misspellings. It tries first to match the query, basically composed of ordered association of words, with any occurrence alike in Bing's database of indexed web pages; if the number of occurrence is high, then the query is considered correct and no correction is to take place.

However, if the query was not found, Bing uses its n-gram statistics to deduce the next possible correct word in the query. Sooner or later, an entire suggestion for the whole misspelled query is generated and displayed to the user in the form of "Including results for spelling-suggestion". For example, searching for the word "conputer" drives Bing to suggest "Including results for Computer". Likewise, searching for "The hord disk sturage" drives Bing to suggest "Including results for the hard-disk storage". Searching for the proper name "jahn cenedy" drives Bing to suggest "Including results for John Kennedy". Figure 1-3 show the spelling suggestions returned by Bing search engine when searching for "conputer", "The hord disk sturage", and "jahn cenedy" respectively.

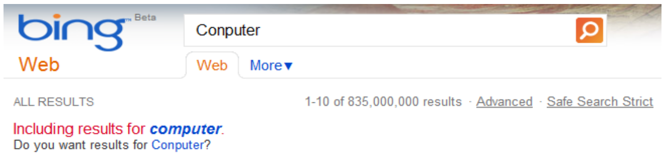


Figure 1. Spelling suggestion for "conputer"

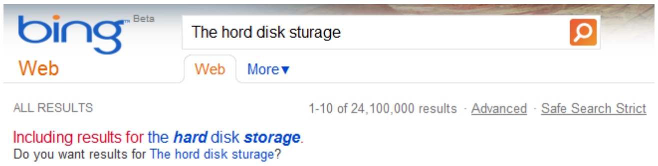


Figure 2. Spelling suggestion for "The hord disk sturage"

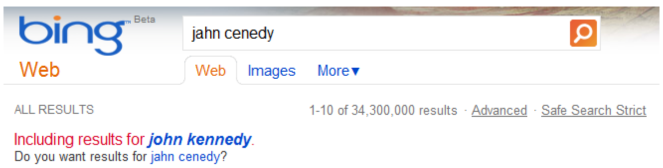


Figure 3. Spelling suggestion for "jahn cenedy"

B. ASR Model

The proposed error correction method is executed during the post-editing stage of an ASR system, and is based on Bing's spelling suggestion. At early stages, prior to post-editing the recognized text using the proposed algorithm, a standard ASR system is fed by an input waveform that represents the speech to be recognized. Then, the signal is digitally processed in order to extract its spectral features and audible phones. Afterwards, the likelihood of an observed phoneme given an extracted spectral feature is computed by the acoustic model (AM) and its Hidden Markov Model (HMM) and phonetic lexicon. In parallel, the language model (LM) computes the probability of the obtained phone to occur in the language. Finally, a decoding module statistically infers the spoken words and generates the final output text.

The proposed model further processes the obtained output and adds a post-editing stage to the system with the purpose of detecting and correcting any possible misspelled words that were generated during the recognition process. In essence, the output text that is obtained from the decoding module is broken down into a collection of tokens, each made out of six words. In a sequential fashion, these tokens are sent one after the other to Bing search engine as search parameters. If Bing does not return a spelling suggestion, then it is evident that the query contains no misspelled words; and thus no correction is needed for this particular token and no changes is to occur for the original text.

On the other hand, if Bing returns a spelling suggestion, then definitely the query contains some misspelled words; and thus a correction is required for this particular token of words. The correction consists of replacing the original token in the text by the Bing's suggested correction. Figure 4 depicts a block diagram for a generic ASR system, however modified by adding to it a post-editing layer to perform error correction using Bing's spelling suggestion.

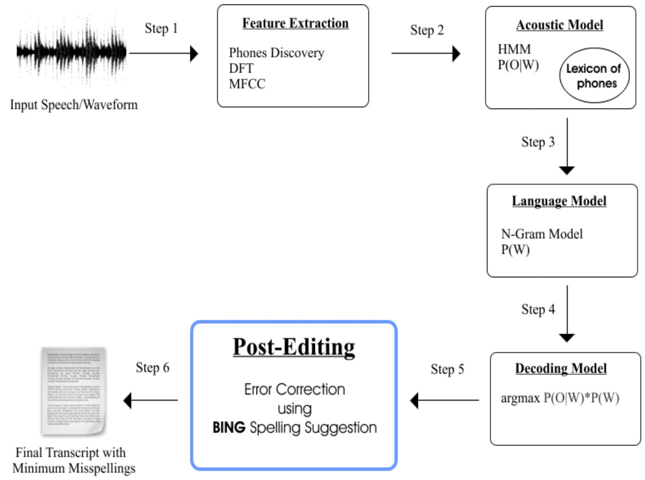


Figure 4. ASR system with a post-editing layer for error correction

C. The Error Correction Algorithm

The proposed algorithm comprises several steps to be executed in order to detect and correct *ASR* misspellings. The algorithm takeoffs by dividing the recognized output transcript into several tokens $T=\{ t_1 \ldots t_n \}$, each composed of 6 words, $t_i=\{w_0,w_1,w_2,w_3,w_4,w_5\}$ where t_i is a particular token and w_j is a single word or term in that token. Then, every t_i is sent to be validated using Bing search engine. The search results returned by Bing are then parsed to identify whether or not they contain the "*Including results for* c_i" spelling suggestion message, where c_i is the suggested correction for t_i. If true, then token t_i must contain a certain misspelled word; and hence, t_i is replaced by c_i. Ultimately, after all tokens get validated, all original correct tokens $O=\{ t_1 \ldots t_k \}$, plus the corrected ones $C=\{ c_1 \ldots c_p \}$ are concatenated together, yielding to a new text with fewer misspellings represented formally as $V=\{ v_1 \ldots v_{k+p} \}$. The post-editing process then finishes and the algorithm halts. Figure 5 summarizes the flow of execution for the proposed algorithm.

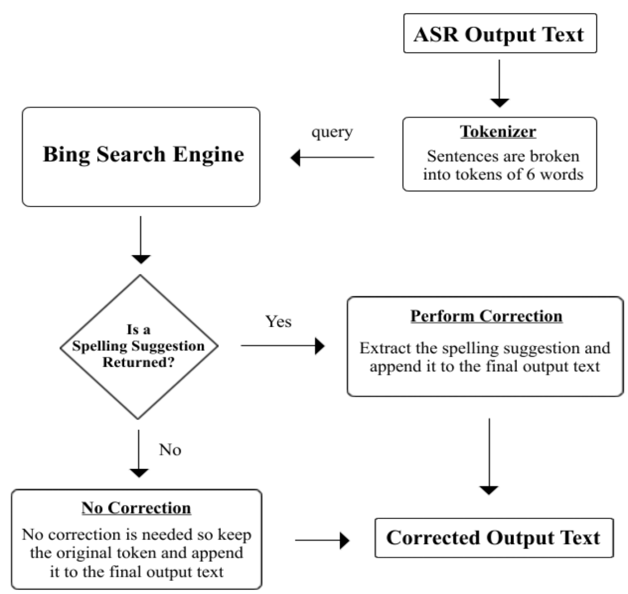


Figure 5. Flowchart illustrating the different steps of the proposed algorithm

D. The Pseudo-Code

The following pseudo-code describes the entire logic behind the proposed algorithm, independently of any programming language platform.

// the purpose of this procedure is to correct ASR spelling errors using Bing spelling suggestion

```
// INPUT: ASR recognized text possibly containing errors and misspellings
// OUTPUT: Corrected text
START
Procedure Post-Editing(asr_text)
{
   // breaks the asr_text into blocks of 6 words each
   tokens ← Tokenize(asr_text, 6)

   // iterates until all tokens are exhausted
   for (i←0 to tokens_length)
   {
     // send tokens[i] to Bing search engine
     results ← BingSearch(tokens[i])
                    if(results contains("Including results for")
       // indicates some misspellings in tokens[i]
       output ← getSuggestion(results)
       // extract correction and append it to output file
     else
       output ← tokens[i]
       // no misspellings so add the original tokens[i]
   }
   RETURN output
}
FINISH
```

The procedure *Post-Editing()* contains one for loop that is executed *n* times, where *n* is the total number of tokens in the *ASR* text. Considering "*results ← BingSearch(tokens[i])*" as the basic operation, the time complexity of the algorithm is described as follows:

$$\sum_{i=0}^{n} 1 = n \text{ and thus the algorithm is of time complexity } O(n)$$

Since the basic operation is to be executed *n* times regardless of the content of the input *ASR* text, $C_{Best}(n)= C_{Worst}(n)= C_{Average}(n)= n$ = *number of tokens in the original ASR text*

V. EXPERIMENTS & RESULTS

In the experiments, speech recognition was performed on two speeches in two different languages: English [23] and French [24]. Bing.com was used to post-edit the English speech, while Bing.fr was used to post-edit the French speech. As for the ASR software, a custom proprietary application program based on Microsoft Speech Application Programming Interface (SAPI 5.0) engine [25] was utilized to perform the speech recognition of the two input speeches.

The proposed post-editing algorithm was implemented using MS C# 4.0 under the MS .NET Framework 4.0 and the MS Visual Studio 2010.

The following paragraph is the input English speech to be processed by the ASR software.

Virtual machine applications such as VMWare Workstation, Sun Virtualbox, and Microsoft Virtual PC now allow you to boot the second operating system on top of your main OS, eliminating the need and hassle of rebooting into another OS. Installing a NICs driver into a Windows, Macintosh, or Linux system is easy: just insert the driver CD when prompted by the system. Unless you have a very offbeat NIC, the operating system will probably already have the driver preinstalled, but there are benefits to using the driver on the manufacturer CD. IEEE could use the traditional Physical layer mechanisms defined by the Ethernet standard. But, there was already in place a perfectly usable 10 Gbps fiber network, called SONET, used for wide area networking (WAN) transmissions. Microsoft pushed the idea of a single client tunneling into a private LAN using software. Cisco, being the router king that it is, came up with its own VPN protocol called Layer 2 Tunneling Protocol

The subsequent paragraph represents the output transcript generated by the ASR system along with all the misspellings (underlined) that were produced during the recognition process.

Virsual machine applications such as VWare Workstation, Sun Virualbox, and Micro soft Virsual PC now allow you to boot the second operaing system on tat of your main OS, eliminating the need and hassl of reboting into another OS. Installing a NICs driver into a Windoos, Makintosh, or Linix system is easy: just insert the driver CD when promptd by the system. Unless you have a very offbeet NIC, the operating system will probably already have the driver pre-installed, but there are benefits to using the driver on the manufachurer CD. IEEE could use the traditional Physical layer mechanisms defined by the Ethernit standard. But, there was already in place a perfectly usable 10 Gbps fiber network, called SONETT, used for wide area networking (WAAN) transmissions. Micro soft pushed the idea of a single client tulleling into private LAAN using software. Ciskow, being the router king that it is, came up with its own VPN protocol called Layer 2 Tunneling Protocol

Next is the same previous transcript, however, error-corrected using the proposed post-editing error correction algorithm. Underlined are the words that were not corrected.

Virtual machine applications such as VMWare Workstation, Sun Virtualbox, and Microsoft Virtual PC now allow you to boot the second operating system on tat of your main OS, eliminating the need and hassl of rebooting into another OS. Installing a NIC driver into a Windows, Macintosh, or Linux system is easy: just insert the driver CD when prompts by the system. Unless you have a very offbeat NIC, the operating system will probably already have the driver preinstalled, but there are benefits to using the driver on the manufacturer CD. IEEE could use the traditional Physical layer mechanisms defined by the Ethernet standard. But, there was already in place a perfectly usable 10-Gbps fiber network, called SONETT, used for wide area networking (WAN) transmissions. Microsoft pushed the idea of a single client tulle long into a private LAN using software. Cisco, being the router king that it is, came up with its own VPN protocol called Layer 2 Tunneling Protocol

The English recognized transcript comprehended 23 misspelled words out of 161 total words (number of words in the whole speech), making the error rate close to E = 23/161 = 0.142 = 14.2%. Several of these errors were proper names such as “Microsoft”, others were technical words such as “LAN”, “Macintosh”, “Linux”, “VMWare”, and “Ethernet”, and the remaining ones were regular English words such as “virtual”, “operating”, “rebooting”, “hassle”, etc. When the proposed post-editing error correction algorithm was applied, 18 misspelled words out of 23 were corrected successfully, leaving only 5 non-corrected errors and they were as follows: “promptd” was miss-corrected as “prompts”, “tulleing” was miss-corrected as “tuelle long”, and “tat”, “hassl”, and “SONETT” were not corrected at all. As a result, the error rate using the proposed algorithm was close to E = 5/161 = 0.031 = 3.1%. Consequently, the improvement can be calculated as I = 0.142/0.031 = 4.58 = 458%, that is increasing the rate of error detection and correction by a factor of 4.58.

Another experiment was conducted on a French speech and it is delineated below:

Enfin pour nuancer les sens attribué à ces quatre directions de l’espace, M. Monod a proposé une combinaison du haut et du bas avec l’orientation à droite et à gauche. Les deux zones gauches sont charactérisées par des élément plus passifs dans la psychologie de l’individu et sont associées au passé de celui ci dans la zone bas gauche. Dans la zone droite on retrouve la même distinction entre les facteurs les plus dynamiques en haut droit et les processus de socialisation plus anciens en bas droit. Les crayons de couleur constituent un stimulus banal d’ou son impact sur le sujet est moins fort que celui des planches d’encre de Rorschach.

The subsequent paragraph represents the output transcript generated by the ASR system along with all the misspellings (underlined) that were produced during the recognition process.

Enfin pur nuancer les sence attribué à ces quatre directions de l’espace, M. Mono a proposé une combinaison du haut et du bas avec l’orientation à droite et à gouche. Les deux zones gouches sont charactérisées par des élément plus passivs dans la psycologie de l’inditu et sont associées au passé de celui ci dans la zone bas gouche. Dans la zone droite on retruve la même distinction entre les facdeurs les plus dynamiques en haot droit et les processuse de socialisation plus anciens en bas droit. Les craiyons de couleur constituent un stimulus banal dou son impact sur le sujet est moins fort que celui des planches d’encre de Roschah.

Next is the same previous transcript, however error-corrected using the proposed post-editing error correction algorithm. Underlined are the words that were not corrected.

Enfin pour nuancer les sens attribué à ces quatre directions de l’espace, M. Mono a proposé une combinaison du haut et du bas avec l’orientation à droite et à gauche. Les deux zones gauches sont charactérisées par des élément plus passive dans la psychologie de l’individu et sont associées au passé de celui ci dans la zone bas gauche. Dans la zone droite on retrouve la même distinction entre les facteurs les plus dynamiques en haut droit et les processus de socialisation plus anciens en bas droit. Les crayons de couleur constituent un stimulus banal dou son impact sur le sujet est moins fort que celui des planches d’encre de Rorschach.

The French recognized transcript comprehended 16 misspelled words out of 110 total words (number of words in the whole speech), making the error rate close to E = 16/110 = 0.145 = 14.5%. Several of these errors were proper names such as “Rorschach”, and others were regular French words such as “pour”, “gauche”, “retrouve”, “crayons”, etc. When the proposed post-editing error correction algorithm was applied, 13 misspelled words out of 16 were corrected successfully, leaving only 3 non-corrected errors and they were as follows: “passivs” was miss-corrected as “passive”, and “Mono” and “dou” were not corrected at all. As a result, the error rate using the proposed algorithm was close to E = 3/110 = 0.027 = 2.7%. Consequently, the improvement can be calculated as I = 0.145/0.027 = 5.37 = 537%, that is increasing the rate of error detection and correction by a factor of 5.37.

VI. Experiments Evaluation

The experiments conducted on the proposed ASR post-editing error correction algorithm evidently showed a 458% improvement in the error correction rate for English speech and 537% for French speech. In other terms, around 4.5 times more English errors were detected and corrected, and 5.3 times more French errors were detected and corrected. On average, the proposed algorithm improved the error correction rate by I= (458% + 537%) / 2 = 497%, that is increasing the overall rate of error detection and correction by a factor of 4.9. Table 1 summarizes the experimental results obtained for the proposed ASR error correction algorithm before and after post-editing.

TABLE I. Experimental Results Before and After Post-Editing

	English Document Total words = 161	**French Document Total words = 110**
Number of errors resulted before post-editing	23	16
Number of errors resulted after post-editing	5	3
Error rate before post-editing	14.2%	14.5%
Error rate after post-editing	3.1%	2.7%
Improvement ratio	4.58 (458%)	5.37 (537%)

VII. Conclusions

This paper presented a new ASR post-editing error correction method based on Bing’s online spelling suggestion technology. The backbone of this technology is a large dataset of words and sentences indexed by Bing and originally extracted from several online sources including web pages, documents, articles, and forums. This allows Bing to suggest common spellings for queries containing errors and linguistic mistakes. For this reason, the proposed algorithm excelled in detecting and correcting ASR errors as it fully harnessed Bing’s online spelling suggestion to spell-check the ASR recognized output text. Experiments carried out, indicated a noticeable reduction in the number of ASR errors, yielding to an outstanding improvement in the ASR error correction rate.

VIII. Future Work

As future work, various ways to parallelize the proposed algorithm are to be investigated so as to take advantage of multiprocessors and distributed computers. The projected outcome would be a faster algorithm of time complexity O(n/p), where n is the total number of word tokens to be spell checked and p is the total number of processors.

Acknowledgment

This research was funded by the Lebanese Association for Computational Sciences (LACSC), Beirut, Lebanon under the "Web-Scale Speech Recognition Research Project – WSSRRP2011".

References

[1] L. Rabiner and B.H. Juang, Fundamentals of Speech Recognition, Prentice Hall PTR, Englewood Cliffs, New Jersey, 1993.

[2] T. Dutoit, An Introduction to Text-to-Speech Synthesis (Text, Speech and Language Technology), Springer, 1st ed, 1997.

[3] W. A. Lea, "The value of speech recognition systems", Readings in Speech Recognition, Morgan Kaufmann Publishers, Inc, San Francisco, CA, pp. 39–46, 1990.

[4] L. Deng and X. Huang, "Challenges in adopting speech recognition", Communications of the ACM, vol. 47, No.1, pp.60–75, 2004.

[5] D. Jurafsky, J. Martin, Speech and Language Processing, 2nd ed, Prentice Hall, 2008.

[6] A. Rudnicky, A. Hauptmann, and K. Lee, "Survey of Current Speech Technology", Communications of the ACM, vol. 37, No. 3, pp.52-57, 1994.

[7] Microsoft Corporation, Microsoft Bing Search Engine, URL: http://www.bing.com, 2011.

[8] J.R. Lewis, "Effect of Error Correction Strategy on Speech Dictation Throughput", Proceedings of the Human Factors and Ergonomics Society, pp.457-461, 1999.

[9] Tomokiyo, L. M, "Recognizing non-native speech: Characterizing and adapting to non-native usage in speech recognition", Ph.D. thesis, Carnegie Mellon University, 2001.

[10] L.L. Chase, "Error-Responsive Feedback Mechanisms for Speech Recognizers", PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 1997.

[11] D. Gibbon, R. Moore, and R. Winski, "Handbook of Standards and Resources for Spoken Language Systems", Mouton de Gruyter, 1997.

[12] B. Suhm, "Multimodal Interactive Error Recovery for Non-Conversational Speech User Interfaces", PhD thesis, University of Karlsruhe, 1998.

[13] A. R. Setlur, R. A. Sukkar, and J. Jacob, "Correcting recognition errors via discriminative utterance verification", In Proceedings of the International Conference on Spoken Language Processing, pp.602–605, Philadelphia, PA, 1996.

[14] Z. Zhou, H. M. Meng, and W. K. Lo, "A multi-pass error detection and correction framework for Mandarin LVCSR", In Proceedings of the International Conference on Spoken Language Processing, pages 1646–1649, Pittsburgh, PA, 2006.

[15] L. Mangu and M. Padmanabhan, "Error corrective mechanisms for speech recognition", In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 1, pp.29–32, Salt Lake City, UT, 2001.

[16] S. Kaki, E. Sumita, and H. Iida, "A method for correcting errors in speech recognition using the statistical features of character co-occurrence", In COLING-ACL, pp.653–657, Montreal, Quebec, Canada, 1998.

[17] S. Jung, M. Jeong, and G. G. Lee, "Speech recognition error correction using maximum entropy language model", In Proceedings of the International Conference on Spoken Language Processing, pp.2137–2140, Jeju Island, Korea, 2004.

[18] A. Sarma and D. D. Palmer, "Context-based speech recognition error detection and correction", In Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, pp.85–88, Boston, MA, 2004.

[19] E. K. Ringger and J. F. Allen, "Error correction via a post-processor for continuous speech recognition", In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol.1, pp.427–430, Atlanta, GA, 1996.

[20] Eric K. Ringger and James F. Allen, "A Fertility Channel Model for Post-Correction of Continuous Speech Recognition", In Proceedings of the Fourth International Conference on Spoken Language Processing, vol.2, pp. 897-900, 1996.

[21] R. L. Brandow and T. Strzalkowski, "Improving speech recognition through text-based linguistic post-processing", United States Patent 6064957, 2000.

[22] A.A. Markov, "Essai d'une Recherche Statistique Sur le Texte du Roman 'Eugène Oneguine'", Bull. Acad. Imper. Sci. St. Petersburg, vol. 7, 1913.

[23] M. Meyers, CompTIA Network+ All-in-One Exam Guide, 4th ed, McGraw-Hill Osborne Media, 2009.

[24] N. Kim-Chi, La Personnalité et l'épreuve de dessins multiples, PUF edition, 1989.

[25] Microsoft Corporation, Microsoft Speech Application Programming Interface (SAPI 5.0) engine, URL: http://www.microsoft.com/download/en/details.aspx?displaylang=en&id=10121, 2011.

Segmentation of the Breast Region in Digital Mammograms and Detection of Masses

Armen Sahakyan
Digital Signal and Image Processing Lab
Institute for Informatics and Automation Problems of NAS of RA
Yerevan, Armenia

Hakop Sarukhanyan
Digital Signal and Image Processing Lab
Institute for Informatics and Automation Problems of NAS of RA
Yerevan, Armenia

Abstract—**The mammography is the most effective procedure for an early diagnosis of the breast cancer. Finding an accurate and efficient breast region segmentation technique still remains a challenging problem in digital mammography. In this paper we explore an automated technique for mammogram segmentation. The proposed algorithm uses morphological preprocessing algorithm in order to: remove digitization noises and separate background region from the breast profile region for further edge detection and regions segmentation.**

Keywords - mammography; image segmentation; ROI; masses detection; breast region

I. Introduction

Breast cancer stays in the first place among women malignant neoplasia structures list (about 30%). According to Worldwide Health Corporation it is being #1 of the fundamental reasons of the women's average age mortality. The National Cancer Institute estimates that one out of eight women develops breast cancer at some point during her lifetime [1].

The goal of mammography is to provide early detection of breast cancer through low-dose imaging of the breast. Mammography is considered to be the most efficient technique for identifying lesions when they are not palpable and when there are structural breast modifications [2]. It shows to the physician differences in breast tissue densities and these differences are fundamental to a correct diagnosis. At present, there are no effective ways to prevent breast cancer, because its cause remains unknown [3]. Therefore urgency and importance of mammography image processing is obvious.

Computer-Aided Detection (CADe) and Diagnosis (CADi) systems are continuously being developed aiming to help the physicians in early detection of breast cancer. These tools may call the physician's attention to areas in the mammography that may contain radiological findings. In digital mammography, segmentation is the process of partitioning mammograms into regions, aiming to produce an image that is more meaningful and easier to analyze [4]. After being segmented, the mammogram or the mass lesion region can be further used by physicians, helping them to take decisions that involve their patients' health.

This paper is organized as followed. Section II gives some knowledge about image segmentation. Section III describes an image segmentation technique presented in this paper. In Section IV are shown experimental results of described techniques. In the next section the conclusion and future work are given.

II. Mammogram Segmentation

Mammogram segmentation usually involves classifying mammograms into several distinct regions, including the breast border [5], the nipple[6] and the pectoral muscle. The principal feature on a mammogram is the breast border, otherwise known as the skin-air interface, or breast boundary. The breast contour can be obtained by partitioning the mammogram into breast and non-breast regions. The extracted breast contour should adequately model the soft-tissue/air interface and preserve the nipple in profile.

There are a number of problems associated with the accurate segmentation of the breast region. Firstly, owing to the nature of x-ray each pixel in a mammogram represents two or more tissues; indeed all pixels contain a component due to attenuation by skin. Superimposition of different tissue types makes it difficult to differentiate between different regions. As a result of the mammogram acquisition process, there is a region of decreasing contrast near the breast contour where the breast tapers off. This region constitutes the uncompressed region of the breast commonly referred to as the "breast edge", and is caused by a lack of uniform compression of the breast tissue. This tapering effect causes a lack of visibility along the peripheral region of the mammogram, making it difficult to perceive the breast contour and identify the nipple position. The process of digitization may further decrease this visibility through the addition and accentuation of noise.

There have been various approaches proposed to the task of segmenting the breast profile region in mammograms. Some of these have focused on using threshold [7][8], gradients [9], modeling of the non-breast region of a mammogram using a polynomial [10], or active contours [11].

III. Segmentation Algorithm

Digital mammogram images were acquired from the mini-MIAS database [12].

Images acquired consist of left and right breast images of fatty, fatty-glandular and dense-glandular breasts. The acquired mammogram images are classified into three major cases: malignant, benign and normal, all of which are subdivided into five categories as follows: 1) Circumscribed masses 2) Speculated masses 3) Ill-defined masses 4) Architecturally distorted masses 5) Asymmetrical masses. The images are digitized at 200 micron pixel edge and padded in order to obtain all images with a size of 1024 × 1024 pixels.

A. *Digitization Noise Removal*

Digitization noises such as straight lines (see Fig. 1(a)) present in the majority of acquired mammogram images are filtered using a two-dimensional (2D) Median Filtering approach in a 3-by-3 neighborhood connection [13]. Each output pixel contains the median value in the 3-by-3 neighborhood around the corresponding pixel in the input images. The edges of the images however, are replaced by zeros. Fig. 1(a) shows the digitization noise present in a mammogram image and Fig. 1(b) shows the same image after noise removal.

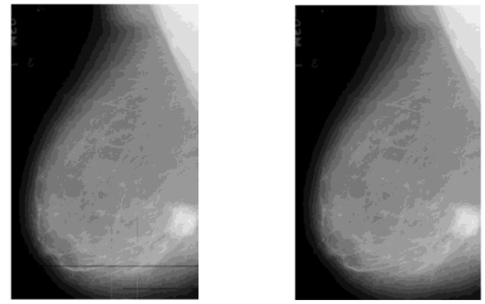
B. *Image Enhancement*

Problems with image acquisition such as scanner-induced artifacts, excessive background noise, scratches and dust artifacts could influence the reliability of this algorithm. The mammogram presented in Fig. 2(a) has a highly non-uniform background and very little contrast in the area above the core breast tissue region. So image enhancement is required before segmentation. As enhancement technique was selected contrast-limited adaptive histogram equalization (CLAHE) [14]. It is a well-known technique of adaptive contrast enhancement. The normal and adaptive histogram equalization may over-enhance the noises and sharp regions in images due to the integration operation. It yields large values in the enhanced image for high peaks in the histogram of the nearly uniform regions in the original image. To solve this problem, the CLAHE uses a clip level to limit the local histogram in order to limit the amount of contrast enhancement for each pixel. This clip level is a maximum value of the local histogram specified by users. An interactive binary search process is used to redistribute the pixels which are beyond the clip level. The CLAHE algorithm has following steps: 1) divide the original image into contextual regions; 2) obtain a local histogram for each pixel; 3) clip this histogram based on the clip level; 4) redistribute the histogram using binary search; 5) obtain the enhanced pixel value by histogram integration. The result of CLAHE technique is shown in Fig. 2(b).

C. *Background Separation*

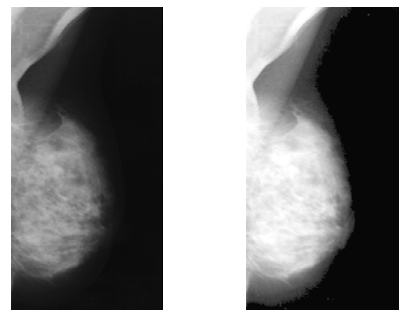
Radiopaque artifacts such as wedges and labels in the mammograms images are removed using threshold technique and morphological operations [15][16].

Fig. 3(a) shows a mammogram image with a radiopaque artifact present. After the grayscale mammogram images are converted into binary, as shown in Fig. 3(b) for the image in Fig. 3(a).



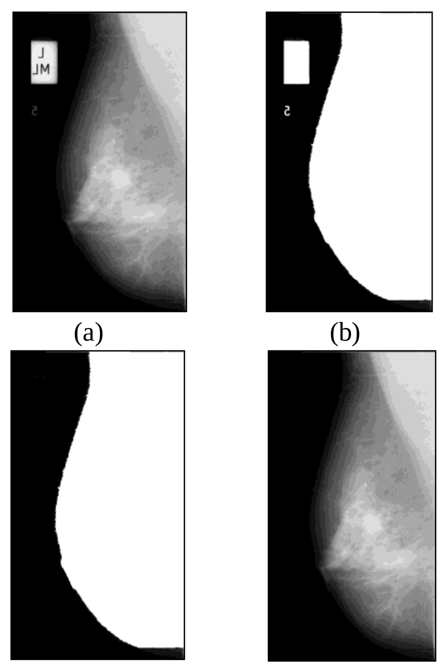
(a) (b)

Figure 1. Mammogram digitization noise removal using 2D median filtering. (a) Original image (b) Filtered image after noise removal



(a) (b)

Figure 2. Mammogram enhancement using CLAHE. (a) Original image (b) Enhanced image



(a) (b) (c) (d)

Figure 3. Suppression of artifacts and labels from a mammogram (a) Original image (b) Threshold image (c) Largest area (object) selected from threshold image 3(b) (d) Mammogram image with radiopaque artifacts suppressed

The algorithm steps to find artifacts and labels and to separate breast profile are as follows:

- Convert grayscale mammogram into binary using threshold technique.
- Binary image objects are labeled and number of pixels in all objects is counted.
- All binary objects are cleaned except the largest one: breast profile (Fig. 3(c)). After which morphological operation to remove isolated pixels is applied.
- After algorithm checks all pixels in a binary image and sets a pixel to 1 if five or more pixels in its 3-by-3 neighborhood are 1's, otherwise, it sets the pixel to 0.
- The resulting binary image is multiplied with the original mammogram image to get the final image without artifacts and etc.

D. Edge Detection

Edge detection is used for getting edge map of the breast region. In the first step horizontal scanning is performed. If any change of pixel intensity is observed it is marked by a black pixel indicating a horizontal edge point. This process is continued for all rows of pixel data to obtain a horizontal edge map. In the next step, image is scanned vertically. Continuing the process for all the columns a vertical edge map image is obtained. Finally, the horizontal edge map is merged with vertical edge map by performing a logical OR operation on the two image files, to obtain the edge map of mammogram image. The algorithm steps for horizontal image map are:

- Scan the image array horizontally from left-most pixel to right-most pixel from first row to last row and take the first pixel intensity value as a reference value.
- Compare intensity of subsequent pixels with the reference value. If the same value go on to next pixel.
- If the value differs, change the value of reference value to the pixel intensity value and mark the pixel black.
- If the last row and column pixel is not reached then go to Step2.

Steps for horizontal image map are the same as implemented for rows.

E. Segmentation of Breast ROI

After previous steps we get the ROI of the Breast. We now consider the edge map that corresponds only to the ROI of the breast. The edge map indicates various closed structures within the breast region that corresponds to the different anatomical regions of the breast. The objective is to identify these regions on the mammogram image. The algorithm starts by identifying the left baseline of the breast image from the edge map. Then a line is drawn vertically from top to bottom identifying the left boundary of the breast. Then the breast boundary is scanned on the right side to locate the rightmost pixel on the breast contour. After the pixel is located another vertical line is drawn from top to bottom passing through the rightmost pixel thus partitioning the image only to the breast ROI.

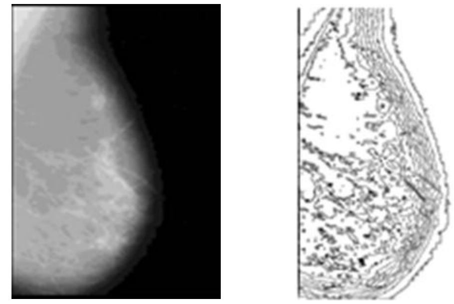


Figure 4. Mammogram after edge detection algorithm

This process optimizes the algorithm and increases the processing efficiency. At this stage is tried to locate all edge paths that are circular or terminate either on the left base line or the bottom of the image, forming a closed structure. Locating all the edge paths that originates from the top margin line namely the first row of the image is started.

The algorithm travels each individual path and stores them on the plotting list. This list is plotted on another image if the edge path is circular or end on the last row of the image or the vertical line representing the left boundary of the breast.

After all pixel paths on the first row are traversed the algorithm repeats similar scanning and traversal of all pixel paths from the row that is indicated by dividing the image vertically into sixteen segments. The first row of each segment is used for locating pixel paths for traversal. All complete paths are then plotted on another image thus providing regions of the breast. The algorithm steps to detect regions are:

- Scan the image from the left side of the image to locate the leftmost pixel of the breast region and draw a vertical line along this pixel from top to bottom representing the left baseline or boundary.
- Scan the image from the right side of the image to locate the rightmost pixel of the breast region and draw a vertical line along this pixel from top to bottom.
- Partition the obtained rectangle horizontally into sixteen segments and start with the first row of the first segment.
- Scan the enclosed rectangle from the right side to left, from the first row of the segment.
- Obtain a pixel that is black indicating an edge path, traverse the pixel path by considering all the surrounding pixels in a clockwise priority and consider the pixel with the highest priority.
- The pixels that surrounded the edge pixel, but are of lower priority are stored in a history stack to be used only if the traversal process reaches a dead end.
- If a dead end is reached, pop out from the history stack a lesser priority pixel and continue with the traversal process.
- Store the pixels traversed in a plotting list for plotting.

- Traversal continues to the next pixel till it reaches the left baseline or the bottom of the image or the start position is reached.
- If the traversal is terminated, the plotting list is erased and continues from Step5. Else plot pixels from the plotting list.
- Continue to Step4 till all black pixels, indicating an edge path, is traversed.
- Move to the first row of the next segment and continue from Step4 to Step9.

IV. EXPERIMENTAL RESULTS

To demonstrate the robustness of the algorithm it has been tested on mammograms with differing breast tissue densities. Overall, for the mammograms evaluated, the mean values for the quality measure for breast region detections were 0.95, signifying that the algorithm seems extremely robust with respect to density types. In Fig. 5 is shown experimental results of detection of breast contours of mammogram. In Fig. 6 is shown results of edge detection and regions segmentation of mammogram.

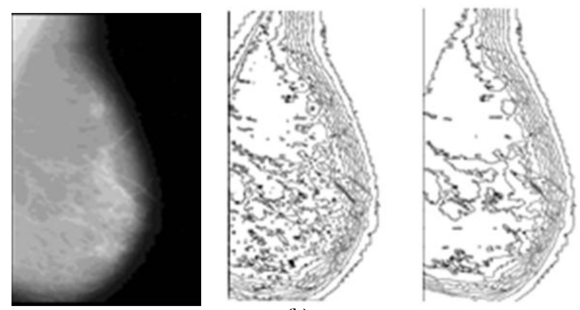
V. CONCLUSION AND FUTURE WORKS

There are a number of factors which make it difficult to postulate the exact effect digital mammograms on a particular segmentation algorithm. The first of these relates to acquisition parameters, such as exposure time and energy level, which influence the quality of the image registered on film. Secondly, segmentation of the breast region from the background is further hampered by the tapering nature of the breast.

As we can see from the results, mammography segmentation using technique presented in this paper is efficient. Edge detection and regions segmentation algorithms work more efficient when images are preprocessed. For the future work it is planned improvement of the algorithm to derive a smoother breast region contour for image preprocessing. Improvement of edge detection and region segmentation algorithms and abnormalities detection (mass, tumors or calcifications) in segmented images is planned.

(a) (b) (c)

Figure 5. Detection of breast contour of the mammogram (a) Original, (b) Segmented mask, (c) Breast contour superimposed on (a).



(a) (b) (c)

Figure 6. Segmentation of the breast region: (a) original mammogram, (b) mammogram after edge detection, (c) mammogram after regions segmentation.

REFERENCES

[1] L.-M. Wun, R. M. Merrill, and E. J. Feuer, "Estimating Lifetime and Age-Conditional Probabilities of Developing Cancer," Lifetime Data Analysis, vol. 4, pp. 169-186, 1998.

[2] Swanson, G. M. (1992). Breast cancer in the 1990's. Journal of American Medical Women's Association, v. 47, p. 140-148.

[3] D. E. Stewart, et al., "Attributions of cause and recurrence in long-term breast cancer survivors," Psycho-Oncology, vol. 10, pp. 179-183, 2001.

[4] Shapiro, Linda G. & Stockman, George C. (2002). "Computer Vision". Prentice Hall. ISBN 0-13-030796-3

[5] R. Chandrasekhar, and Y. Attikiouzel, "Automatic Breast Border Segmentation by Background Modelling and Subtraction", in 5th International Workshop on Digital Mammography (IWDM), (Yaffe M. ed.), Medical Physics Publishing, Madison, USA, pp. 560–565.

[6] R. Chandrasekhar, and Y. Attikiouzel, "A Simple Method for Automatically Locating the Nipple on Mammograms", IEEE Transactions on Medical Imaging, vol. 16, pp. 483–494, Oct. 1997.

[7] U. Bick, M.L. Giger, R.A. Schmidt, R.M. Nishikawa, D.E. Wolverton, and K. Doi, "Automated Segmentation of Digitized Mammograms", Academic Radiology, vol. 2, no. 2, pp. 1–9, 1995.

[8] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3 ed.: Pearson Prentice Hall, 2007.

[9] A.J. Mendez, P.J. Tahoces, M.J. Lado, M. Souto, J.L. Correa, and J.J. Vidal, J.J, "Automatic Detection of Breast Border and Nipple in Digital Mammograms", Computer Methods and Programs in Biomedicine, vol. 49, pp. 253–262, 1996.

[10] R. Chandrasekhar, and Y. Attikiouzel, Y. Automatic, "Breast Border Segmentation by Background Modeling and Subtraction", in: M.J. Yaffe (Ed.), Proceedings of the 5th International Workshop on Digital Mammography (IWDM), Medical Physics Publishing, Toronto, Canada, 2000, pp. 560–565.

[11] M. A. Wirth, A. Stapinski, "Segmentation of the breast region in mammograms using active contours", in Visual Communications and Image Processing, pp.1995–2006.

[12] J. Suckling et al., "The Mammographic Image Analysis Society digital mammogram database", Exerpta Medica., vol. 1069, pp. 375– 378, 1994.

[13] T. Huang, G. Yang, G. Tang, " A fast two-dimensional median filtering algorithm", Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 27, 1979

[14] S. M. Pizer, et al., "Adaptive histogram equalization and its variations," Computer Vision, Graphics, and Image Processing, vol. 39, pp. 355-368, 1987.

[15] A. Sahakyan, "Segmentation of Mammography Images Enhanced by Histogram Equalization", Mathematical Problems of Computer Science 35, Armenia, pp. 109 - 115, 2011.

[16] A. Sahakyan, H. Sarukhanyan, "Automatic Segmentation of the Breast Region in Digital Mammograms", Computer Science and Information Technologies, Proceedings of the Conference, pp. 386 - 389, Yerevan, Armenia, September 26-30, 2011.

Wavelet de-noising technique applied to the PLL of a GPS receiver embedded in an observation satellite

Dib Djamel Eddine
Laboratory of Telecommunications and Digital Signal Processing Department of Electronics
Djillali Liabes University
Sidi Bel Abbes, Algeria.

Djebbouri Mohamed
Laboratory of Telecommunications and Digital Signal Processing Department of Electronics
Djillali Liabes University
Sidi Bel Abbes, Algeria.

Taleb Ahmed Abddelmalik
Laboratory of automatics, Mechanical and Computer Science
Valenciennes University.
France.

***Abstract*—In this paper, we study the Doppler effect on a GPS(Global Positioning System) on board of an observation satellite that receives information on a carrier wave L1 frequency 1575.42 MHz .We simulated GPS signal acquisition. This allowed us to see the behavior of this type of receiver in AWGN channel (AWGN) and we define a method to reduce the Doppler Effect in the tracking loop which is wavelet de-noising technique.**

Keywords-GPS; Doppler frequency; PLL; wavelet packet de-noising;

I. Introduction

A GPS (Global Positioning System) is a geo-localization system operates globally .It includes three segments: Space, Control, user.. GPS signals are transmitted on one frequency, called L1, which contains the code acquisition called «coarse" (C / A), and the various navigation messages (L1 = 1575.42 MHz). Firstly we define the satellite of observation, it belongs to the LEO satellites. Low earth orbit is defined as an orbit within a locus extending from the earth's surface up to an altitude of 2000 Km [1].

Attributing to their high speeds, data transmitted through LEO is handed off from one satellite to another as satellites generally move in and out of the range of earth-bound transmitting stations. Due to low orbits, transmitting stations are not as powerful as those that transmit to satellites orbiting at greater distances from earth's surface.

As LEO orbits are not geostationary, networks of satellites are required to provide continuous coverage. In our case the average of altitude of satellite of observation is 700km. The period of an observation satellite is 101 minutes [2], than we can compute the velocity of the satellite of observation:

Vob=Rob×$d\theta/dt$ = (700+6368) km=7330 m/s (1)

The period of a GPS satellite is 11h, 58min, 2.05s [3], as above we compute the velocity of the satellite GPS:

Vgps=Rgps×$d\theta/dt$ =26560km=3874 m/s (2)

II. Doppler Effect

The change in frequency observed when there is a relative movement between the source and the observer is called the Doppler Effect. It can be given by:

$$f_d = \beta f_e \quad (3)$$

Where:

f_d : Doppler frequency

f_e : Transmission frequency, and

$$\beta = Vd/C \quad (4)$$

Vd is the velocity which causes the Doppler and C is the celerity.

We have to give extreme cases of the shift Doppler in an observation satellite and it is given by:

$$\overrightarrow{vs} = \overrightarrow{Vob} + \overrightarrow{V\ GPS}. \quad (5)$$

$$vs = \sqrt{Vob^2 + V\ GPS^2 - 2VobV\ GPS\ \cos\theta}. \quad (6)$$

(6) Is Generalized Pythagorean relationship where θ is the angle between two vectors of velocities of the two satellites (GPS and observation).

v_s is the relative velocity between the satellite GPS and the satellite of observation. The maximum speed between the satellites is when the two satellites are in opposition and the minimum speed is when the two satellites are in the same direction.

Because:

if θ= 0(cos θ >0) ,Vs achieves its minimum value (the two satellites are in the same direction)

And if $\theta = \pi$ ($\cos\theta < 0$) Vs awaits its maximum value (the two vectors are in opposite directions).

Therefore

Vs min=7330-3874=3456m/s.
Vs max=7330+3874=11204m/s.
Now we define:

$$Vd = \frac{vsrecos\theta}{\sqrt{re^2+rs^2-2erssin\theta}} \quad [4] \qquad (7)$$

The velocity Vd which causes the Doppler with:

Vs: relative velocity between the two satellites
re: the radius of rotation of the satellite observation
rs: the radius of rotation of the GPS satellite
θ: Is shown in Figure 1:

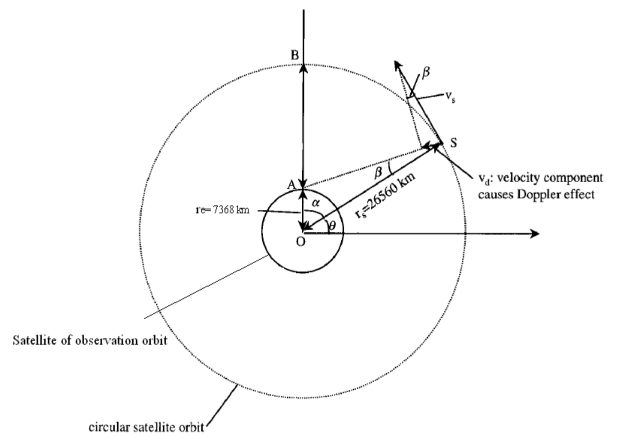


Figure 1. Doppler frequency caused by the velocities of the two satellites.

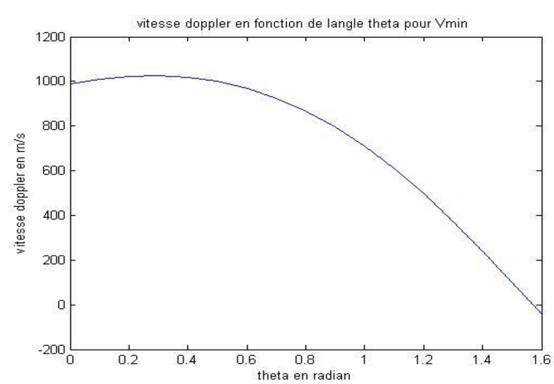


Figure 2. Doppler velocity vs. angle theta at Vs min.

We draw the equation of Vd with the two value of Vs.

After computing we have
Vd min=1025m/s
Vd max=3135m/s
Which give two shift Doppler

In the first case **Fd=5.4Khz.**

In the second case **Fd=16.5Khz.**

In the first case we observe that it likes a shift Doppler in a GPS in the earth and its graph of acquisition is as shown in figure 4.

In the second case the process of acquisition have a lot of noise because the large bandwidth applied on the PLL filter.

In this section we define firstly the principle of a PLL; then

we give an idea on the wavelet de-noising technique and how we can apply this technique in a PLL.

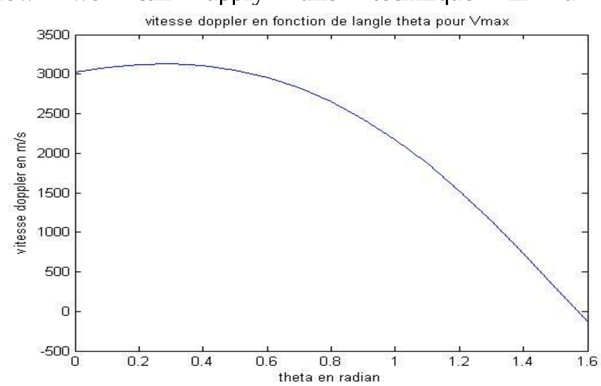


Figure 3. Doppler velocity vs. angle theta at Vs max.

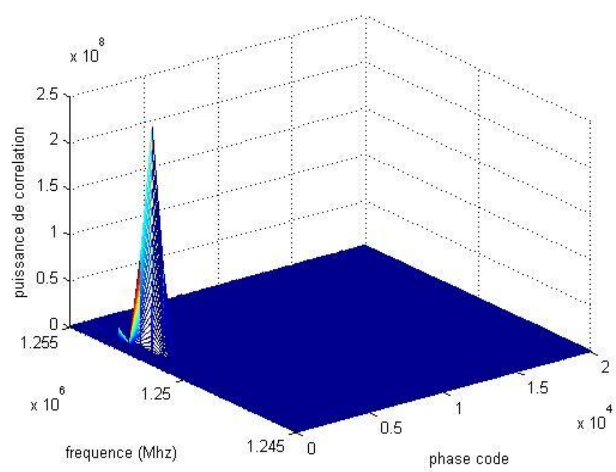


Figure 4. Acquisition of GPS signal at Fd=5.4 kHz.

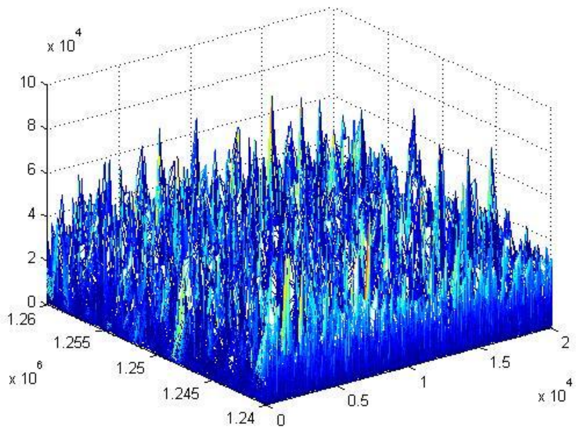


Figure 5. Acquisition of GPS signal at Fd=16.5 MHz.

III. Basic Principle Of PLL

A PLL is a control loop which synchronizes its output signal $u_0(t)$ (generated by a voltage controlled oscillator) with a reference or input signal $u_i(t)$ in frequency as well as in phase. The PLL generates a control signal which is related to the phase error to control VCO to generate the signal frequency which will be closer to the input signal ,until the output frequency of a PLL is exactly same as the input signal and the error signal $u_d(t)$ between the oscillator's output signal and the reference signal is zero ,or remains constant .In such a control mechanism ,the PLL always adjusts the phase of the output signal to lock to the phase of the reference signal [5].A typical PLL block diagram is shown in figure 6 .It consists of three basic function components :a discriminator or a phase detector (PD),a loop filter (LF) and a voltage controlled oscillator (VCO).LF is used to filter the result from the PD and generate the control signal $u_f(t)$ to control the VCO to generate the output signal.

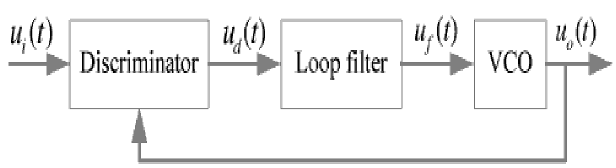


Figure 6. A typical PLL block diagram

IV. Principle Of Wavelet Packet De-Noising

Signal f(t) can be decomposed into coefficients $\{\alpha_\lambda(t)\}$ which are on another space based on the wavelet packet basis .The signal f(t) can be expressed by the linear superposition of these coefficients:

$$f(t) = \sum_{\lambda \epsilon A} a_\lambda \beta_\lambda (t) \tag{8}$$

Where $\{\beta_\lambda(t)\}$are the basis of the other space .We can abstract the characteristic of the signal from the coefficients a_λ.So the processing of the signal can be replaced by the processing of a_λ [6].In this paper ,the wavelet packet decomposition was used.

The wavelet packets transform the signal in time domain into the coefficients in the inner product space of wavelet packet. Define the following notation: $\varphi(x)$ is an scaling function and $\phi(x)$ is the corresponding wavelet function, ,$\{V_k\}$ is a multi-resolution space ,also called scale space generated by $\varphi(x)$. $\{W_k\}$ is a wavelet space generated by$\phi(x)$, W_{k-1} is an fill space differences between V_{k-1} and V_k ,so Lévesque space $L^2(R)$ can be decomposed as [7] :

$$L^2(R) = \cdots \oplus W_{-2} \oplus W_{-1} \oplus W_0 \oplus W_1 \oplus W_2 \oplus \ldots$$

And $\{2^{\frac{k}{2}}\phi(2^k t - l): l\epsilon Z\}$ is a group of bases of W_k.wavelet packet transform can be carried out by followed ways:

Define a sequence of functions as follows [8] :

$$\begin{cases} \varphi_{2n}(x) = 2^{j/2} \sum_k h_k\, \varphi_n(2^{j/2} - k) \\ \varphi_{2n+1}(x) = 2^{j/2} \sum_k g_k\, \varphi_n(2^{j/2} - k) \end{cases} \tag{9}$$

Where j,kεZ (Z is an integer set) and j is a scale parameter is time or location parameterεN (N is a non –negative integer set),h_k is a low –pass filter coefficient,$g_k$is a high –pass filter coefficient .moreover $\{h_k\}$ and $\{g_k\}$ are a group of conjugate mirror filters .their relationships are as followed :

$$\sum_{n\epsilon Z} h_{n-2k} h_{n-2l} = \delta_{k,l} \sum_{n\epsilon Z} h_n = \sqrt{2}\, g_k = (-1)^k h_{l-k} \tag{10}$$

Then two –scale equations of wavelet packet transform can be achieved:

$$\begin{cases} \varphi_{2n}(x) = \sqrt{2} \sum_k h_k \varphi_n(2x-k) \\ \phi_{2n}(x) = \sqrt{2} \sum_k g_k \varphi_n\,(2x-k) \end{cases} \tag{11}$$

The signal can be expressed by the following wavelet packet bases function:

$$f(t) = \sum_{n,j} C_n^j\,(k) 2^{\frac{k}{2}} \phi_n \left(2^{\frac{k}{2}} t - j\right). \tag{12}$$

Where:

$$C_n^j(k) = 2^{\frac{k}{2}} \int_R \ f(t)\, \phi_n \left(2^{\frac{k}{2}} t - j\right) dt. \tag{13}$$

Based on the theory of local maxima value of the wavelet transformation, the characteristic information is concentrated in a few coefficients .So, de-noising process can be done by saving the character coefficients and threshold other coefficients.

After threshold, the modified coefficients can be used to reconstruction of signal .The reconstruction formulation of a discrete signal is done by [8]:

$$C_m^j(k) = \sum_n \bar{h}_{k-2n} C_{2m}^{j+1}(n) + \sum_n \bar{g}_{k-2n} C_{2m+1}^{j+1}(n) \tag{14}$$

$\bar{h}_{k-2n}$ and $\bar{g}_{k-2n}$ can be obtained by reversing order of h_{k-2n} and g_{k-2n}.

Wavelet packet decomposition can decompose the signal to different frequency bands in different levels .If we have sufficient decomposed levels and data samples the beginning and the ending of a frequency band can be acquired [6].

In general wavelet packet de-noising is done in the following steps :

Firstly, decompose the signal based on the selected wavelet packet basis .The signal is decomposed into several layers of wavelet packet coefficients as a tree [6].

The structure of the tree is as figure 7:

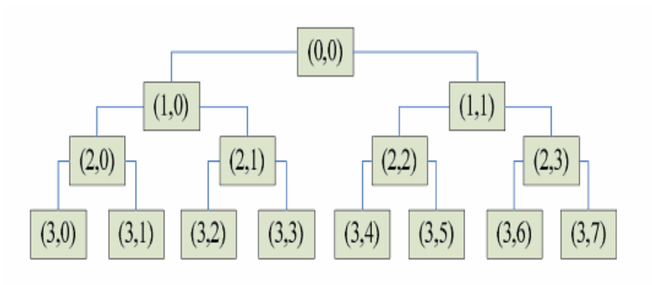


Figure 7.

Secondly, compute the best tree based on the entropy of Shannon, which is computing the best wavelet packet basis.

Thirdly, a threshold is computed and then a soft-threshold is applied to the coefficients. The threshold can be calculated as follows [8]:

$$\lambda = \sigma\sqrt{2 * log_2 N} \tag{15}$$

There are two kinds of threshold functions, hard threshold and soft threshold functions. In this paper, soft threshold function was selected .It is defined as [8]:

$$\widehat{W}_{j,k} \begin{cases} sign(W_{j,k})(|W_{j,k}| - \lambda), & |W_{j,k}| \geq \lambda \\ 0 & , |W_{j,k}| < \lambda \end{cases} \tag{16}$$

Fourthly, signal reconstruction is performed using the modified detail coefficients.

APPLYING WAVELET PACKET DE-NOISING FOR A PLL

The purpose of applying the wavelet packet de-noising technique in the PLL is to reduce the noise level before the loop filter .the block diagram is as figure 8.

The main function of the wavelet packet de-noising is to reduce the noise level within the bandwidth of the loop filter .the loop filter could filter out the noise beyond the bandwidth of the loop filter ,but the noise within the bandwidth still could pass through and affect the NCO tracking performance.

It should be noted that the wavelet packet de-noising technique can only reduce the noise level rather than eliminate the noise totally.

The remaining noise still can affect the NCO tracking performance but at a lower level[9].

The wavelet packet de-noising technique may cut off some useful signals which are smaller than the threshold .This disadvantage will make the PLL spend more time to lock due to the loss of some control signals and output from PLL distorted. However the decrease of the noise will produce smaller phase error that will help the PLL to maintain locked and smooth the output [6].

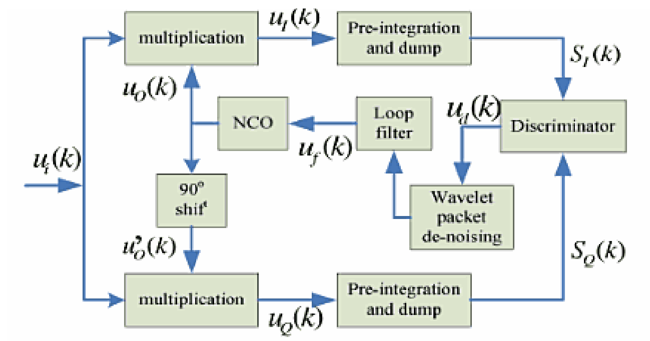


Figure 8.

V. Test And Results

In our case we have 16.5 KHZ such as a Doppler frequency and we will see the behavior of the loop of tracking with the both configuration

1-without applying of wavelet packet de-noising

2-with applying of wavelet packet de-noising

Data_nav=1 1 -1 1 1 -1 -1 1 1 1 -1 1 1 -1 1

A. *without applying of wavelet packet de-noising:*

We will see the signal at the input of loop filter (this signal will command the NCO) and also the bits after tracking.

We observe that because the lots of noise in the signal at the input of loop filter we have bad results (weak signal) so we cannot extract the bits.

B. *with applying of wavelet packet de-noising:*

We apply the wavelet packet de-noising just before the loop filter and we have these results:

This method de-noise the signal to have good performance especially when we have a lot of noise because the higher Doppler frequency and we observe that the bits tracked are similar to the original signal GPS.

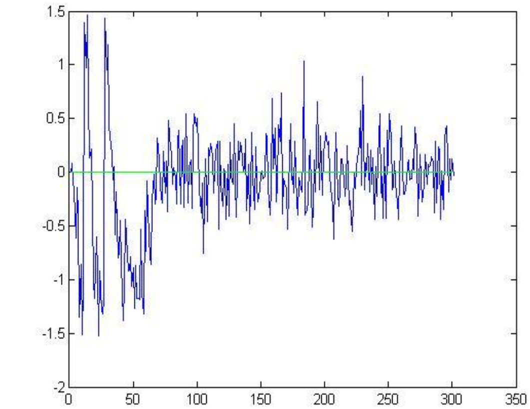


Figure 9. The signal at the input of loop filter without applying wavelet packet de-noising

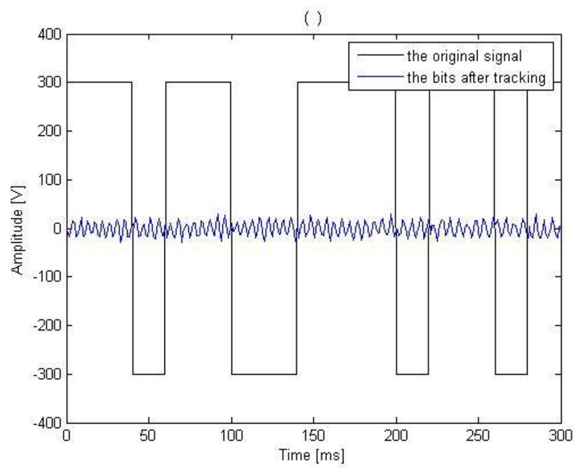


Figure 10. The bits after tracking without applying wavelet packet de-noising.

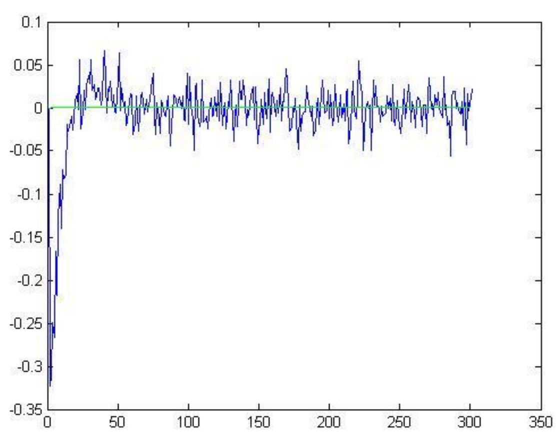


Figure 11. The signal at the input of loop filter with applying wavelet packet de-noising

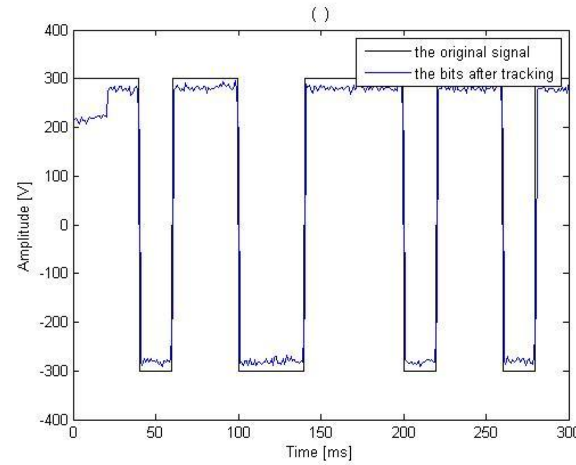


Figure 12. The bits after tracking with applying wavelet packet de-noising

CONCLUSION

Applying the wavelet packet de-noising technique in the PLL helps to reduce the noise within the bandwidth of the loop filter and therefore, a less noisy tracking performance can be obtained from the NCO, we observe that by using this method we have less error in the bits in the end the process of tracking .than the ordinary method that do not use the wavelet packet de-noising.

This method is recommended in the GPS embedded in the satellite of observation because the high effect of Doppler frequency.

REFERENCES

[1] IEEE journal on selected areas in communications .vol.13 .no.2 February 1995

[2] www.astronoo.com/satelliteobservation

[3] Fundamentals of Global Positioning System Receivers: A Software Approach James Bao-Yen Tsui Copyright © 2000 John Wiley & Sons, Inc. Print ISBN 0-471-38154-3 Electronic ISBN 0-471-20054-9 page 37-49

[4] Gardner, F.M: Phase lock techniques, 3rd demijohn Wiley and sons, Inc, USA (2005)

[5] Qian H-m,Ma,J-c,Li,Z-y :fiber optical gyro de-noising based on wavelet packet soft –threshold algorithm .journal of chinese Inertial technology 15(5),602-605(2007)

[6] daubechies ,I :orthonormal bases of compactly supported wavelets .commun.pure appm.41,909-996(1998)

[7] peng,Y,weng ,X.H :thresholding –based wavelet packet methods for doppler ultrasound signal denoising in :IFMBE proceedings,APCMBE 2008 ,vol 19,pp.408-413(2008)

[8] Lian,P,Lachapelle G,Ma,C.L :Improving tracking performance of PLL in high dynamics applications ION.NTM,1042-1058(2005).

An Improved Squaring Circuit for Binary Numbers

Kabiraj Sethi
Department of Electronics and Telecommunication Engineering,
VSS University of Technology, Burla
Formerly UCE, Burla-768018. (India)

Rutuparna Panda
Department of Electronics and Telecommunication Engineering,
VSS University of Technology, Burla
Formerly UCE, Burla-768018. (India)

***Abstract*—In this paper, a high speed squaring circuit for binary numbers is proposed. High speed Vedic multiplier is used for design of the proposed squaring circuit. The key to our success is that only one Vedic multiplier is used instead of four multipliers reported in the literature. In addition, one squaring circuit is used twice. Our proposed Squaring Circuit seems to have better performance in terms of speed.**

Keywords-Vedic mathematics; VLSI; binary multiplication; hardware design; VHDL.

I. Introduction

Multiplication and squaring are most common and important arithmetic operations having wide applications in different areas of engineering and technology. The performance of any circuit is evaluated mainly by estimating the silicon area and speed (delay). Hence, continuous efforts are being made to achieve the same. In order to calculate the square of a binary number, fast multipliers such as Braun Array, Baugh-Wooley methods of two's compliment, Booth's algorithm using recorded multiplier and Wallace trees are in use. Recursive decomposition and Booth's algorithm are the most successful algorithms used for multiplication. Other methods include Vedic multipliers based on 'Urdhva Tiryagbhyam' and the "Duplex" properties of 'Urdhva Tiryagbhyam'. Therefore, the main motivation behind this work is to investigate the VLSI Design and Implementation of Squaring Circuit architecture with reduced delay. Interestingly, only one multiplier is used here instead of four multipliers reported in the literature. Here, one squaring circuit is used twice to reduce delay.

Vedic mathematics was reconstructed from Vedas by Sri Bharati Krisna Tirthaji (1884-1960) after his eight years of research on Vedas [1-3]. According to him, Vedic mathematics is mainly focused on sixteen very important principles or word-formulae, which are otherwise known as Sutras. Note that the most important feature of the Vedic mathematics is its coherence. The entire system is wisely interrelated and unified. The general multiplication scheme can easily be reversed to achieve one-line divisions. Similarly, the simple Squaring Scheme can easily be reversed to produce one-line Square Roots. These methods are very easy to understand. This paper discusses a possible application of Vedic mathematics to design multipliers and squaring circuits. General idea for design of digital multipliers is described in [4, 5].

The idea of using Vedic mathematics for design of multipliers has been discussed in [6-13]. In [14], 'Urdhva Tiryagbhyam' Sutra is shown to be an efficient multiplication algorithm as compared to the conventional counterparts. Authors of [15] have also shown the effectiveness of this Sutra to reduce N×N multiplier structure into an efficient 4×4 multiplier structure. However, they have mentioned that 4×4 multiplier section can be implemented using any efficient multiplication algorithm. In [16], authors have presented an array multiplier architecture using Vedic Sutra. More or less the coding is done in VHDL and synthesis is done in Xilinx ISE series [17,18].

Recently, a squaring circuit has been reported in the literature [19]. This may be noted that designing Vedic multipliers using array multiplier structures as discussed in above references provide us less delay and, thus, they are treated as high speed multipliers as compared to Booth's algorithm using recorded multipliers and Wallace trees. However, we can reduce delay further using carry save adders (CSA). The idea of using CSA for design of digital multipliers is explained in [3,4]. This has motivated us to design Vedic multipliers based on CSA [20]. One more crucial issue with the earlier proposed methods is that they use four numbers of such Vedic multipliers to evaluate squaring of a n-bit binary number. Here, we have more focus on the issue and tried to use only one Vedic multiplier instead of four for evaluating square of a n-bit binary number.

We apply Vedic Sutras to binary multipliers using carry save adders. In particular, we develop an efficient binary multiplier architecture that performs partial product generations and additions in parallel. With proper modification of the Vedic multiplier algorithm, the squaring circuit is developed. Here, the computation time involved is less. The combinational delay and the device utilizations obtained after synthesis is compared. Our proposed Vedic multiplier based Squaring Circuit seems to have better performance in terms of speed. The hardware architecture of the squaring circuit is presented.

II. The Multiplier Architecture

Booth's multipliers [5] are normally used for squaring of binary numbers. The modified Booth multiplier considered uses four components. Booth Encoder, Partial Product Generator, Wallace Tree and Binary Adders are used for Booth multiplier architecture. Booth multiplier uses two main ideas to increase the speed of the multiplication process. First attempt is to reduce the number of partial products.

Then the second attempt is to increase the speed at which the partial products are added. The partial products are reduced using Booth encoder. Time for partial product additions is reduced using Wallace Tree. Further, Binary adder is used for addition of final sum vector and carry vector.

In this section, we propose an efficient multiplier architecture using Vedic mathematics. The 'Urdhva Tiryagbhyam' (Vertically and Crosswise) sutra [2] has been traditionally used for the multiplication of two numbers in the decimal number system. In this paper, we apply this Sutra to the binary number system. The motivation behind the extension to binary number system is to make it compatible with the digital hardware circuits. This Sutra is illustrated with the help of a numerical example, where two decimal numbers are multiplied.

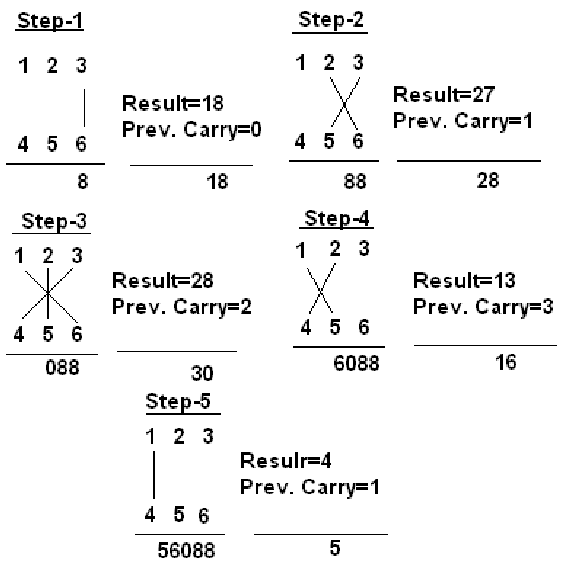


Figure 1. Line diagram for multiplication.

For more clarity, line diagram for the multiplication of two decimal numbers (123×456) is displayed in Fig. 1. Note that the digits on two ends of the line are multiplied and then results are added with the previous carry, as shown in the figure. When we find more lines in one step as shown in steps 2 to 4, all results are added to the previous carry. Interestingly enough, the least significant digit of the number, thus, obtained acts as one of the result digits. The rest digit acts as the carry for the next step. Note that here the initial carry is zero.

We now extend the above idea (for multiplication) to binary number system. Computer arithmetic [3,4] usually deals with binary number systems. Thus, there is a strong need to develop efficient schemes for multiplication of binary numbers. This may be noted that multiplication of two bits A0 and B0 is nothing but an AND operation.

Interestingly, this operation can easily be implemented using a two input AND gate used in digital circuits. Such types of multipliers using digital circuits similar to array multipliers are discussed in [21,22]. In order to illustrate this coveted multiplication scheme in binary number system, here we consider the multiplication of two binary numbers (4 bits) A3A2A1A0 and B3B2B1B0. As the result of this multiplication would be more than 4 bits, we express it as ……R3R2R1R0. As an illustration, line diagram for multiplication of two 4-bit binary numbers is shown in Fig. 2. The same idea can be extended to higher bits. It is noteworthy to mention here that Fig.2 is simply the mapping of Fig.1 in binary system. For the sake of simplicity, each bit is represented by a cross enclosed by a circle. Least Significant Bit (LSB) R0 is obtained by multiplying the LSBs of the multiplicand and the multiplier. Here, the multiplication process is carried out according to steps displayed in Fig. 2. Further, digits on both sides of the line are multiplied and added with the carry from the previous step. All seven steps shown in Fig.2 are important. This generates one of the bits of the result (Rn) and a carry (Cn). This carry is added in the next step and, hence, the process goes on. If more than one line are there in one step, all results are added to the previous carry. In each step, least significant bit acts as the result bit and other bits act as carry. To be more specific, if in some intermediate step, the sum is '110', then '0' acts as the result bit and '11' as the carry (which is denoted as Cn in this paper). It is noteworthy to mention here that Cn may be a multi-bit number. Thus, we get the following expressions:

$$R_0 = A_0B_0 \quad (1)$$
$$C_1R_1 = A_1B_0 + A_0B_1 \quad (2)$$
$$C_2R_2 = C_1 + A_2B_0 + A_1B_1 + A_0B_2 \quad (3)$$
$$C_3R_3 = C_2 + A_3B_0 + A_2B_1 + A_1B_2 + A_0B_3 \quad (4)$$
$$C_4R_4 = C_3 + A_3B_1 + A_2B_2 + A_1B_3 \quad (5)$$
$$C_5R_5 = C_4 + A_3B_2 + A_2B_3 \quad (6)$$
$$C_6R_6 = C_5 + A_3B_3 \quad (7)$$

with C6R6R5R4R3R2R1R0 being the final product. Partial products are calculated in parallel and, hence, the delay involved is just the time it takes for the signal to propagate through the gates [8].

The multiplier architecture is explained below. Both 2X2 Vedic multiplier module and 4X4 Vedic multiplier architecture are displayed below. The motivation is to reduce delay. Multiplier design ideas are well explained in [3,4]. Here, 'Urdhva Tiryagbhyam' (Vertically and Crosswise) sutra [2] is used to propose such an architecture for the multiplication of two binary numbers.

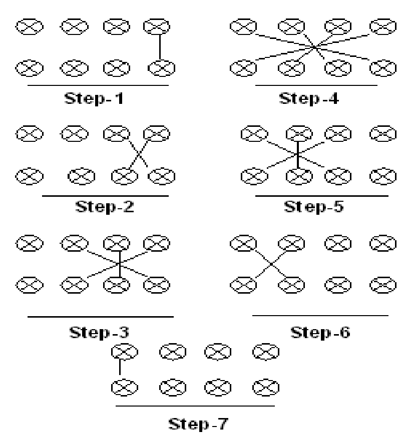


Figure 2. Line Diagram for multiplication of two 4-bit binary numbers.

A. 2x2 Vedic Multiplier Module For Binary Numbers

Here, an efficient Vedic multiplier using carry save adder is presented. The 2X2 Vedic multiplier module is implemented using two half-adder modules and is displayed in Fig. 3. Very precisely we can state that the total delay is only 2-half adder delays, after final bit products are generated.

It is wise to write Implementation Equations of 2X2 Vedic multiplier module for simulation. The implementation equations are written as:

R_0 (1-bit) = $A_0.B_0$ (8)

R_1 (1-bit) = $A_1.B_0 + A_0.B_1$ (9)

R_2 (2-bits) = $A_1.B_1 + R_1$ (1) (10)

Product = R_2 & R_1 & R_0 (11)

where '&' denotes concatenation operation. Note that final result (product) is obtained using Eq.(11).

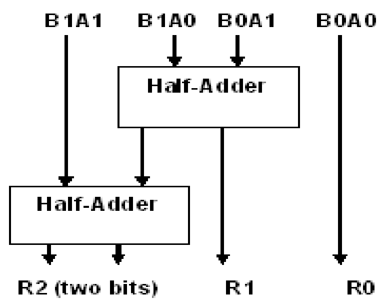


Figure 3. Architecture of 2X2 multiplier.

B. 4x4 Vedic Multiplier Module

The 4X4 Vedic multiplier architecture is displayed in Fig.4. This is implemented using four 2X2 Vedic multiplier modules as discussed in Fig. 3. The beauty of Vedic multiplier is that here partial product generation and additions are done concurrently. Hence, it is well adapted to parallel processing. The feature makes it more attractive for binary multiplications. This, in turn, reduces delay.

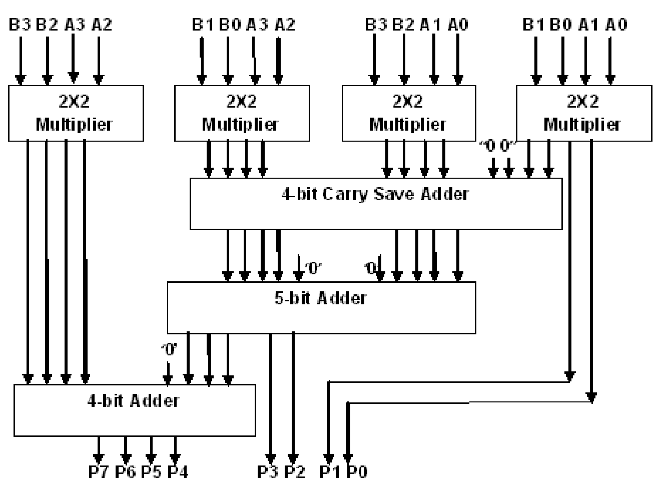


Figure 4. Architecture of 4X4 multiplier

In this section, we describe the architecture of 4X4 multiplier using Vedic method discussed above (Eqns.8-11). To get final product (P7P6P5P4P3P2P1P0), one 4-bit carry save adder, one 5-bit binary adder and one 4-bit binary adder are used. In this proposal, the 4-bit carry save adder (CSA) is used to add three 4-bit operands, i.e. concatenated 4-bit ("00" & most significant two output bits of right hand most of 2X2 multiplier module as shown in Fig.3) and two 4-bit operands we get from the output of two middle 2X2 multiplier modules. It may be noted that the outputs of the CSA (sum and carry) are fed into a 5-bit binary adder to generate 5-bit sum, as desired. Many more interesting ideas can be revoked here.

It may be reiterated the fact that the middle part (P3P2) denotes the least significant two bits of 5-bit sum obtained from the 5-bit binary adder. Finally, as shown in Fig.4, the 4-bit output of the left most 2X2 multiplier module and concatenated 4-bits ('0' & the most significant three bits of 5-bit sum) are fed into a 4-bit binary adder. In this architecture, the P7P6P5P4 express the sum.

The proposed Vedic multiplier can be used to reduce delay. Early literature speaks about Vedic multipliers based on array multiplier structures. On the other hand, we proposed a new architecture, which is efficient in terms of speed. The arrangements of CSA and binary adders shown help us to reduce dely. Interestingly, 8X8 and 16X16 Vedic multiplier modules are implemented easily by using four 4X4 and four 8X8 multiplier modules, respectively. Further, the proposed 4X4 Vedic multiplier can also be used for squaring of a 4-bit binary number.

C. 2-Bit Squaring Circuit

The 2X2 Vedic multiplier architecture is modified as shown in Fig. 5 to realise the 2-bit squaring circuit. Here, one half adder and one AND gate are utilized instead of two half-adders as shown in Fig. 3.

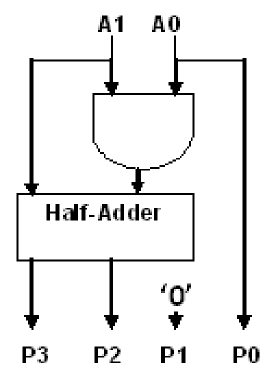


Figure 5. Block Diagram of 2-bit Squaring Circuit.

This is the basic module. Note that a 4-bit squaring circuit is implemented using two 2-bit squaring circuits (as shown in Fig.5) and one 2X2 Vedic Multiplier (as displayed in Fig.3) instead of four 2X2 Vedic Multiplier modules used in Fig.4.

In the same manner, 8-bit squaring circuit and 16-bit squaring circuits are implemented using Vedic multiplier module and squaring circuits of 4-bit and 8-bit, respectively. Likewise, n-bit squaring circuit can be implemented taking one (n/2)-bit Vedic multiplier module and two (n/2)-bit squaring circuits.

D. n-Bit Squaring Circuit

Taking the architectural concept of 4X4 Vedic multiplier module, general block diagram of the newly proposed n-bit squaring circuit is shown in Fig. 6.

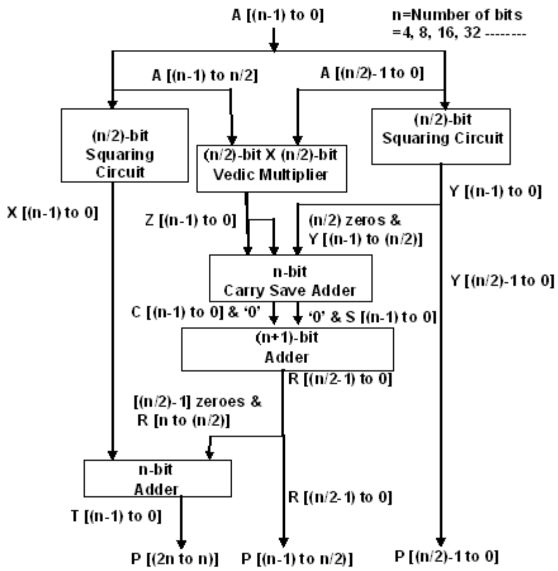


Figure 6. Architecture of n-bit Squaring Circuit.

Let us describe the proposed architecture of n-bit Squaring Circuit in a tabular form. Table-I explains the idea.

TABLE I. N-BIT SQUARING CIRCUIT DESCRIPTION

Bit size	Description	Numbers used
2 bit	AND Gate XOR Gate	2 1
4 bit	2X2 Vedic Multiplier 2 bit squaring circuit CSA Binary Adder	1 2 1 2
8 bit	4X4 Vedic Multiplier 4 bit squaring circuit CSA Binary Adder	1 2 1 2
16 bit	8X8 Vedic Multiplier 8 bit squaring circuit CSA Binary Adder	1 2 1 2
32 bit	16X16 Vedic Multiplier 16 bit squaring circuit CSA Binary Adder	1 2 1 2
64 bit	32X32 Vedic Multiplier 32 bit squaring circuit CSA Binary Adder	1 2 1 2

III. VERIFICATION AND IMPLEMENTATION

In this work, 4-bit, 8-bit, 16-bit, 32-bit and 64-bit squaring circuits are implemented in VHDL [17]. Logic synthesis and simulation are done in Xilinx - Project Navigator and Modelsim simulator [18].

We compare our synthesis results with the method recently presented by Prabha et al [19]. The results are displayed in Table 2 and Table 3 for squaring circuits of different bit size. These Tables show the difference in combinational delays and the device utilization. Squaring circuits of different bit size are considered for simulation. Comparison of combinational delay in nano seconds (ns) is displayed in Table-II.

TABLE II. COMPARISON OF COMBINATIONAL DELAY (NS)

Device: Vertex4vlx 15sf363-12	Modified Booth Multiplier [19]	Prabha et al [19]	Ours
4 bit	8.154	4.993	4.993
8 bit	15.718	14.256	12.781
16 bit	36.657	33.391	15.994
32 bit	74.432	68.125	18.272
64 bit	141.982	129.867	22.905

TABLE III.
COMPARISON OF DEVICE UTILISATION (4 INPUT LUTS)

Device: Vertex4vlx 15sf363-12	Modified Booth Multiplier [19]	Prabha et al [19]	Ours
4 bit	32	6	6
8 bit	186	35	64
16 bit	880	294	366
32 bit	2760	1034	1267
64 bit	6854	4535	5361

Comparison of device utilization (4 input LUTs) is shown in Table-III. The performance of all squaring circuits are evaluated on the same device Vertex4vlx15sf363 with a speed grade of -12.The results suggest that the proposed architecture is faster than "Modified Booth Multiplier" and "the method recently presented by Prabha et al" [19]. Here, we see significant speed improvement though there is a small increase in area. Simulation results obtained are shown in figures 7-9 for verification. Simulation results for 32-bit and 64-bit are not presented here to avoid consuming space. However, they are also verified and found correct.

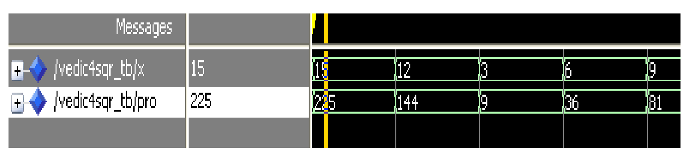


Figure 7. 4-bit Squaring Circuit.

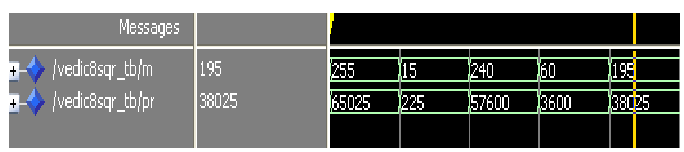


Figure 8. 8-bit Squaring Circuit.

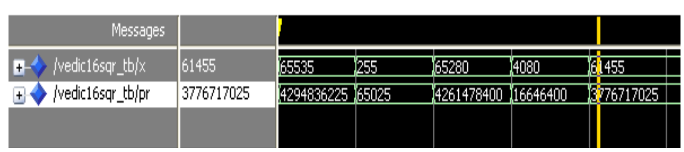


Figure 9. 16-bit Squaring Circuit.

It is worthy to mention here that the results displayed in Table-II are quite expected. The delay is about 2.5 times when we go from 4-bit to 8-bit, in our case. However, the increase in combinational delay is less for higher bits, which is due to inherent parallelism.

To be very precise, the implementation equations (Eqs.8-11) are well adapted to parallel processing. For 8-bit squaring, we are using two 4-bit squaring circuits and one 4X4 Vedic multiplier followed by one CSA and two binary adders as shown in Table-I. Hence, increase in delay is more when we move from 4-bit to 8-bit. But we could exploit the benefit of parallelism while implementing squaring circuits of higher bit size, i.e. 16-bit, 32-bit and 64-bit as displayed in Table-II.

Thus, our method outperforms other methods in terms of speed. The proposed squaring circuit may be useful for the design of hardware for computer arithmetic.

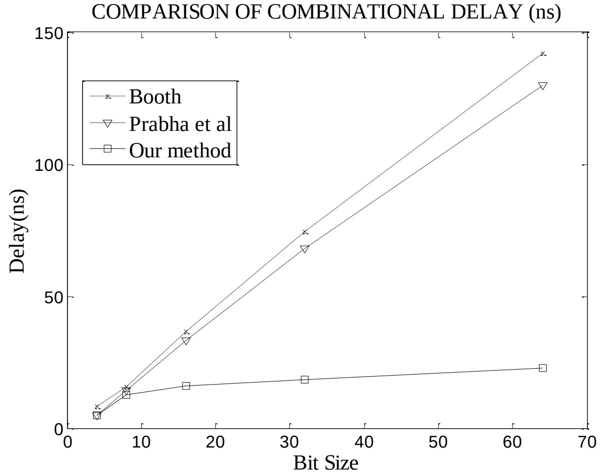


Figure 10. Comparison of Combinational Delay (ns).

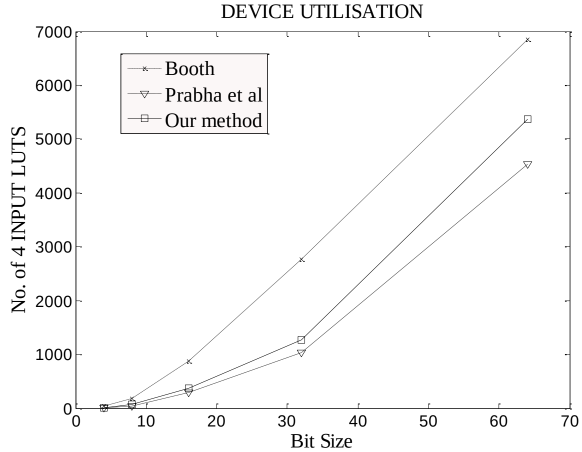


Figure 11. Comparison of Device Utilisation (4 Input LUTs).

To make things explicitly clear, delays for different bit size are displayed in Fig.10. It is observed that the delay is significantly less in our method, particularly for bit size more than or equal to 16. Therefor there is a significant speed improvement in our case.

The reason may be due to the fact that only one multiplier is used instead of four multipliers reported in [19]. However, an engineering tradeoff is observed between Fig.10 and Fig.11. Device utilisation curves are displayed in Fig.11 for a comparison. Space requirement is slightly more in our case as compared to the scheme proposed by Prabha et al [19]. The reason may be due to the fact that we use two squaring circuits of size (n/2)-bits and one CSA.

IV. Conclusion

The performance of the proposed squaring circuit using Vedic Mathematics proved to be efficient in terms of speed. Due to its regular and parallel structure, it can be realised easily on silicon as well. Squaring of binary numbers of bit size other than powers of 2 can also be realized easily. For example, squaring of a 24-bit binary number can be found by using 32-bit squaring circuit with 8 MSBs (of inputs) as zero. The idea proposed here may set path for future research in this direction. Future scope of research is to reduce area requirements.

References

[1] Maharaja, J.S.S.B.K.T., “Vedic mathematics,” Motilal Banarsidass Publishers Pvt. Ltd, Delhi, 2009.

[2] Swami Bharati Krisna Tirtha, “Vedic Mathematics,” Motilal Banarsidass Publishers, Delhi, 1965.

[3] Vedic Mathematics [Online]. Available: http://www.hinduism.co.za/ vedic.htm.

[4] B. Parhami, “Computer Arithmetic Algorithms and Hardware Architectures,” 2nd ed, Oxford University Press, New York, 2010.

[5] Kai Hwang, “Computer Arithmetic: Principles, Architecture and Design,” New York: John Wiley & Sons, 1979.

[6] S. Akhter,, “VHDL implementation of Fast NxN Multiplier Based on Vedic Mathematics,” Proc. of IEEE Conference, pp.472-475, 2007.

[7] P.D. Chidgupkar, and M.T. Karad, “The Implementation of Vedic Algorithms in Digital Signal Processing,” Global Journal of Engng. Educ., vol.8 , pp.153-158, 2004.

[8] H.S. Dhillon, and A. Mitra, “A Reduced-Bit Multiplication Algorithm for Digital Arithmetic,” International Journal of Computational and Mathematical Sciences, pp.64-69, 2008,

[9] P. Mehta, and D. Gawali, “Conventional versus Vedic Mathematical Method for Hardware Implementation of a Multiplier,” Proc. Int Conf. on Advances in Computing, Control, and Telecommunication Technologies, Trivandrum, Kerala, India, pp.640-642, 2009.

[10] P. Nair, D. Paranji, and S.S. Rathod, “VLSI Implementation of Matrix-Diagonal Method of Binary Multiplication,” Proc. of SPIT-IEEE Colloquium and Int Conf., Mumbai, India, pp.55-58, 2008.

[11] M. Ramalatha, K.D. Dayalan, P. Dharani, ,and S.D. Priya, “High Speed Energy Efficient ALU Design using Vedic Multiplication Technique,” Lebanon , pp. 600-603, 2009.

[12] H. Thapliyal, and M.B. Srinivas, “VLSI Implementation of RSA Encryption System Using Ancient Indian Vedic Mathematics,” Proc. of SPIE VLSI Circuits and Systems II, pp.888-892, 2005.

[13] H.D. Tiwari, G. Gankhuyag, C.M. Kim, and Y.B. Cho, “Multiplier Design Based on Ancient Indian Vedic Mathematics”, Proc. Int SoC Design Conf., pp.65-68. 2008.

[14] P.D. Chidgupkar, and M.T. Karad, “The Implementation of Vedic Algorithms in Digital Signal Processing,” Global J. of Engg. Edu., vol.8, no. 2, 2004.

[15] H. Thapliyal and M. B. Srinivas, “High Speed Efficient N×N Bit Parallel Hierarchical Overlay Multiplier Architecture Based on Ancient Indian Vedic Mathematics,” Enformatika Trans., vol. 2, pp. 225–228, Dec. 2004.

[16] R. Pushpangadan, V. Sukumaran, R.Innocent, D. Sasikumar, and V. Sundar, "High Speed Vedic Multiplier for Digital Signal Processors," IETE Journal of Research, vol.55, pp.282-286, 2009.

[17] V.A. Pedroni, "Circuit Design with VHDL," 2008.

[18] 'Xilinx ISE User Manual', Xilinx Inc, USA, 2007

[19] Prabha S., Kasliwal, B.P. Patil and D.K. Gautam, "Performance Evaluation of Squaring Operation by Vedic Mathematics", IETE Journal of Research, vol.57, Issue 1, Jan-Feb 2011.

[20] Devika, K. Sethi and R.Panda, "Vedic Mathematics Based Multiply Accumulate Unit," 2011 International Conference on Computational Intelligence and Communication Systems, CICN 2011, pp.754-757, Nov. 2011.

[21] http://www.scribd.com/doc/29045484/A digital multiplier architecture using Urdva Tiryagbhyam sutra of Vedic mathematics

[22] http;//webspace.utexas.edu/hd3496/www/Downloads/ DhiMit_IC3_2007.pdf

AUTHORS PROFILE

Mr. Kabiraj Sethi is presently a Senior Faculty in the Department of Electronics and Telecommunication Engineering, VSS University of Technology Burla. His area of research interests includes – VLSI and Digital signal processing.

Dr. Rutuparna Panda was born in 1963. He received B.Sc Engg. and M. Sc. Engg. degrees from UCE Burla in 1985 and 1988, respectively. He obtained Ph.D.(Engg) degree from IIT,Kharagpur in 1998. He is currently a Professor in the Department of Electronics and Telecommunication Engineering, VSS University of Technology Burla. He has guided 24 M.Tech Theses and 4 Ph.D. Theses. He has over 70 papers in International/National Journals and conferences. His area of research interests includes – Bioinformatics, Biomedical Image Fusion, Digital signal/image processing, VLSI Signal Processing.

Forks impacts and motivations in free and open source projects

R. Viseur

Teaching Assistant, Department of Economics and Innovation Management,
Faculty of Engineering, University of Mons, 20, Place du Parc, 7000 Mons, Belgium.

Senior Research Engineer, Centre of Excellence in Information and Communication Technologies,
29/3, Rue des Frères Wright,
6041 Charleroi, Belgium.

***Abstract*— Forking is a mechanism of splitting in a community and is typically found in the free and open source software field. As a failure of cooperation in a context of open innovation, forking is a practical and informative subject of study. In-depth researches concerning the fork phenomenon are uncommon. We therefore conducted a detailed study of 26 forks from popular free and open source projects. We created fact sheets, highlighting the impact and motivations to fork. We particularly point to the fact that the desire for greater technical differentiation and problems of project governance are major sources of conflict.**

Keywords- open source; free software; community, co-creation, fork.

I. Introduction

Bar and Fogel define forks as situations occurring when developers "make a separate copy of the code and start distributing their own divergent version of the program" [2]. Free and open source software has four freedoms: the freedom to run, to study, to redistribute copies and modify the software (gnu.org). The free and open source software licenses guarantee the four freedoms, which involve the provision of source code [20]. Forks are usually observed in the field of free software. Forking is indeed a right that stems from the four freedoms associated with the software.

Mateos Garcias and Steinmueller distinguish mechanisms of forking and hijacking. The hijacking occurs when individuals "depose the project leader who has resisted the revision, leaving this original leader with no followers" [18]. In this paper, we will use "fork" for "forking" or "hijacking".

The title of Rick Moen's essay, "Fear of Forking", is characteristic of the fear of forks among entrepreneurs [19]. When he announced the LibreOffice fork (from OpenOffice.Org), Bruce Guptill, consultant for the analyst firm Saugatuck (www.saugatech.com), estimated for example that "the nature of open source leads to fragmentation, itself leads to uncertainty". As a failure of cooperation, forks are an interesting research topic.

The paper is organized as follows.

We will explore the concept of forks. We will then study a set of forks that occurred within popular free and open source software projects, and identify their motivations and impacts. Finally we will discuss the results, and propose ways to better prevent forks.

II. Background

A. Perception of fork

If the fear of forks is visible with companies, Gosain also points to the sensitivity of the open source community beside the forks and the fragmentation of projects [10].

Bar and Fogel estimate that forks are often the result of a management mismatch [2]. They recommend forking only if necessary and if able to do better job. If the motivation for forking is the slowness of patches release, they recommend producing patches instead. Fogel notes, however, the scarcity of forks and a preference for trying to reach an agreement [8].

Eric Raymond estimates that forking "spawns competing projects that cannot later exchange code, splitting the potential developer community" [29]. He also distinguishes the case of "pseudo-forks", i.e. distinct projects that share a large common code base (this is for instance the case of GNU/Linux distributions). Weber considers that specialization may, in some cases, be managed through a system of patches, so as to avoid fragmentation of the project [39].

B. Forks and governance

For Hemetsberger and Reinhardt, management of online collaboration is less a question of coordinating tasks than overcoming conflicts arising from the contradictions between collective strategy and individual actions [13]. The voluntary nature of contributions often prevents the enforcement of duties or decisions (principle of consensus). Dahlander and Magnusson also consider that capture of network externalities requires specific skills (it has a cost) and that gains associated with the opening decrease when the number of players increases [5]. They highlight the difficulty in aligning business and community strategies. Bowles and Gintis distinguish the operating logic of a community, and the ones of companies and states [4]. The tensions that may result do not necessarily cause a fork. However the example of Netscape illustrates the difficulty of finding a tradeoff between a company and a community [36].

Implementation of common rules and effective governance structures should limit the tensions and especially their consequences. Eric Raymond distinguishes several structures for the management of free and open source projects [28]. First, a single developer can work on the project and take all decisions alone. He is expected to pass the torch in case of failure to maintain the project.

Second, multiple developers can work under the direction of a "benevolent dictator". This structure is found in the Linux kernel (Linus Torvalds) or in Emacs (Richard Stallman). The potential for conflict is higher. Authority comes from responsibility and some developers become in practice responsible for one or more parts of the software. Another principle complements this rule: seniority prevails. The title of benevolent dictator may be passed on to another developer, as in the Perl project. Third, the decisions can be made by a panel of voters. This is for example the case for the Apache project.

The 2000s have seen the increasing involvement of businesses in the development of free and open source softwares, by initiating projects, freeing existing projects or collaborating with well-established communities [34, 38]. The increasing size of projects and cooperation between sometimes competing businesses (coopetition) also contributed to the creation of more complex and formal governance structures.

C. Forks and licenses

In practice, project license modulates the interest in whether to fork, even if no free and open source license cancels the risk [32]. Two major types of free licenses exist: permissive licenses (also named academic or unrestrictive licenses) and copyleft licenses (also names reciprocal or restrictive licenses) [1, 16, 20, 35]. A permissive license allows the user to apply a different license, possibly a proprietary license, to derivative works (thus also to forks).

A copyleft license "links the rights to the obligation to redistribute the software and its changes only under the same license as that by which the licensee has obtained those rights" [20]. In case of copyleft licensed software, exchanging source code is still possible between the original software and its forks. In case of a permissive free software license, the license can change and forbid the exchange of source code. In particular, the exchange will be impossible if the new software is published under a proprietary license, and one-way if it is published under a copyleft license (due to the fact that copyleft imposes conservation of the original license) [20]. St. Laurent considers other legal provisions limiting forkabily (or, if not, the consequences), such as brand protection in the Apache license [32]. Incompatibilities between licenses, sometimes due to apparently innocuous terms in legal texts, also reduce the opportunities for exchange and combination of source code [9, 32, 35]. Yamamoto, Matsushita, Kamiya and Inoue show, through a study of source code similarities applied to BSD (BSD-Lite, FreeBSD and NetBSD), a progressive divergence of the source code, despite the license compatibility and the similarity of features [41]. St. Laurent also considers this divergence as inevitable with time [32].

Finally, a copyleft license would also limit the financial incentives to fork as it is not possible to create a proprietary branch from the original development [40].

Elie considers unstable (and subject to a higher risk of fork) projects characterized by the coexistence of free release of the software and a second version published under a proprietary license (dual licensing, delayed publication,...) [7]. Elie names "hybrid model" this principle of "discrimination between users". Dahlander and Magnusson estimate on the contrary that the detention of copyright (and other controls) hampers forks initiatives (and allows the return to a proprietary development in case of insufficient network externalities) [5]. The technical complexity of the software would also reduce the risk of fork [33].

Note that the hybrid model suggested by Elie is distinct of the hybrid model described by Muselli [7, 22, 23, 35]. The later indicates a strategy of openness, promoting greater distribution while allowing to retain control over the project. This approach is supposed to facilitate the capture of value by the company and nullify the risk of fork. Muselli gives Sun Microsystems SCSL license as an example.

D. Forks impacts

Wheeler shades the presumed dangerousness of the fork and associates it with a system of healthy competition [40]. He compares it to the principle of a censure motion in parliament or to a strike. The fork would allow the developers community to attract the leaders' attention on the requests that are not taken into account. Some authors even see an "invisible hand" that guarantees the projects sustainability and continuity [26]. The ability to fork would also keep "the communities vibrant, and the companies honest" [21]. Elie sees the fork as "a fundamental right" but also insists on the risk of being cut from the wealth of the core [7]. He often sees in forks the consequence of "ill-defined control systems". Merit in free software communities would come from charisma and ability to live in the conflict rather than technical competences.

Wheeler recognizes that too many forks can cause a weakening of a projects family in the long term [40]. Spinellis and Szyperski see it as a waste of efforts and a source of confusion for the community [31]. Wheeler also distinguishes the forks as variants of software created with a goal of experimentation. A "winning mutation" can finally be accepted as constituting the best approach to a problem. Wheeler sees four possible outcomes to a fork:

- The fork does not convince and disappears.
- The original project and the fork evolve and gradually diverge.
- The original project and the fork merge after a period of cohabitation.
- The original project disappears.

III. RELATED WORKS

Nyman and Mikkonen, in a study of 566 projects hosted on Sourceforge.net and presented by their maintainers as forks, identify motivations classifiable into four categories: technical motivations (adding features, specialization, porting, improving), license changes, local adaptations (language or regional differences) and revival of abandoned projects [25]. Open source company Smile also mentions disagreements about technology directions and licensing, but adds disagreement on trade policy as possible cause of fork [30].

Many forks benefit from more or less extensive studies (or are briefly discussed) in the literature. It includes the family of BSD operating systems [39, 41], KHTML [11], Roxen [5], GCC [8], CVS [2], NCSA HTTPd [34, 38] or SPIP [7]. These results will be used in this study.

IV. METHODOLOGY

We have studied 26 forks of popular free and open source projects. Popular projects have been found more likely to provide usable observations. We relied on existing documents: books, scientific articles, press releases, news on portals about open source and computer science, or projects pages. We have not considered forks leading to the creation of proprietary software, like Kerberos [32].

For each fork we gather relevant information in dedicated forms (fact sheets). They describe the chronology of each fork, its actors and their motivations. The results were summarized in a table, including the initial project name, fork name, fork motivation(s) and its impact on the original project. The impact was evaluated according to the possible outcomes identified by Wheeler [40].

The influence of the license type and the degree of openness of the project management structure were also observed. We assigned a score for openness on a scale from 1 to 4:

- the project is under a free and open source license but centrally managed,
- the project is managed by a team and the rules are informal,
- the decision-making procedures are planned, but favor core team,
- the procedures are documented, decisions and appointments are subject to the votes of active community members.

Note that the management structure may be difficult to precisely determine when the fork is old and/or a project has been completely abandoned.

V. RESULTS

Six motivations to fork have been identified: death of the original project (19%), technical motivations –e.g. new specialization, divergent technical views, different technical objectives,...– (42%), license change (15%), conflict over brand ownership (12%), problems of project governance (38%), cultural differences (8%) and searches for new innovation directions (4%).

In practice, the case studies show that the successful forks (which are likely to be harmful to the original publisher, if there is one) usually start for an important reason.

Stopping the support of popular free and open source software often leads to a fork (see NCSA HTTPd, 386BSD, Red Hat Linux or Roxen). The open source fork succeeds but usually can coexist with a closed version of the product (see Red Hat Linux or Sourceforge).

A fork can occur after the emergence of technical differences. The BSD systems have thus often adopted different technical specializations such as portability or security [31]. This is the most common cause (42%).

Project governance is a source of conflicts for nearly half of the studied cases (38%). The problem is usually a lack of openness of development teams: slowness for taking external contributions into account (see OpenOffice.org), discussion of project objectives (see Sodipodi), maintainer's reluctance to switch to a community development process (see OpenOffice.org, Dokeos, PHP Nuke),... This is therefore the second most common cause of fork.

Brand ownership also appears as a source of conflict (see Claroline, Mambo and OpenOffice.org). It may be related to the issue of governance as the trademark allows the software editor to keep a check on the progress of the project. The brand then crystallizes the tensions between an editor and a community once their objectives diverge.

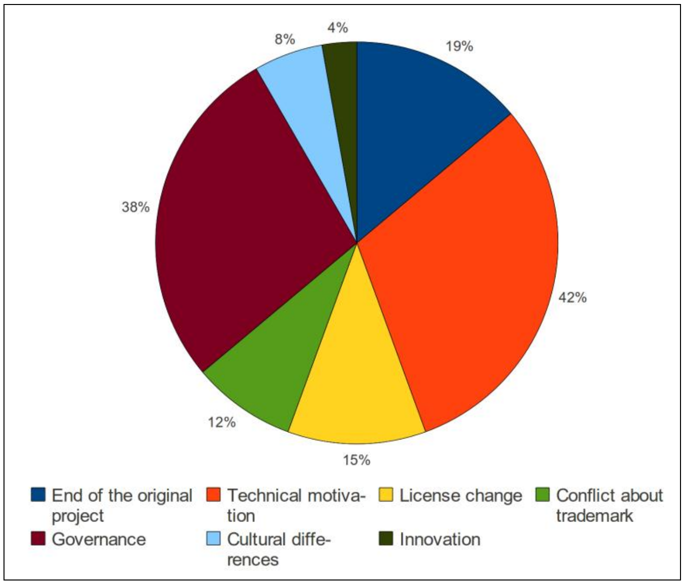


Figure 1. Motivations to fork.

Licensing problems sometimes cause a fork. It may not affect the type of license (see Xfree86) but rather increase (see Ext JS) or reduce (stop the free branch) the software freedom. Licensing software under the GPL or AGPL can facilitate exchanges between projects, since the original license can hardly be changed. The license change is not a dominant motivation to fork (15%).

Forks that have been raised by Theo de Raadt, leader of OpenBSD, can be justified, at least in part, by political or ideological positions. This configuration seems quite marginal in the free and open source landscape. Culture shocks between community and company (see KHTML) or community and administration (see Spip) appear as a possible cause (8%) and illustrate the difficulty in aligning business and community strategies.

The case studies show that the majority of forks do not cause the extinction of the original project (81%). Exception made of the Apache server, X.Org, Joomla or Inkscape, cohabitation appears in more than half of the cases studied (54%). In some cases, the exchange of source codes exists (see FreeBSD, NetBSD, OpenBSD). Subsequent projects fusion (see GCC and EGCC) is possible. The progressive divergence may hamper the merger (see Webkit and KHTML). The complete failure of a fork occurs in less than one case out of five (19%).

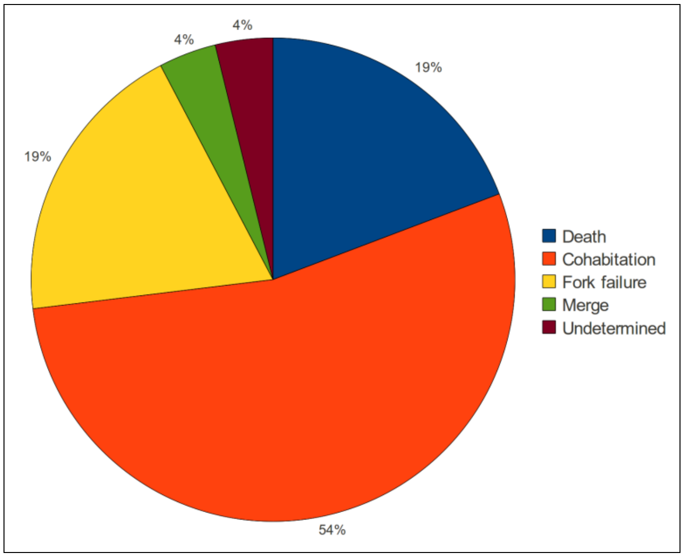


Figure 2. Forks impacts on original projects.

Finally, we find that nearly eight out of ten forks adopt a governance structure characterized by comparable or greater openness than in the original project. The formal rules of processes can give a biased impression of openness, that complaints made against the source code contribution mechanisms may moderate. The OpenOffice.org project (before entering the incubator of the Apache Foundation) is an example.

VI. Discussion

Compared to the study of Nyman and Mikkonen, our research groups several motivations under the label of “technical motivations” and highlights three additional causes: governance issues, difficulties associated to culture differences (already mentioned in state of the art) and conflicts over the ownership of a brand [25]. The changes in technical guidance also occupy a prominent place in our study, although proportionally less. The recovery of stopped projects is most frequent. These differences may be explained by the wider spectrum of motivations considered in our study but also by the different nature of considered projects. Nyman and Mikkonen are based on a set of projects taken on Sourceforge.net, which hosts many small projects, whereas our study was based on popular and mature projects. These have already an active community that plays a role in regulating and empowering the actors.

Many beliefs are refuted by our study. First, the use of copyleft licenses does not reduce the risk of forks. More than six out of ten studied forks were indeed published under a copyleft license (about 75% of free and open source projects are released under a copyleft license [16]). Second, hybrid models do not seem particularly subject to forks (except Chamilo). Third, the fear of a fork driven by competition (and perceived as an act of predation) seems exaggerated: only the case of OpenBravo could possibly be taken as such.

Privatization of popular free and open source software often results in a free software fork. However, the transition from a more permissive free license to a less permissive free license may also lead to a fork. The license change, regardless of its meaning, very often raised tensions in the community. The license choice must be well thought out from the beginning.

The risk of fork due to technical divergences is high. However, it may be limited by adopting a suitable architecture from the beginning. MacCormak, Rusnak and Baldwin recommend a modular architecture [17]. They point to the need for an “architecture for participation” to ease the comprehensibility of the code and the contribution. Mozilla project is an good example. The code left by Netscape was made more modular, and that contributed to attract patches from community [6].

The “kernel-extension model” is an example of modular architecture. It allows the improvement of the software without impacting its core. The editor then guarantees the performance of a core incorporating common features. Integrators and advanced users improve the functionality by developing extensions [3]. This approach can also reduce conflicts with the development team because the integrators need only understand the software interfaces for extensions development. Understanding the specifics of the kernel is not needed. Conflicts may occur on the other hand between community extensions and proprietary extensions sold by the editor.

Promotion of such an architecture underpins the creation of application programming interfaces (APIs), and reminds of the “user toolkits for innovation” described by Von Hippel [37]. These toolkits permit a form of outsourcing to users for innovation tasks requiring deep understanding of customers' needs. The expected benefit is a better satisfaction of customers and, in a free software project, a lower risk of tensions around the project orientations.

Samba illustrates the “killer of innovation” side due to the quality requirement when a large user base exists. This example highlights the value of incubators, such as Apache incubator, allowing experimentation next to the main project. In a way Samba TNG plays an incubator role. A similar effect can be achieved by creating experimental branches in the repository (see Linux).

VII. Conclusion

The goal of this proposal is to shed some light on the motivations and impact of the fork mechanism in free and open source software projects. This paper identified the main motivations to fork, that are technical divergences and governance mismatches. Other causes were highlighted: end of the original project, license change, conflict about trademark and strong cultural differences.

We discussed some ways to manage tensions and prevent project splitting, for example by improving software modularity.

VIII. FUTURE WORKS

The governance issues generally relate to a lack of communication with the community.

However, it seems difficult to conclude definitely on the choice of a specific governance model. Indeed, some projects governance structures appear to be open (cf. FreeBSD, KHTML, OpenOffice.org,...) but are also subject to forks. Moreover some successful projects are build on main developers' strong authority. Thus Mozilla community enforces code ownership (e.g.: module owner) despite the risk of disputes in the community [15, 24].

A more detailed study of these structures, and in particular their interactions with developers, should therefore be considered. The analyze of messages exchanged between developers before, during and after forks would maybe allow to identify specific reasons for the schisms. Data could be extracted (for qualitative or quantitative researches) from public collaborative tools such as mailing lists and bugtrackers (e.g.: [6, 14, 15, 27]).

REFERENCES

[1] T.A. Alspaugh, H.U. Asuncion, and W. Scacchi, “Intellectual Property Rights Requirements for Heterogeneously-Licensed Systems”, 17th IEEE International Requirements Engineering Conference (RE '09), September 2009.

[2] M. Bar and K. Fogel, Open Source Development with CVS, Paraglyph Press, 2003.

[3] P. Bertrand, “Les extensions, arme fatale des solutions Open Source”, Journal du Net, 28 juin 2010.

[4] S. Bowles and H. Gintis, “Social Capital and Community Governance”, Economic Journal, Royal Economic Society, vol. 112 (483), November 2002, pp. 419-436.

[5] L. Dahlander and M. Magnusson, “How do firms make use of open source communities?”, Long Range Planning, vol. 41, n°6, December 2008, pp. 629-649.

[6] J.-M. Dalle and M.L. den Besten, “Voting for Bugs in Firefox”, FLOSS Workshop 2010, July 1, 2010.

[7] F. Elie, Économie du logiciel libre, Eyrolles, 2006.

[8] K. Fogel, How To Run A Successful Free Software Project - Producing Open Source Software, CreateSpace, 2004.

[9] D.M. German and A.E. Hassan, “License integration patterns: Addressing license mismatches in component-based development”, IEEE 31st International Conference on Software Engineering, ICSE 2009, May 2009, p. 188-198.

[10] S. Gosain, “Looking through a window on Open Source culture : lessons for community infrastructure design”, Systèmes d'Information et Management, 2003, 8:1.

[11] A. Grosskurth and M.W. Godfrey, “Architecture and evolution of the modern Web browser”, Preprint submitted to Elsevier Science, June 20, 2006.

[12] B. Guptill, “OpenOffice Changes Highlight Fragmentary Nature and Future of Open Source”, Saugatuck Technology, 2010.

[13] A. Hemetsberger and C. Reinhardt, “Collective Development in Open-Source Communities: An Activity Theoretical Perspective on Successful Online Collaboration”, Organization studies, vol. 30 n°9, September 2009, pp. 987-1008.

[14] J. Howison, K. Inoue, and K. Crowston, “Social dynamics of free and open source team communications”, IFIP International Federation for Information Processing, vol. 203, 2006, pp. 319-330.

[15] S. Krishnamurthy, “About Closed-door Free/Libre/Open Source (FLOSS) Projects: Lessons from the Mozilla Firefox Developer Recruitment Approach”, Libre software as a field of study, Upgrade, vol. VI, issue n°3, June 2005.

[16] J. Lerner and J. Tirole, “The Scope of Open Source Licensing”, Journal of Law, Economics, and Organization, vol. 21, issue 1, 2005, pp. 20-56

[17] A. MacCormack, J. Rusnak, and C.Y. Baldwin, “Exploring the Structure of Complex Software Designs: An Empirical Study of Open Source and Proprietary Code”, Management Science, vol. 52 (7), 2006, pp. 1015-1030.

[18] J. Mateos Garcia and W.E. Steinmueller, “Applying the Open Source Development Model to Knowledge Work”, INK Open Source reasearch working paper n°2, January 2003.

[19] R. Moen, “Fear of Forking - How the GPL Keeps Linux Unified and Strong”, Linuxcare, November 17, 1999.

[20] E. Montero, Y. Cool, F. de Patoul, D. De Roy, H. Haouideg, and P. Laurent, Les logiciels libres face au droit, Cahier du CRID, n°25, Bruylant, 2005.

[21] G. Moody, “Who Owns Commercial Open Source – and Can Forks Work?”, Linux Journal, April 2, 2009.

[22] L. Muselli, “Les licences informatiques : un outil de modulation du régime d’appropriabilité dans les stratégies d’ouverture. Une interprétation de la licence SCSL de Sun Microsystems”., 12ème Conférence de l'Association Information et Management, Lausanne, juin 2007.

[23] L. Muselli, “Le rôle des licences dans les modèles économiques des éditeurs de logiciels open source”, Revue française de gestion, n° 181, 2008, pp. 199-214.

[24] M. Nurolahzade, S.M. Nasehi, S.H. Khandkar, and S. Rawal, “The role of patch review in software evolution: an analysis of the Mozilla Firefox”, Proceedings of the joint international and annual ERCIM workshops on Principles of software evolution (IWPSE) and software evolution (Evol) workshops, 2009.

[25] L. Nyman, and T. Mikkonen , “To Fork or Not to Fork: Fork Motivations in SourceForge Projects”, IFIP Advances in Information and Communication Technology, Vol. 365, 2011, pp. 259-268.

[26] L. Nyman, T. Mikkonen, J. Lindman, and M. Fougère, “Forking: the Invisible Hand of Sustainability in Open Source Software”, Proceedings of SOS 2011: Towards Sustainable Open Source, 2011.

[27] L. Prechelt and C. Oezbek, “The search for a research method for studying OSS process innovation”, Empirical Software Engineering, vol. 16 (4), 2011, pp. 514-537.

[28] E.S. Raymond, “Homesteading the Noosphere”, First Monday 3 (10), October 1998, pp. 90-91.

[29] E.S. Raymond, “The Cathedral & the Bazaar (Musings on Linux and Open Source by an Accidental Revolutionary)”, O'Reilly Media, 2001.

[30] Smile, Comprendre l'open source et les logiciels libres, Livre blanc, www.smile.fr.

[31] D. Spinellis and C. Szyperski, “How is Open Source affecting software development?”, IEEE Software, January-February 2004, pp. 28-33.

[32] A.M. St.Laurent, Understanding Open Source and Free Software Licensing, O'Reilly Media, 2004.

[33] M. Välimäki, “Dual licensing in open source software industry”, Systèmes d'Information et Management, vol. 8, n°1, 2003, pp. 63-75.

[34] R. Viseur, “Gestion de communautés Open Source”, 12ème Conférence de l'Association Information et Management, Lausanne, juin 2007.

[35] R. Viseur, “La valorisation des logiciels libres en entreprise”, Jeudis du Libre, Université de Mons, 15 septembre 2011.

[36] R. Viseur, “Associer commerce et logiciel libre : étude du couple Netscape / Mozilla”, 16ème Conférence de l'Association Information et Management, Saint-Denis (France), 25-27 mai 2011.

[37] E. Von Hippel, “User toolkits for innovation”, Journal of Product Innovation Management, vol. 18 (4), July 2001, pp. 247-257.

[38] E. von Hippel and G. von Krogh, “Open Source Software and the “Private-Collective” Innovation Model: Issues for Organization Science”, Organization Science, vol. 14 no. 2, March / April 2003, pp. 209-223.

[39] S. Weber, The success of open source, Harvard University Press, April 30, 2004.

[40] D.A. Wheeler, “Why Open Source Software / Free Software (OSS/FS, FLOSS, or FOSS)? Look at the Numbers !”, www.dwheeler.com, 2007.

[41] T. Yamamoto, M. Matsushita, T. Kamiya, and K. Inoue, "Measuring similarity of large software systems based on source code correspondence", Product Focused Software Process Improvement, vol. 3547, Springer Berlin / Heidelberg, 2005, pp. 530–544.

AUTHORS PROFILE

Robert Viseur was born in Mons, Belgium, in 1977. He graduated from the Faculty of Engineering of Mons. He earned a Ph.D. in Applied Sciences in 2011.

He is Teaching Assistant at the Department of Economics and Innovation Management in the Faculty of Engineering (University of Mons, Belgium) and Senior Research Engineer at the Centre of Excellence in Information and Communication Technologies (Charleroi, Belgium).

Energy-Efficient Dynamic Query Routing Tree Algorithm for Wireless Sensor Networks

Si Gwan Kim
Dept. of Computer Software
Kumoh Nat'l Inst. of Technology
Gumi, Korea

Hyong Soon Park
Gumi Electronics & Information Technology Research
Gumi, Korea

Abstract— **To exploit in answering queries generated by the sink for the sensor networks, we propose an efficient routing protocol called energy-efficient dynamic routing tree (EDRT) algorithm. The idea of EDRT is to maximize in-network processing opportunities using the parent nodes and sibling nodes. In-network processing reduces the number of message transmission by partially aggregating results of an aggregate query in intermediate nodes, or merging the results in one message. This results in reduction of communication cost. Our experimental results based on simulations prove that our proposed method can reduce message transmissions more than query specific routing tree (QSRT) and flooding-based routing tree (FRT).**

Keywords- sensor networks; routing trees; query processing.

I. INTRODUCTION

Wireless sensor networks have emerged as an innovative class of networked systems due to the union of smaller, cheaper embedded processors and wireless interfaces with sensors based on micro-mechanical systems (MEMS) technology. Each node is equipped with one or more sensors, storage and processing resources, and communication subsystems. Each sensor is specialized to monitor a specific environmental parameter such as thermal, optic, acoustic, seismic, or acceleration. The nodes are distributed in the sensing phenomenon. Typical sensor networks incorporate into a variety of military, medical, environmental, and commercial applications.

Sensor networks often contain one or more sinks that provide centralized control. A sink typically serves as the access point for the user or as a gateway to another network. Large sensor networks can be composed of thousands of sensor nodes deployed in the field to observe a region. Sensor networks have several major constraints: limited processing power, limited storage capacity, limited bandwidth, and limited energy. Researchers are working to solve many of the limitations affecting sensor nodes and networks. Some researchers are working to improve node design; others are developing improved protocols associated with a sensor network; still others are working to resolve security issues.

Energy efficiency has been a major concern in sensor networks because most sensor nodes have limited power. If used without care, they will deplete their power quickly [1][2][3][4]. It is known that message communication among sensor nodes is a main source of energy consumption. Typically, wireless communication consumes several thousand times more energy than computation [5]. In the tree-based approach [6][7] a spanning tree rooted at the sink is constructed first. Subsequently this tree is exploited in answering queries generated by the sink. This is done by performing in-network aggregation along the aggregation tree by proceeding level by level from its leaves to its root. The main idea of in-network processing is to reduce volumes of data in the network by partially aggregating sensed values or merging intermediate data. For aggregation queries such as MAX, SUM and COUNT, an intermediate node may aggregate them and send only a newly computed value instead of just forwarding all values received from its children. For example, for a SUM query, an intermediate node forwards only the added value among the values received from its children. These aggregate queries reduce the number of messages, thus reducing power consumption.

In this paper, we propose a query-based routing tree, called energy-efficient dynamic routing tree (EDRT) that is separately constructed for each query by utilizing the query information. The main objective of the EDRT is to minimize the number of hops by increasing the amount of data merge processing, thus reducing the total number of generated messages to reach the destination. The EDRT is constructed in such a way that messages generated from sensor nodes can be merged more often and earlier.

This paper is organized as follows. Section 2 discusses the related works; Section 3 formally defines the EDRT and describes how to construct EDRT in sensor networks. Experimental evaluation of EDRT is presented in Section 4. Finally Section 5 concludes the paper.

II. RELATED WORKS

There has been a lot of work on query processing in distributed database systems, but major differences exist between sensor networks and traditional distributed database systems[8][9][10][11][12]. As sensor networks have limited capabilities such as energy consumption and computation, query processing in sensor networks must take into account these constraints. Much work in construction of efficient routing trees in sensor networks has been done in sensor network applications [13][14][15][16].

When centralized querying is employed in WSN, the base station acts as the point where the query is introduced and results are gathered. The TinyDB Project at Berkeley [17], which is largely used for data gathering in sensor networks, uses spanning trees for the data retrieval, but does not rely on any other in-network data to optimize queries. This centralized technique may not be feasible for self-organizing sensor networks since a query may be initiated from any node in the network and propagating the query to the base station would cost too much. A semantic routing tree (SRT) is a routing tree used in query dissemination to route a query to the nodes that have a possibility to generate tuples for the query. By sending a query only to the nodes that need to receive the query, the SRT can reduce communication cost in query dissemination.

In [18], the minimum distance tree (MD-tree) is separately constructed for each query by utilizing the query information. The MD-tree can increase the amount of in-network processing by constructing the tree in such a way that messages generated from sensor nodes can be merged more often and earlier, thus minimizing the energy consumption. In [19], a query routing trees are formed by balancing the data load to be transmitted from one tree level to the next.

The goal is to balance the data received and relayed by each node in the network. The energy savings in this tree are mostly theoretical since they do not deal with collisions occurring from many nodes trying to communicate with the same parent. Reference [20] proposes the design of a distributed index that scalably supports multi-dimensional range queries. Distributed index for multi-dimensional data (or DIM) uses a novel geographic embedding of a classical index data structure, and is built upon the GPSR geographic routing algorithm. DIFS [21] extends traditional binary-tree and quad-tree by allowing multiple parents and multiple roots. In DIFS, a node may have several parents, which may be located far away. This leads to distance sensitivity problem.

Thus constructing the DIFS tree and update operations are expensive. But DIFS scales well to large-scale networks by using a multiply rooted tree and a geography/value coverage tradeoff that balances communication overhead over many nodes.

III. ENERGY-EFFICIENT DYNAMIC ROUTING TREE

In this section, we present our energy efficient routing algorithm based on dynamic routing tree.

A. *Definition*

We model a sensor network as an undirected graph $G = (V, E)$ where V is a set of nodes and E is a set of edges. A root node can be act as a base station. An edge (v_i,v_j) is in E if two nodes v_i and v_j can communicate each other. Fig. 1 shows a graph for a sensor network with 8 nodes.

The distance from v_i to v_j in graph G for a sensor network is defined to be the length of a path from v_i to v_j with the minimum number of edges. The distance from the root node to v_i is called the "distance of v_i" . In Fig. 1, v_1 is a root node and distance of v_7 is 3.

Parent candidate set CP_i and sibling candidate set CS_i for sensor node i is defined as follows.

$CP_i = \{ v_i \mid v_j \text{ is a neighbor of } v_i \text{ and } l_j = l_i - 1\}$

$CS_i = \{ v_i \mid v_j \text{ is a neighbor of } v_i \text{ and } l_j = l_i\}$

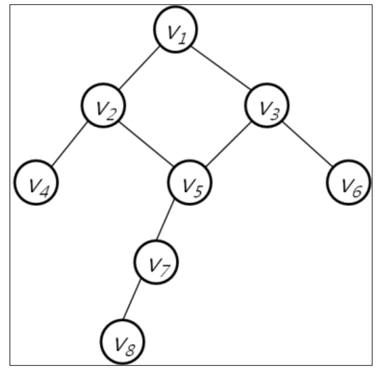


Figure 1. Example of a sensor network

In other words, parent candidate set CP_i is a set of neighbour node that is lower level by one than the given node i. And sibling candidate set CS_i is a set of neighbour node that is same level with the given node i.

A query node is a node which satisfies the query qualification conditions in the WHERE clause of the query. For convenience, the root node is considered as a query node for every query regardless of satisfying the qualification of the query.

The minimum distance of node i for query Q, denoted by $MD_{i,Q}$ is defined as follows:

$$MD_{i,Q} = \begin{cases} 0, \text{if } i \text{ is root node or candidate node} \\ min\{ MD_{j,Q} \mid j \in CP_i \} + 1, \text{otherwise} \end{cases}$$

In other words, if sensor node i is a root node or a candidate node for a query, $MD_{i,Q}$ is 0. Otherwise, $MD_{i,Q}$ is added by 1 the smallest value of the parent candidate set. We use the term *md* instead of $MD_{i,Q}$ for brevity if node i for query Q is known in advance. *Candidate parent md set* $MD_{i,Q}^{CP}$ for node i is defined to be a collection of $MD_{i,Q}$ for CP_i. Each member of this set consists of node *id* and *md* value. *Candidate sibling md set* $MD_{i,Q}^{CS}$ for node i is a collection of $MD_{i,Q}$ for CS_i. Each member of this set consists of node *id* and *md* value as in $MD_{i,Q}^{CP}$. But, if *md* value is not 0, *md* – 1 is stored.

The first node to be received for query Q, denoted as $P_{i,Q}$, is a node which has the smallest *md* value among candidate parent and candidate sibling set. In other words, $P_{i,Q}$ = *MinDistId* ($MD_{i,Q}^{CP} \cup MD_{i,Q}^{CF}$), where *MinDistId* is a function which returns the *id* of the smallest *md* value. If there is more than one node which has the smallest value, the smaller level is selected, and if levels are same, random node is selected.

B. *Our Algorithm*

In this section, we present the process of our algorithm. This process consists of two stages.

- *Candidate Set Decision Stage*: This stage determines the parent candidate set and sibling candidate set for each node.
- *Query Dissemination and EDRT Construction Stage: When a user requests a query, the EDRT for the query is constructed through the query dissemination. Each sensor node calculates the md value and sends the*

query message with this value to neighbor nodes which has the smallest md value.

1) Candidate Set Decision Stage

In this stage, parent and sibling candidate sets are determined for each node. Candidate decision message, denoted as *CDM*, includes *dest_id*, *src_id* and *level*, where *dest_id* is the destination node identifier, *src_id* is the sender node identifier and level is the level of sender node. The level of root node is 0.

In Fig. 2, the path taken by the candidate decision messages are shown in arrows and candidate sets CP_i and CS_i are shown. In Fig. 3, candidate decision processes are shown.

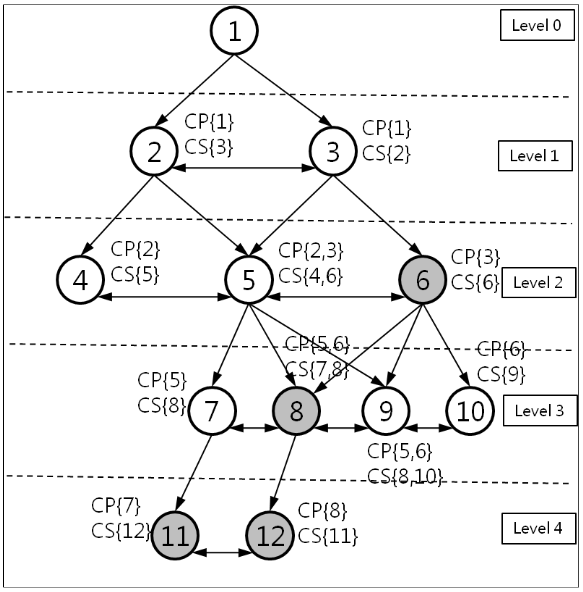


Figure 2. Example of Candidate Set Decision

2) Query Dissemination and EDRT Construction Stage

When a user requests a query, the EDRT for the query is constructed through the query dissemination and candidate set decision stage. In this stage, a query message containing query information and *md* value of a sender floods from the root node down the network. The format of query messages is as follows: <*dst_id*, *src_id*, *md*, *query*>, where *dst_id* is the destination identifier, *src_id* is the sender identifier, *md* is the minimum distance of the sender, and *query* is the query information that contains the query identifier, query, and so on.

Fig. 4 shows the example of how query dissemination and EDRT construction is processed when a user requests a query. In Fig. 4, *md* value is decided through the query dissemination. *md* values are specified on the lines between sibling nodes. These values are shown in pairs, meaning an *md* value for a node is for the other sibling node.

For example, for node 5, *md* value is 0 for the sibling node 6, while for node 6, *md* value is 1 for the sibling node 5.

Input:
CDM (*dest_id* / *src_id* / *level*),
Node i with $level_i$ = *INVALID_VALUE*, $CP_i = \emptyset$ and $CS_i = \emptyset$
Output:
Node i with *level*, CP_i, CS_i
Step :
1. Sink node transmits candidate decision message to root node.
2. Root node broadcasts the message with its identifier value *src_id* and its *level*.
3. When a node receives the message, it checks the following case.

```
if (level of node i == INVALID_VALUE) {
   Set level of node i as value of level field in CDM plus 1
   Add src_id to CP_i
   Broadcast the message with its identifier and level
}
else {
   if (level field in CDM == (level of node i) - 1 ) {
       Add src_id to CP_i
   } else if (level field in CDM == (level of node i) ) {
       Add src_id to CS_i
   }
}
```

4. This process is repeated until all the nodes in the network decide their levels, parent and sibling candidate sets.

Figure 3. Candidate Decision Processes

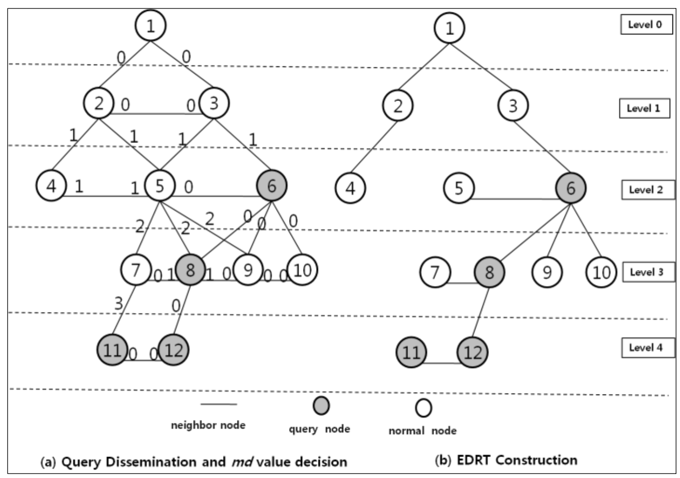


Figure 4. Query Dissemination and EDRT Construction Example

C. Data Gathering in EDRT

Each sensor node sends data, which satisfy the query *Q* that was sent from the sink node, to sink node. While transmitting the result satisfying the query *Q*, each sensor node sends to parent or sibling node along the constructed tree. Each node aggregates the data when receiving the partial result.

Data transmission starts at the bottom of tree up to the root node. Partial aggregation and packet merge operations take place while transmitting packets from the bottom nodes up to the root node. Each sensor node has two transmission opportunities to send. Each sensor node decides the transmission time depending on the status of its parent. Sensor nodes which have some data to send decide the transmission timing depending on the each node's parent node.

In Phase 1, for given query *Q*, sensor nodes with *md* value of parent node is not zero transmit data to the parent node or sibling node. In

Input :
Query Message(*dest_id / src_id / md / query*)
Node *i* with CP_i, CS_i, $MD_{i,Q}^{CP} = \emptyset$, $MD_{i,Q}^{CF} = \emptyset$;
Output:
node N_i ($NextNode_i$ = *MinDistId* ($MD_{i,Q}^{CP} \cup MD_{i,Q}^{CF}$)

1. Sink node delivers query *Q* to root node.
2. Root node broadcasts its id(i.e. *src_id*) and *md* value with 0.
3. If node *i* receives query *Q* message, it checks:
if (*src_id* of query *Q* message $\in CP_i$) {
$MD_{i,Q}^{CP}$ = (*src_id*, *md*);
if ($|MD_{i,Q}^{CP}| == |CP_i|$) {
if (node *i* is candidate node for query *Q*) {
$MD_{i,Q}$ = 0;
} else {
$MD_{i,Q}$= *min*($MD_{i,Q}^{CP}$) + 1;
}
Set Parent node of query *Q* as MinDistId($MD_{i,Q}^{CP}$);
Set its own *src_id* of query *Q* message and broadcast it;
}
} else if (*src_id* of query *Q* message $\in CS_i$) {
if (*md* of query *Q* message == 0)
Set *md* value of sibling node *src_id* of node *i* to 0;
else
Set *md* value of sibling node *src_id* of node *i* to
md value of query *Q* minus 1;
}
4. Repeat step 3 until every node decides its parent node.
5. Each node decides to send its data to node *MinDistId* ($MD_{i,Q}^{CP} \cup MD_{i,Q}^{CF}$).

Figure 5. Query Dissemination and EDRT Construction Process

Phase 2, all sensor nodes that have data to send transmit to parent node only.

Data is transmitted to the node which has the smaller *md* value. If *md* value is same for parent nodes and sibling nodes, node is randomly selected. If *md* value of parent node is same as the sibling node, it is transmitted to the parent node. Fig. 5 shows the sequence of data transmission for same level nodes in the data gathering stage. Nodes 4, 5, and 6 are on the same level, and shaded nodes 2, 5 and 6 have data to send. In Phase 1, node 5 waits because *md* value of its parent node has is 0. Node 6 sends its data to node 5 which has smaller *md* value. In Phase 2, node 5 sends its data to node 2 which has smaller *md* value than node 3, then sends merged data to node 2.

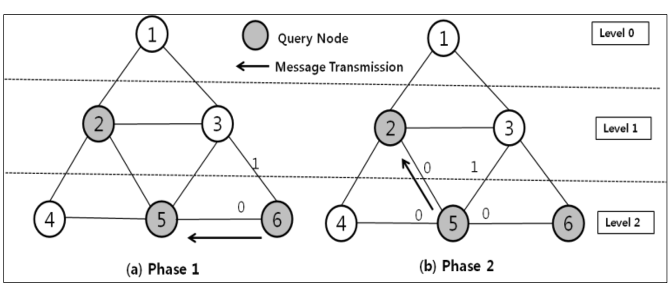


Figure 6. Data transmission sequence in Data Gathering Stage

IV. PERFORMANCE EVALUATION

In this section, we evaluate and compare the performance of three routing schemes among our EDRT, QSRT and naïve FRT. FRT is the general routing tree based on flooding. In FRT, each node selects the parent node which delivers the first query message.

QSRT[18] simply selects the parent node which has the smallest *md* value.

A. Settings

In our simulation experiments, sensor nodes are randomly distributed in a sensor network. A sensor network is of size width *w* and height *h*, with square form. The number of nodes *N* to be distributed in a sensor network depends on the communication range *r* and the number of nodes within the communication range, i.e. node density *d*. The selectivity of a query is the percentage of the query nodes for the query in a sensor network.

Table I summarizes the default values for the parameters used in the simulations. In all the experiments, we have generated 10 sensor networks, executed the simulation 10 times for each sensor network and calculated the average values.

TABLE I. SIMULATION PARAMETERS

Parameter	Value
Density	6 ~ 20
Communication Range	30 m
Query Selectivity	0 ～ 100 %
Initial Energy	2 J
Communication Energy Consumption	50 nJ/bit
Network Size	150m ×150m ~ 1200m×1200 m
Round	10 ～ ∞

Performance metrics are the total number of message transmissions required for one query and the number of messages gathered in the sink node.

We have performed four experiments to evaluate our schemes as follows:

- Query Selectivity : We vary the query selectivity from 1% to 100% to evaluate the effect of various query selectivities among three trees.
- Network Size : In this experiment, we change the network size to evaluate the effect of various network sizes among three trees.
- Node Density : We investigate the effect of various node densities among three trees. We varied the node density from 5 to 19.
- Amount of Data Gathering : We investigate the amount of data gathered in the sink node until the network dies.

TABLE II. NUMBER OF CANDIDATE NODE FOR SELECTIVITY

Query Selectivity (%)	0	10	20	30	40	50	60	70	80	90	100
Number of Candidate Node	0	29	57	86	115	144	172	201	230	259	287

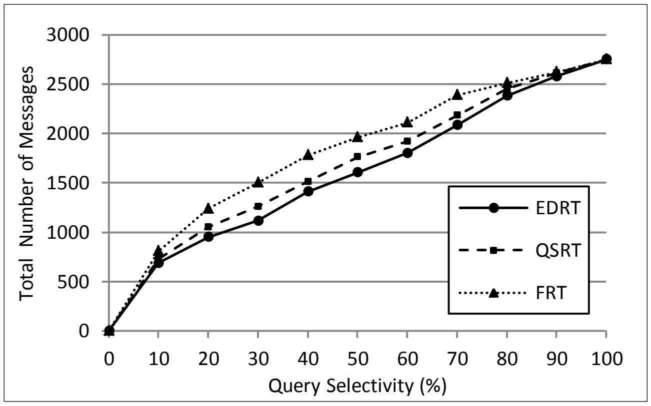


Figure 7. Query Selectivities

B. *Performance of Various Query Selectivities*

We vary the query selectivity from 10% to 100% to evaluate the effect of various query selectivities on the benefit of EDRT over FRT and QSRT. Network size is set to 300m×300m and node density is 9. We used the number of candidate node as in Table II. Fig. 7 shows the simulation results. In the figure, when the query selectivity is less than 20%, the performance of EDRT is similar to that of other trees. This is because a small number of nodes are the query nodes for a query; hence few messages are generated in the network. As the query selectivity increases, the benefit of data aggregates also increases. As the query selectivity approaches 100%, however, the benefit again decreases. This is because all the nodes in the network generate messages: Thus, in-network processing occurs at almost every node in both routing trees. Overall, EDRT outperform other schemes in various query selectivities, with at maximum 25% reduction of message transmissions.

C. *Performance of Various Network Size*

In this experiment, we change the network size from 150m×150m to 1200m×1200m to evaluate the effect of various network sizes on the benefit of EDRT over other trees. Query selectivity is set to 30, and density is set to 9. And Table III shows the number of nodes and the number of candidate nodes with varying size of network for this experiment. Fig. 8 shows the experimental results. In small size networks, the benefit of EDRT is small because there are a small number of nodes in the network. However, as the network size increases, the benefit of EDRT also increases.

When network size is 600m, total number of messages generated for our EDRT is slightly (about 5~6%) less than that of QSRT and 35% less than that of FRT. When the network size is less than 600m, EDRT and QSRT take advantage of in-network processing, thus minimizing the number of generated messages. The reason is that in large sensor networks, messages from sensor nodes are merged within a few hops, rather than transferred up to the base station without being merged. EDRT show better performance over QSRT in various network sizes, with about 10% reduction of message transmissions. But EDRT outperforms than FRT for all the network sizes.

TABLE III. NUMBER OF NODES WITH VARIOUS NETWORK SIZE

Network Size (m)	150	300	450	600	750	900	1050	1200
Number of Nodes	72	287	645	1147	1791	2580	3511	4586
Number of Candidate Node	22	86	194	344	537	774	1053	1376

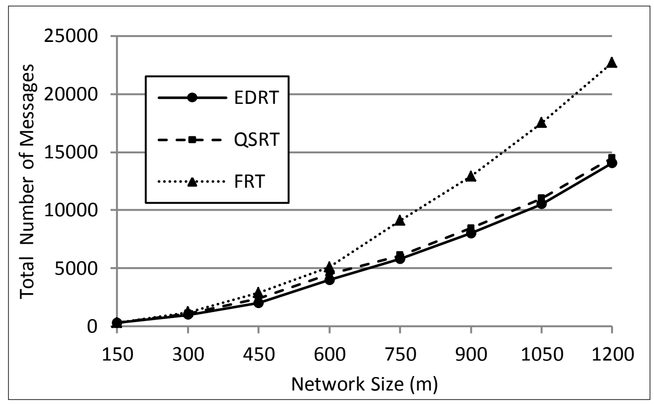


Figure 8. Performance of Various Network Size

D. *Performance of Various Node Density*

We investigate the effect of node densities varying from 5 to 19. Network size is set to 300m×300m and query selectivity is 30. And Table IV shows the number of nodes and the number of candidate nodes with varying node density for this experiment.

Fig. 9 shows the experimental results. As in the figure, the benefit of EDRT over FRT and QSRT increases as the node density increases. In case of low node density, meaning the number of node is small, the probability for aggregates is low. But as the node density increases, the probability for aggregates is high, leading to 12% less messages generated at node density at 13.

E. *Performance of Data Gathering in Sink Node*

In this experiment, we compare the number of messages gathered in the sink node until the sensor network dies after power consumption among three schemes. Network size is 300m x 300m, query selectivity is 30, and density is 9. We generated 10 networks, and each node transmits random messages to sink node.

Fig. 10 shows the experimental results. For less than 4000 rounds, all trees show all the same performance. But as the round reaches near 4000, EDRT performs better than FRT and QSRT. As EDRT requires less hops than FRT and QSRT, this leads to less energy consumption in node, longer network life, and finally more data gatherings in sink node. For above 5000 rounds, EDRT performs 8% better than FRT and 4% better than QSRT.

TABLE IV. NUMBER OF NODES WITH VARIOUS DENSITY

Density	5	7	9	11	13	15	17	19
Number of Nodes	159	223	287	350	414	478	541	605
Number of Candidate Node	48	67	86	105	124	143	162	182

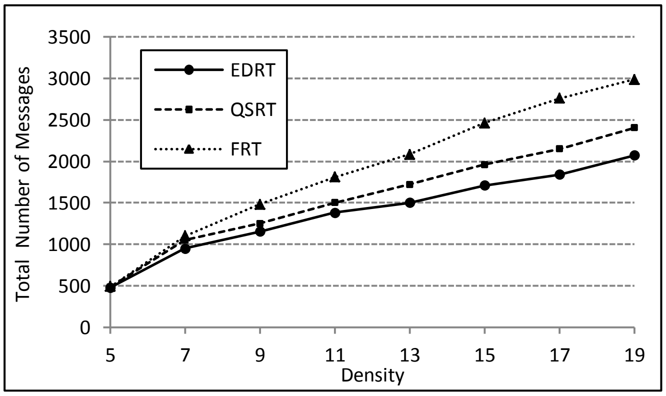


Figure 9. Performance of Various Node Density

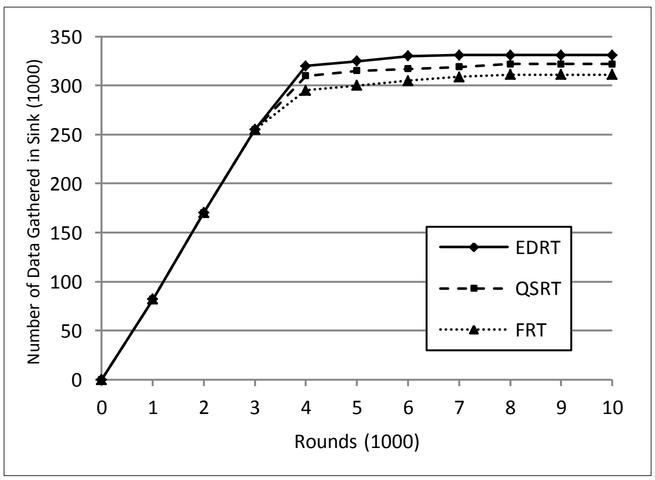


Figure 10. Performance of Data Gathering in Sink Node

V. CONCLUSIONS

In this paper, we proposed a query-based EDRT scheme, which is constructed dynamically for each query. We have designed the EDRT in such a way that data aggregate processing occurs as early as possible in result collection by delivering result messages to the parent and friends node. And we have evaluated the performance of our schemes with other works and have founded our scheme outperforms existing routing trees in various environments. The number of message transmissions for EDRT can be reduced up to 37% and 12%, compared with FRT and QSRT, respectively. And the number of messages received in BS is increased by 8% and 4%, comparing with FRT and QSRT, respectively. For the future research project, we will apply these techniques to the experimental sensor networks for the water pollution surveillance in the reservoir.

ACKNOWLEDGMENT

This paper was supported by Research Fund, Kumoh National Institute of Technology.

REFERENCES

[1] K. Akkaya, M. Younis, "A survey of routing protocols in wireless sensor networks", Ad Hoc Network(Elsevier), p325-349, 2005.

[2] Kay, Römer, Friedemann Mattern, "The Design Space of Wireless Sensor Networks", Wireless Communications, IEEE, Vol. 11, Issue 6, pp.54-61, 2004.

[3] C.K. Toh, "Ad Hoc Mobile Wireless Networks: Protocols and Systems", Prentice Hall PTR, 2002.

[4] Jamal N. Al-Karaki, Ahmed E. Kamal, "Routing Techniques in Wireless Sensor Networks: A Survey", Wireless Communications, IEEE, Vol. 11, Issue 6, pp.6-28, 2004.

[5] W. Heinzelman, A. Chandrakasan and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks", Proceeding of the 33rd Hawaii International Conference on System Sciences (HICSS '00), 2000.

[6] S. Madden, "Database Abstractions for Managing Sensor Network Data," *Proceedings of the IEEE* , vol.98, no.11, pp.1879-1886, Nov. 2010.

[7] J. Gehrke, S. Madden, "Query processing in sensor networks," Pervasive Computing, IEEE , vol.3, no.1, pp. 46- 55, Jan.-March 2004.

[8] Y. Yao and J. Gehrke, "The cougar approach to in-network query processing in sensor networks", ACM SIGMOD Record, Volume 31, Issue 3, pp.9-18, 2002.

[9] F. Ye, H. Luo, J. Cheng, S. Lu, L. Zhang, "A Two-tier data dissemination model for large-scale wireless sensor networks", proceeding of ACM/IEEE MOBICOM, SESSION: Sensor Networks ,pp.148-159, 2002.

[10] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed diffusion: a scalable and robust communication paradigm for sensor networks", Proceedings of ACM MobiCom '00, Boston, MA, pp.56-67, 2000.

[11] Andreou, P., Zeinalipour-Yazti, D., Chrysanthis, P., Samaras, G., "In-network data acquisition and replication in mobile sensor networks", Journal Name: Distributed and Parallel Databases, Vol. 29, No. 1, pp. 87-112, 2011.

[12] Chipara, O., Chenyang L., Stankovic, J.A., Roman, G., "Dynamic Conflict-Free Transmission Scheduling for Sensor Network Queries", IEEE Transactions on Mobile Computing, Vol.10, No. 5, pp.734-748, 2011

[13] S. Madden, M. Frankln, J. Hellerstein, and W. Hong, "TAG: a tiny aggregation service for ad-hoc sensor networks", In Proc. of OSDI, 2003.

[14] Rohm, U., Scholz, B., Gaber, M.M., "On the Integration of Data Stream Clustering into a Query Processor for Wireless Sensor Networks", International Conference on Mobile Data Management, 2007.

[15] De Poorter, E., Bouckaert, S., Moerman, I., Demeester, P., "Broadening the Concept of Aggregation in Wireless Sensor Networks", Second International Conference on Sensor Technologies and Applications, 2008.

[16] Singh, J., Agrawal, R., "Energy efficient processing of location based query in Wireless Sensor Networks", 16th IEEE International Conference on Networks, 2008.

[17] S. Madden, M. Franklin, and J. Hellerstein. TinyDB: an acquisitional query processing system for sensor networks. ACM Trans. on Database Systems, Vol. 30, No. 1, 122-173, 2005.

[18] IC Song, Y Roh, D Hyun, MH Kim, A Query-Based Routing Tree in Sensor Networks , 2nd Int'l Conf. on Geosensor networks (GSN'06), 2006.

[19] T. Yan, Y. Bi, L. Sun, and H. Zhu. "Probability based dynamic load-balancing tree algorithm for wireless sensor networks", Int. Conf. Networking and Mobile Computing, 2005.

[20] L. Xin , Y, Kim, R. Govindan, W. Hong, "Multi-dimensional range queries in sensor networks", Proceedings of the 1st international conference on Embedded networked sensor systems, SenSys '03, 2003.

[21] B. Greenstein, D. Estrin, R. Govindan, S. Ratnasamy, and S. Shenker. "DIFS: A distributed index for features in sensor networks", First IEEE Ineternational Workshop on Sensor Network Protocols and Applications, May 2003.

AUTHORS PROFILE

Si-Gwan Kim received the B.S. degree in Computer Science from Kyungpook Nat'l University in 1982 and M.S. and Ph.D. degrees in Computer Science from KAIST, Korea, in 1984 and 2000, respectively. He worked for Samsung Electronics until 1988 and then joined the Department of Computer Software Engineering, Kumoh National Institute of Technology, Gumi, Korea, as an associate professor. His research interests include sensor networks, mobile programming and parallel processing.

Hyong Soon Park received the B.S. and M.S. degrees in software engineering from Kumoh National Institute of Technology, Gumi, Korea, in 2003 and 2007, respectively. He is with Gumi Electronics & Information Technology Research as a research engineer. His research interests include wireless sensor networks, and mobile computing.

Design of a web-based courseware authoring and presentation system

Engr. Prof Hyacinth C. Inyiama
Electronic and Computer Engineering Dept
Nnamdi Azikiwe University
Awka, Nigeria

Engr. Miss Chioma N. Anozie
Electronic and Computer Engineering Dept
Nnamdi Azikiwe University
Awka, Nigeria

Engr. Dr. Mrs. Christiana C. Okezie
Electronic and Computer Engineering Dept
Nnamdi Azikiwe University
Awka, Nigeria

Engr. Mrs. Nkolika O. Nwazor
Research and Development Department
Electronic Development Institute
Awka, Nigeria

***Abstract*—A Web-based Courseware Authoring and Presentation System is a user-friendly and interactive e-learning software that can be used by both computer experts and non-computer experts to prepare a courseware in any subject of interest and to present lectures to the target audience in full media. The resultant courseware is accessed online by the target audience. The software features automated assessment and grading of students. Top-down methodology was adopted in the development of the authoring software. Implementation languages include Asp.Net and Visual Basic 9.0 and Microsoft Access 2007. An author can use the system to create a web-based courseware on any topic of his choice, while the software platform still remains intact for yet another author. A major contribution of this work is that it eases courseware preparation and delivery for lecturers and trainers.**

Keywords- ***Web; Courseware; Authoring; Presentation; E-Learning; Multimedia.***

I. INTRODUCTION

Paper and pencil correspondence courses have been available for over fifty years, and for motivated students, have been proven to work adequately. However, today's educators of college and university students face new challenges related to the increasing demand for provision of course related resources and documents due to the geometric surge in the population of students in tertiary institutions. Traditional blackboard and paper tutorship is proving to be an inadequate method of ensuring effective learning and assimilation amongst a multitude of students. Also there is a high demand for distance learning in today’s information age which cannot be met by traditional method.

Tertiary institutions in some of the developing and under-developed nations like Nigeria have not only experienced a tremendous increase in the number of students enrolled yearly but have also experienced the disruption of normal school session frequently by strikes. This vast number and the frequent disruption of normal school session have made it difficult for the few lecturers there are to lecture and assess students effectively. Hence there is need for an e-learning solution that enables lecturers to effectively teach and assess students irrespective of the permitted classroom time and the location of the students.

With the growth of information and communication technology, the advent of the world wide web and the introduction of e-learning technology, the development of a solution to extend the classroom to the internet and educate students on a one-on-one basis, is now possible [1]. The web-based courseware authoring and presentation system is such a solution and also incorporates a module for testing the student and hence assessing his performance. It also has an easy-to-use interface, hence ensuring that any lecturer can be an author of a courseware with minimal computer expertise.

II. PROBLEM STATEMENT

To establish the problem statement, the shortcomings of the present method of knowledge dissemination in most developing and under-developed nations (the traditional blackboard and paper system) were identified:

- A student misses a lot of information that cannot be easily passed to him by his fellow students if he is absent from or late to a lecture. This is because the lecturer, who is seasoned by years of experience, skills and education, knows exactly how to describe a topic and what elements to employ to accelerate understanding. Hence when a student misses a lecture, it becomes very difficult to understand the topic that was taught.
- Traditional blackboard lecture delivery method is not tailored to suit the different learning paces of the different students in a classroom.
- The unstable educational environments of tertiary institutions in these nations, caused by strikes and riots make it almost impossible for any lecturer to effectively teach all the topics in his course scheme of work. This poses a big problem as information is mostly relevant in its entirety and not as an incomplete piece.

- The present system is not economical. This is because a lot of people who wish to be educated but are hindered because of distance cannot harness the power of education. Distance learning cannot be conducted in the traditional method of education.

III. Objective

The objective of this paper is to present the design of a web-based courseware authoring and presentation system which is an e-learning system tailored to the needs and challenges (presented in the problem statement) of the educational environment in most developing and under-developed nations of the world.

IV. Definitions And Explanation Of Terms

Courseware is a term that combines the words 'course' with 'software'. Its meaning originally was used to describe additional educational material intended as kits for teachers or trainers or as tutorials for students, usually packaged for use with a computer. The term's meaning and usage has expanded and can refer to the entire course and any additional material when used in reference to an online or 'computer formatted' classroom [1].

The technicalities involved in courseware production deter many teachers from utilizing this new technology. This lead to the development of authoring system for courseware production that will allow novice programmers to produce their own courseware packages [2]. An Authoring System allows a non-programmer to easily create a courseware without going through the rigors of developing the software and writing the codes. The programming features are built in but hidden behind buttons and other tools, so the author does not need to know how to program [3].

A Web-based authoring tool can be considered an e-learning tool because it creates courseware that utilizes the web-hosting facility of the authoring software to be made available online, thereby enabling students to study over the internet irrespective of their location.

V. System Overview

The web-based courseware authoring and presentation system is a platform that can be used by a lecturer/author with little technical skills but basic computer literacy. It is a platform that can be used for the courseware development of any topic or course and can be used multiple times. It is not designed specifically for any particular course and as such can be accessed by different course authors.

As shown in Fig.1, the web-based courseware authoring and presentation system for tertiary institutions has 3 main modules:

- The courseware production module
- The user module
- The online e-learning portal module

The courseware is developed with the aid of the courseware authoring software; it is then published on the web, from where it can be accessed by the appropriate user.

A. The courseware Production

This entails the pre-development of the courseware from course materials. This is normally done offline. Here the courseware author designs the outline of his course (which could be based on international, national or personal schemes). Based on the design, he then categories his work into the following divisions where appropriate:

- Objectives of his course
- The lecture notes/text that would accompany his course
- The lecture media that would shed further insight to his work
- The Quiz questions that would be used to test student’s understanding of his course.

This is a very important stage of courseware production, for any failure or mistake in this pre-development stage will affect the final output of the finished courseware. It may lead to a situation where the intended content is not properly communicated to the students that will use the courseware hence defeating one of the aims of the web-based courseware authoring and presentation system. Text development packages like Microsoft notepad and video development packages like adobe flash are used at this stage to digitize course materials

B. Users

Users, in this context, refer to authorized or registered people with various degrees of access privileges. This means that access to the web-based courseware authoring and presentation system is limited to registered members. Also, the user’s access privilege is determined by his role. The permissible roles in the web-based courseware authoring and presentation system are shown below:

- Lecturer/Author: this user has courseware creation rights as well as courseware presentation rights. Hence he can both create and view previously created courseware.
- Student: this user’s access privilege is limited to viewing available courseware.

C. Online E-learning

This part embodies all events and transactions that occur over the internet. Hence log in of registered members, registration of new members, final development of pre-designed courseware and courseware presentation are under this part. When a registered user logs in to the web-based courseware system, his access privileges are determined and if he has author rights, then upon feeding pre-design courseware materials into the system, the system saves them in the database, from where it can be retrieved for viewing purposes by all users.

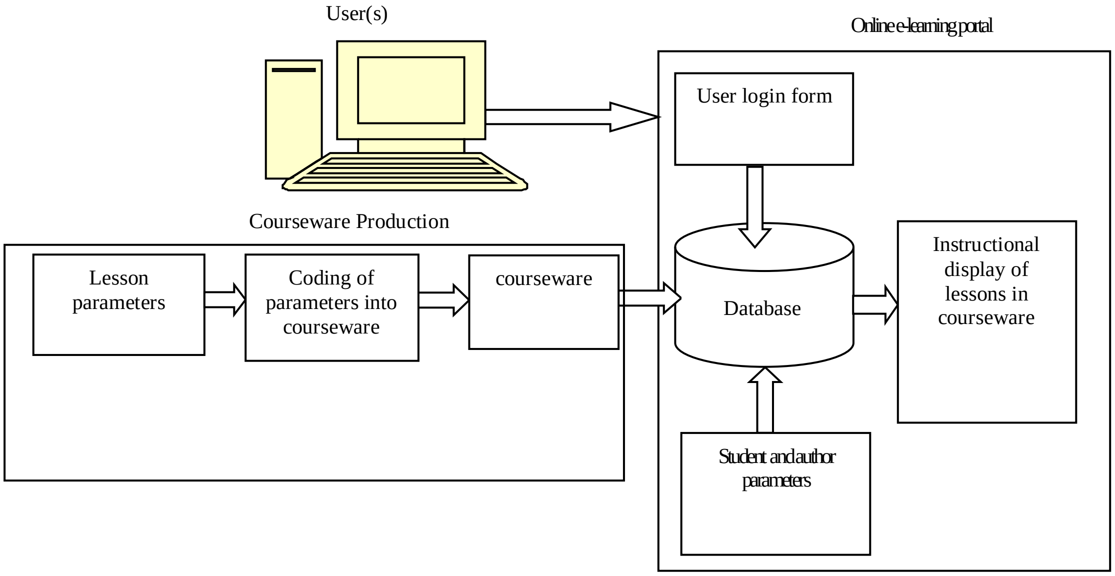


Figure 1. The block diagram representation of the web-based courseware authoring and presentation system for tertiary institutions.

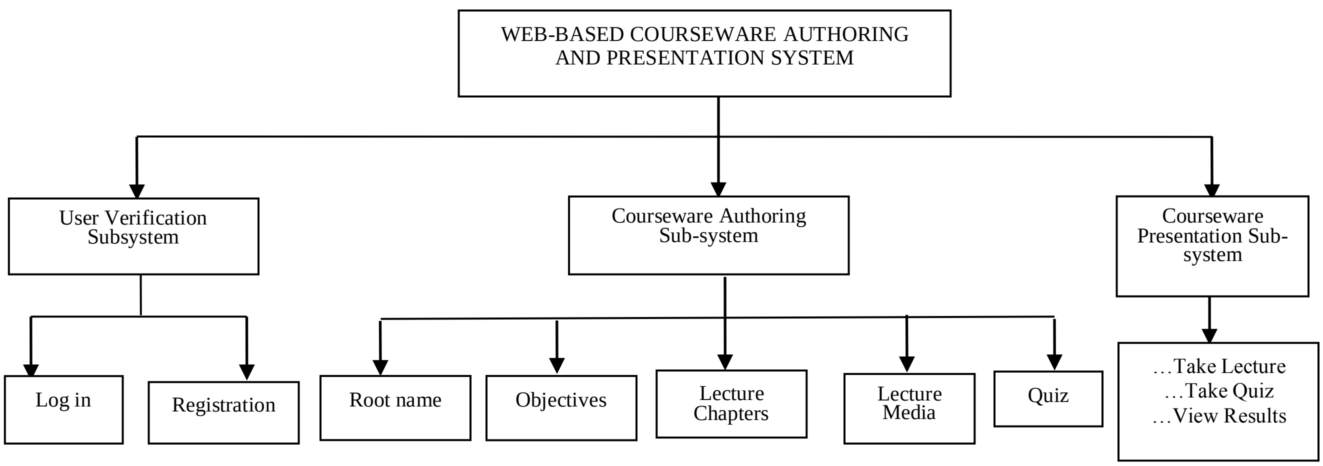


Figure 2. Top-down Design procedures for the Web-based Courseware Authoring and Presentation System

VI. DESIGN SPECIFICATION

The top-down design of the web-based courseware authoring and presentation system (WCAPS) in high level model is shown in Figure2 to illustrate the system specification of the project.

A. *Software sub-systems*

The web-based courseware authoring and presentation system is grouped into four subsystems. These subsystems are further divided into modules. Asp.net and VB.net were combined to develop the courseware authoring and presentation system.

1) *User verification subsystem:* This subsystem has two modules: login module and registration module. It is the foremost subsystem encountered by a user. Access is granted to only those whose data are stored in the database; else the prospective user is redirected to the registration page. The lecturer table and the student table are utilized for the smooth operation of this subsystem. The flowchart of the user verification subsystem is shown in Fig3 and 4

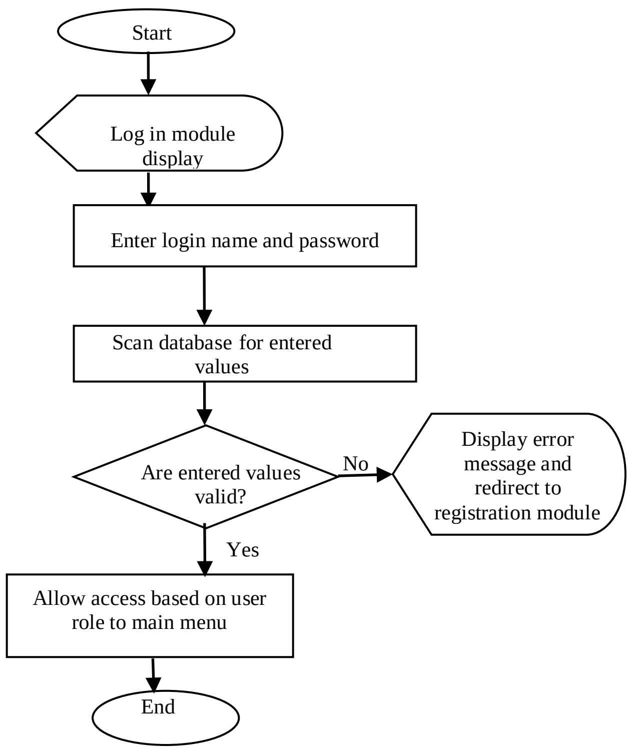


Figure 3. The flowchart representation of the log in module of the user verification subsystem

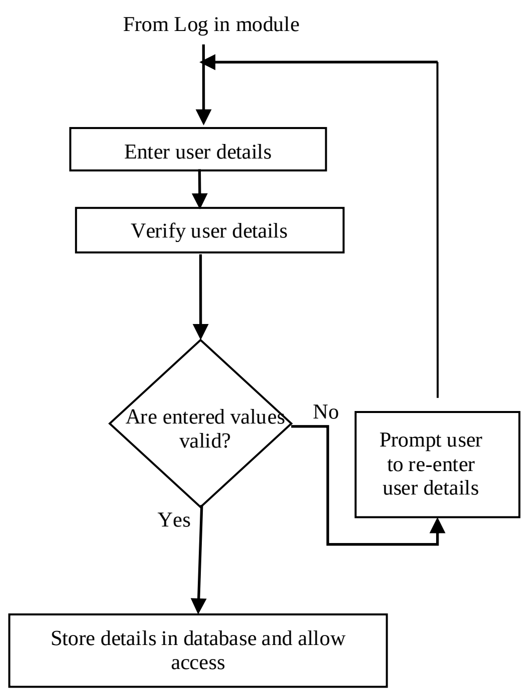


Figure 4. The flowchart representation of the registration module of the user verification subsystem

2) *Courseware authoring subsystem:* In this subsystem, the lecturer/author is able to create his pre-designed courseware. At the beginning, the author supplies at prompting the materials/ parameters necessary for the development of his courseware. These include his identity data, the files necessary for the different modules of his course and the questions necessary for the quiz. This sub-system comprises of five modules: Rootname Generation module, Objectives module, Tutorial (lecture notes) module, Tutorial (media files), Quiz and Score module.

a) *Rootname Generation module:* The rootname for every courseware is coined from a concatenation of the author's name, courseware title, and courseware topic and courseware code. This way a unique name is formed for every courseware that is created irrespective of whether it is by the same author. The rootname is then stored in the courseware table with its corresponding courseware details (topic, author, title, code). Subsequently as the courseware author continues to the next stage, the rootname is used to save all entered data in their respective tables. This is done this way so that, when a courseware is requested, the courseware processor gets the rootname of the requested courseware from the courseware table and then goes to other tables and searches for the locations of files bearing this rootname and uploads those files to make the courseware functional.

b) Objectives module: This section is linked to the Objectives of the courseware, thereby stating what the aim of the courseware is and what a student should be equipped with after studying the courseware. The objectives file is uploaded from the client side computer and then stored in a server-side file called courseware items. The location of this objectives file in the server is saved to the objectives table against the current rootname.

c) Tutorial module: the tutorial module is divided into the lecture notes part and the media part. While the lecture notes (as the name implies) is culled from the lecturer's notes and/textbook, the media files are scenes captured during a lecturer's lecture. The media files help to further emulate the classroom experience. All these files are uploaded as well and saved in the server folder called courseware items. The respective locations of the files are saved in the chapters table (for the lecture notes) and the Multimedia table (for the lecture media) where appropriate, against the current rootname.

d) Question module: This section displays the questions associated to each tutorial topic. In this module, students are exposed to questions on the courseware topic and the answers associated to each question supplied as well. It serves as a good revision venue. During the development of the quiz, the quiz builder in the web-based courseware authoring and presentation system wizard is used to input questions and save them directly to the quiz table along with their options and the right answer. These questions are saved against the current rootname.

The objectives table, chapters table, multimedia table, quiz table and courseware table are used for all data storage in this subsystem. Fig.5 is the flowchart for the courseware authoring subsystem.

3) *Courseware presentation Subsystem:* This subsystem is comprised of a single module and entails a user-friendly

interface for viewing available courseware. The flowchart is shown in Fig.6.

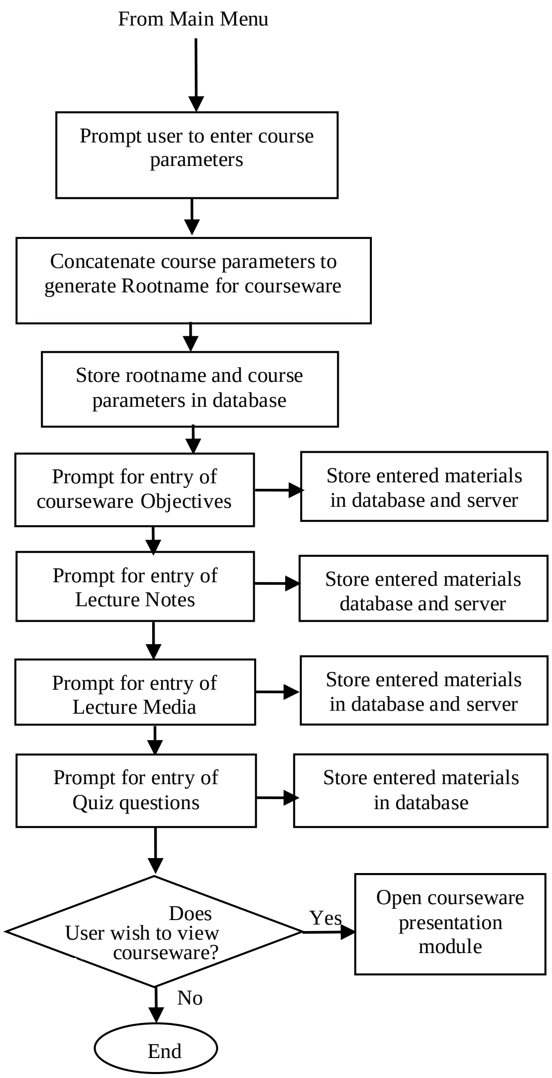


Figure 5. The flow chart of the courseware authoring subsystem

4) *Database Specification:* Database is a collection of logically related data. The Database Specifications are intended to support program coding and database generation by the development group. The database structure, content, data fields, and records were defined. The database for this project was designed using Microsoft Access 2007 database management system. The database of the structured logic design tutor has the following tables:

- Courseware Registration
- Lecturer
- Student
- Objectives
- Chapters
- Multimedia
- Quiz

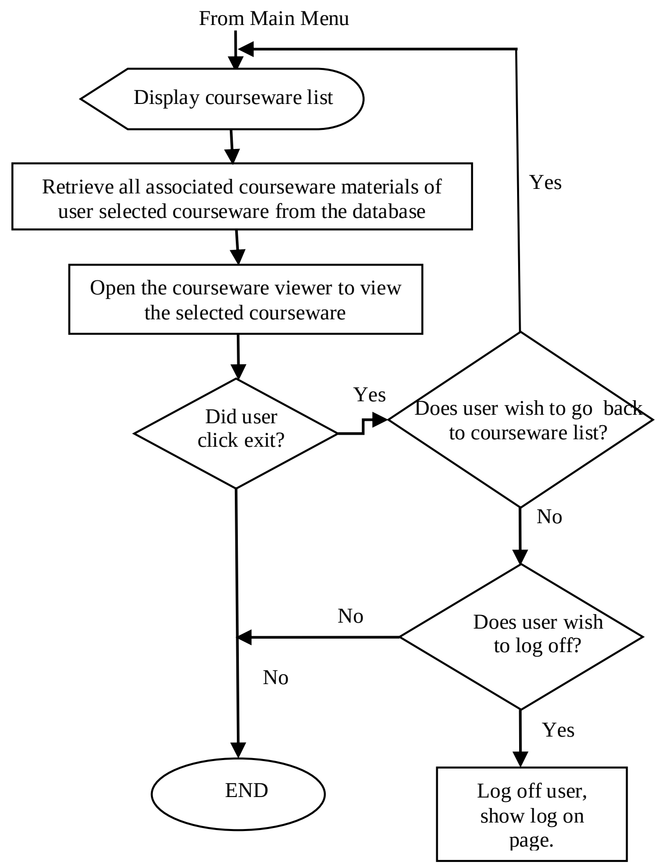


Figure 6. The flowchart of the courseware viewing subsystem

Table1 is a sample of the database table design used in the system

TABLE I. A TABLE DESIGN FOR COURSEWARE REGISTRATION

FIELD NAME	FIELD TYPE	FIELD SIZE	KEY
CoursewareID	Autonumber	Long integer	Primary Key
CoursewareAuthor	Text	100	
CoursewareTitle	Text	150	
CoursewareCode	text	10	
Rootname	memo		

VII. COMPARISON WITH ESTABLISHED COURSEWARE AUTHORING SYSTEMS

There are some courseware's authoring systems available in the market. They include Authorware, Moodle to mention a few.

Authorware is an icon-based multimedia authoring tool which allows the rapid development of complex interactive multimedia projects, particularly courseware and kiosk applications, for both the Macintosh and Microsoft Windows operating systems.

It consists of three main elements: interactive courseware, written in Authorware, which teaches the student basic concepts involved in Authorware programming, and demonstrates the function of each of the icons used to program in Authorware; a tutorial through which students are given the opportunity to use Authorware to incorporate various media elements, including written audio, graphics, video, and text, into their own interactive courseware; and various course materials, including a statement of objectives, study questions, and quiz questions [4].

However, Authorware requires an understanding of its courseware development language. It also requires a license to use and buying it is steep. It also requires an installation process hence meaning that should the courseware author be in a different location and without access to his system, he can't create coursewares. The Web-Based Courseware Authoring System, though lacking the functional complexity of authorware, with its simplicity ensures an author can create coursewares irrespective of his location or access to his personal system. It also requires no knowledge of any programming language.

Moodle is a learning management system that lets you provide documents, graded assignments, quizzes, discussion forums, etc. to your students with an easy to learn and use interface. Moodle is open-source, meaning that the programming code that runs it can be changed to meet the specific needs of users and institutions. Moodle is also free to download and use; there is no licensing fee [5].

However, the drawback is the need to download Moodle to one's personal computer in order to create coursewares. The Web-Based Courseware Authoring System does not need an installation in a personal system. All it needs is internet access and valid authentication data.

VIII. CONCLUSION

During the period of this research work, the limitations of the traditional educational system predominantly used in developing and underdeveloped nations were identified. From there, the functional requirements of the web-based authoring system were developed. It was also necessary to put into consideration the limited programming expertise of the courseware authors. The system prototype was then compared with existing authoring systems like Authorware and Moodle. To ensure a robust system, Visual Basic.Net and Asp.Net was used to develop the Web-Based Courseware Authoring System due to their object-oriented architecture and robust development environment [6][7].

The web-based courseware authoring system was tested with different systems and browsers and it worked seamlessly whilst deploying expected results. However, it does have its shortfalls. In areas such as video conferencing, real-time online blackboard teaching, real-time question and answer forum using chat technology the Web-based Courseware Authoring System is lacking andtherefore further research in these areas is encouraged.

REFERENCES

[1] Routledge, E-Learning: A Framework For Research And Practice. ISBN 0-415-26346-8

[2] Judi Johnson, (2003), Then, Now And Beyond…. A Look At The Past 30 Years Of Educational Software, Learning And Leading With Technology Volume 30 Number 7

[3] Mayer R.E (2001), Multimedia Learning. Cambridge: Cambridge University Press. ISBN 0-521-78239-2

[4] David R. DeVaux (1996), A tutorial in Authorware

[5] Moodle - A Free, Open Source Course Management System for Online Learning. oit.montclair.edu/documentation/moodle/What_is_Moodle.doc

[6] Steven Holzner (2002), Visual Basic .NET Black Book, Paraglyph Press, 2002 pp.800-810

[7] Thiru Thangarathinam(2009),Deploying ASP.NET Applicatios-Part1WebMediaBrands Inc(2009) pp.5-10

The Role of ICT in Education: Focus on University Undergraduates taking Mathematics as a Course

Oye, N. D.
Department of Information Systems
Universiti Teknologi Malaysia

Z.K. Shallsuku,
FUTY-Nigeria

A.Iahad, N.
Department of Information Systems
Universiti Teknologi Malaysia

***Abstract*— This paper examines the role of ICT in education with focus on university undergraduates taking mathematics as a course. The study was conducted at the Federal University of Technology Yola, Adamawa State, Nigeria. 150 questionnaires were administered to the first year students offering MA112 which was completed and returned in the lecture hall. According to the study ICT usage shows that majority (32.7%) use technology once or more in a day. Again the majority of the respondents (35%) said that the greatest barrier to using ICT is technical. The survey shows that there is significant correlation between the students and the use of ICT in their studies. However difficulties facing ICT usage is highly significant also. This shows that students have negative attitudes towards using ICT in their academic work. This is a foundational problem which cannot be over emphasized. Most of the students have never practice using ICT in their primary and secondary schools. Recommendations were made, that the government should develop ICT policies and guidelines to support all levels of education from primary schools to university. ICT tools should be made more accessible to both academic staff and students.**

Keywords- University undergraduates; ICT; Education; ICT tools; ICT infrastructure.

I. INTRODUCTION

E-learning is a unifying term used for online learning, web-based training and technology delivered instruction. E-learning is an approach to facilitate and enhance learning both through computer and communication technology. The use of ICT to support teaching and learning within mathematics remains underdeveloped. While there are examples of good practice, there are significant inconsistencies between schools as well as within mathematics departments. Majorities of teachers are still not confident in the use of ICT and requires further training. In some schools and colleges, access to ICT facilities, including graphing calculators, is too limited and an appropriate range of software has not been made available. In other places, where resources are adequate, they are often not used frequently enough or to promote better teaching and learning. Computers are seen to have the potential to make a significant contribution to the teaching and learning of mathematics. In particular, when students are working on computers, it is generally recognized they are more able to focus on patterns, connections between multiple representations, interpretations of representations and so on [1-4]. This sort of computer use may 'enable a deeper, more direct mathematical experience'. In the UK there is a statutory requirement for ICT to be used in mathematics teaching and learning at all levels; the requirements are summarized in statements such as Students should be offered the following opportunities … using ICT such as spreadsheets, dynamic geometry or graphing software, calculators (DCSF, 2008) In other countries, similar encouragement is given to teachers, and is described by as incentive action and institutional support. However, despite these efforts to embed computers in mathematics teaching and learning, and despite the growing numbers of computers in schools [1], there are some concerns about the degree to which computers have actually become embedded in mathematics classrooms. These concerns fall into three areas: computers are used only marginally in mathematics classes; integration of computers progresses slowly, [3, 5] and where computers are used, they are often used by the teachers in whole class teaching rather than by the students. The growth of ICT has drastically reshape the teaching and learning processes in our universities ([6, 7]. ICT for education is more critical today than before [8]. The higher education institutions around the globe have increasingly adopted ICT as tools for teaching, cirriculm development, staff development, and student learning [9, 10].

According to the [11] much of our curricula and education systems are still products from a mechanistic past, in which predetermined knowledge was delivered in a linear format to a mass audience. The focus was on transferring information in a controlled sequence without accounting for the contextual settings of the different learners. The Universities in Nigeria need to align its teaching and learning methods with best practices found both nationally and globally. Adopting the use of ICT and IS within higher education seems inevitable as digital communication and information models become the preferred means of storing, accessing and disseminating information.

Ease of Use

II. ICT IN EDUCATION

The term, information and communication technologies (ICT), refers to forms of technology that are used to transmit, store, create, share or exchange information. This broad definition of ICT includes such technologies as: radio, television, video, DVD, telephone (both fixed line and mobile phones), satellite systems, computer and network hardware and software; as well as the equipment and services associated with these technologies, such as videoconferencing and electronic mail. ICT in education means teaching and learning with ICT. Researches globally have proved that ICT can lead to improve students' learning and better teaching methods. A report made by the National Institute of Multimedia Education in Japan, proved that an increase in student exposure to educational ICT through curriculum integration has a significant and positive impact on student achievement, especially in terms of "Knowledge Comprehension", "Practical skill" and "Presentation skill" in subject areas such as mathematics, science, and social study. While we recognize that the use of instructional technology in the higher education teaching and learning processes is still in its infancy in Nigeria, ICT instructional use is vital to the progress and development of faculty and students alike. Higher education institutions, especially those in the West, have adopted ICT as a means to impart upon students the knowledge and skills demanded by 21st century educational advancement [9, 10, 12]. ICT also adds value to the processes of learning and to the organization and management of learning institutions. Although some HEIs have the zeal to establish effective ICT education programmes, they are faced with the great problems of proper implementation of the programme. The most important of these is poor ICT penetration and usage among Nigerian higher education practitioners. Almost all African countries' basic ICT infrastructures are inadequate; a result of a lack of electricity to power the ICT materials and poor telecommunication facilities. Above all, this lack of access to infrastructure is the result of insufficient funds[13, 14]. Many cities and rural areas in Nigeria still have fluctuation in their supply of electricity which makes the implementation of ICT in education most difficult. Furthermore most Nigerian universities do not have access to basic instructional technology facilities, which also make the integration of instructional technology in the delivery of quality education difficult. Therefore, computer related telecommunication facilities might not be very useful for most Nigerian students and faculty members, as computers are still very much a luxury in institutions, offices and homes. This has made the integration of essential on-line resources (e-mail, world-wide-web, etc.) into higher education most difficult. In higher education, an important aspect of the shift in technological processes has been to the acceptance and use of ICT for teaching and learning[15]. According to the Commonwealth of Learning International (2001)," another serious challenge facing higher education in Nigeria is the need for integration of new ICT literacy knowledge into academic courses and programs. In this regard, professionals in Nigeria have not been able to benefit from international assistance, international networking and cooperation, or from courses, conferences and seminars abroad, because of lack of funding."

Maintaining the Integrity of the Specifications

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

Research Questions

1. Do students use ICT to support their studies?
2. What is the greatest barrier to using ICT by the students?

III. METHODOLOGY

The research sample are the undergraduate offering MA112 (introduction to calculus) as a major course in the school of pure and applied science FUTY-Nigeria. 150 questionnaires were administered to the first year students which was completed and returned in the lecture hall. In the questionnaire ICT refers to the application of digital equipment to all aspects of teaching and learning which encompasses (PC, TV, Radio, Cellular Phones, Laptop, Overhead Projectors, Slides Projectors, Power-Point Projectors, Electronic Boards, Internet, Hardware, Software, and any technological equipment for teaching and learning). The student is to rate each question based on 1-5 Likert scale, where (1) is Strongly Disagreed and (5) is Strongly Agreed. Descriptive statistics were used to answer the demographic statement. Our statistical results are obtained using SPSS version 17. Using correlation analysis, this paper wants to verify the significant relationship between the undergraduate students and the use of ICT to assist them in their studies.

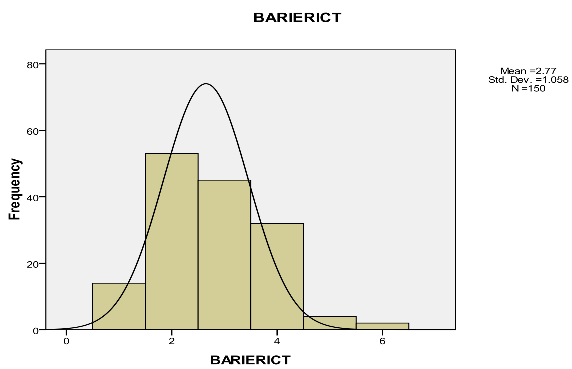
IV. SUMMARY OF THE DEMOGRAPHIC STATEMENT

- The demographic of the survey participants for the gender shows that (60%) are male and (40%) are female.
- Most of the students are within the age bracket of (19-25years) which is (77.3%).
- Is ICT mandatory or voluntary at your institution (ICTMV)? The majority of the students (70.7%) responded that ICT is voluntary. While 24% of the student responded that ICT is mandatory. This may be due to the fact that most of the first year students have graduated from colleges where ICT in education is not encouraged or ICT facilities are not available.
- The question, how often do you use ICT (HUSEICT)? The summary of the technology usage are as follows: (32.7%) use technology once or more in a day, (23.3%) use it once a week, while (20.7%) claimed that they have never use technology.
- Most of the respondents are in level 1-2, (89%). Here we have carry over students they are those that cannot pass the course in their level 1. Again we have some levels: 3(5%); 4(3%) and 5(3%) also repeating the course MA112.

What are the greatest barriers to using ICT to you as an undergraduate? The majority of the respondents (35.3%) said that their problem is technical; on the other hand (30%) said that the problem is cost. Others respondents (21.3%) said that training are their problems, another group (9.3%) said that they need time and the final group (1.3%) said that, it does not fit their programme.

TABLE I. BARRIERS TO USING ICT(BARIERICT)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	TIME	14	9.3	9.3	9.3
	TECHNICAL SUPPORT	53	35.3	35.3	44.7
	COST	45	30.0	30.0	74.7
	TRAINING	32	21.3	21.3	96.0
	COMPENSATION	4	2.7	2.7	98.7
	DOES NOT FIT MY PROGRAM	2	1.3	1.3	100.0
	Total	150	100.0	100.0	



ICT development programme among academic staff of educational institutions especially at the tertiary level is faced by number of obstacles. Prominent among them is the lack of training opportunities for staff. The same problem is recurring in this study again. In a study by[16] lack of interest, limited access to ICT facilities and lack of training opportunities were among the obstacles to ICT usage among academic staff. [17] Opined that inadequate ICT facilities, excess workload and funding were identified as major challenges to ICT usage among academic staff in Nigerian universities. This is affecting the students because they primarily depend on the academic staff.

V. CORRELATION

You will find (Tables 2,3, 4, 5, 6, and 7) in appendix A:

- From table2, ICTEDU1 is significant to ICTEDU2, with r =33.4 and the p-value .000. Correlation result reveals that students' interest to use ICT has significant relation with educational system supporting ICT use in education.
- ICTEDU1 is highly significant to ICTEDU3, with r = 36.7 and p-value .000. Here the correlation result reveals that students' interest on the use of ICT has significant relation with the use of ICT in the education system.
- ICTEDU1is highly significant to ICTEDU4, with r = 37.9 and p-value .000. The correlation result reveals that students' interest to use ICT has significant relation with ICT making education easier for the students.
- ICTEDU3 is also significant to ICTEDU4, with r = 24.8 and p-value .002. Correlation result reveals that the use of ICTs in education is significant to ICT making education easier.
- ICTEDU3 is significant to ICTEDU5, with r = 28.4 and p-value .000.Correlation result reveals that, the use of ICTs in education is significant with promoting education.
- ICTEDU4 is significant to ITCEDU2, with r = 32.0 and p-value .000.Correlation result reveals that ICT making education easier is significant to ICT promoting education.
- ICTEDU5 is significant ICTEDU2, with r = 16.6 and p-value .043. Correlation result reveals that ICT promoting education is significant to educational system supporting ICT use in education.
- From table 3, USEICT1 is significant to USEICT2, with r =20.4 and p-value .012.The correlation result reveals that ICT as a necessary part of education has significant relation with student training on the use of ICT in education.
- USEICT1 is significant to USEICT3, with r = 35.1 and p-value .000. The correlation result reveals that ICT as a necessary part of education has significant relation with use of ICT in the education system.
- USEICT1 is significant to USEICT4, with r = 26.5 and p-value .000. The correlation result reveals that ICT as a necessary part of education has significant relation with ICT making students independent and self learners.
- USEICT2 is significant to USEICT3, with r = 18.7 and p-value .022. The correlation reveals that students training on ICT have significant relation with ICT assisting students in their work.
- USEICT3 is significant to USEICT4, with r = 20.6 and p= value .012. The correlation reveals that ICT assisting students in their work has significant relation with student independent and self-learners.
- From table 4, AICTT1 is significant to AICTT2, with r = 32.4 and p-value .000. The correlation reveals that use of ICT depends on the available tools and this has significant relation with lack of ICT tools.
- AICTT2 is significant to AICTT3, with r = 16.3 and p-value .047. The correlation reveals that lack of ICT tools has significant relation with availability of on-line courses.
- AICTT3 is significant to AICTT4, with r = 31.6 and p-value .000. The correlation reveals that availability of on-line courses has significant relation with availability of ICT tools and facilities.

- From table 5, ICTA1 is significant to ICTA2 with r = 32.8 and p-value .000. The correlation result reveals that students' awareness of ICT has significant relation with existing methods to support ICT.
- ICTA3 is significant to ICTA5, with r = 26.1 and p-value .001. The correlation reveals that ICT awareness for teaching and learning has significant relation with importance of using ICT in education.
- ICTA3 is significant to ICTA4, with r = 35.1 and p-value .000. The correlation reveals that ICT awareness for teaching and learning has significant relation with awareness of ICT among students.
- ICTA4 is significant to ICTA5, with r = 19.2 and p-value .018. The correlation reveals that improved awareness of ICT among students has significant relation with institutions using ICTs in education.
- From table 6, DFICTU1 is highly significant to DFICTU2 with r = 47.6 and p-value .000. The correlation result reveals that teachers' factor has significant relation with the lecture hall facilities and size.
- DFICTU1 is significant to DFICT3, with r = 25.7 and p-value .002. The correlation reveals that teachers factor has significant relation with energy related problem.
- DFICTU1 is significant to DFICTU4, with r = 26.1 and p-value .001. The correlation reveals that teachers factor has significant relation with inadequate internet connectivity.
- DFICTU1 is significant to DFICTU5, with r = 27.5 and p-value .001. The correlation reveals that teachers factor has significant relation with student knowledge of computer usage.
- DFICTU1 is significant to DFICTU7, with r = 25.1 and p-value .002.The correlation reveals that teachers factor has significant relation with students' attitudes towards computer.
- DFICTU2 is significant to DFICTU3, with r = 30.7 and p-value .000.The correlation shows that lecture hall factor has significant relation with energy related problem and power supply.
- DFICTU2 is significant DFICTU4, with r = 29.8 and p-value .000. The correlation shows that lecture hall factor has significant relation with inadequate internet connectivity.
- DFICTU2 is significant to DFICTU5, with r = 33.1 and p-value .000.The correlation shows that lecture hall factor has significant relation with students' knowledge of computer usage.
- DFICTU2 is significant to DFICTU6, with r = 24.9 and p-value.022. The correlation shows that lecture hall factor has significant relation with the availability of computer.
- DFICTU2 is significant to DFICTU7, with r = 26.0 and p-value .001. The correlation reveals that lecture hall factor has significant relation with students' attitudes towards computer.
- DFICTU3 is highly significant to DFICTU4, with r = 48.9 and p-value .000. The correlation reveals that electricity power supply has significant relation with inadequate internet connectivity.
- DFICTU3 is significant to DFICTU5, with r = 36.8 and p-value .000. The correlation shows that electricity power supply has significant relation with knowledge of computer usage.
- DFICTU4 is significant to DFICTU5, with r = 45.9 and p-value .000. The correlation result reveals that inadequate internet connectivity has significant relation with knowledge of computer usage.
- DFICTU4 is significant to DFICTU6, with r = 42.1 and p-value .000. The correlation result shows that inadequate internet connectivity has significant relation with availability of computers.
- DFICTU5 is highly significant toDFICTU6, with r = 46.9 and p-value .000. The correlation result reveals that Knowledge of computer usage has significant relation with availability of computers.
- DFICTU5 is significant to DFICTU7, with r = 38.3 and p-value .000. The correlation shows that knowledge of computer usage has significant relation with students' attitudes towards computer.
- DFICTU6 is significant to DFICTU7, with r = 26.8 and p-value .001. The correlation result reveals that availability of computers has significant relation with students' attitudes towards computer.

VI. DISCUSSION

From table 7, we can summarize that for ICT in Education (ICTEDU), students are interested to use ICTs in education. They believe that ICT will make education easier. Again they believe that the use of ICT will make education system more effective. In the case of (USEICT), the students see ICT as necessary part of education which can assist students in their educational pursue .On the issues of difficulties facing ICT usage (DFICTU). The students consider the energy related problem to be one of the factors affecting inadequate internet connectivity. Hence they believe that if power supply is stable, then internet connectivity will improve. Another problem noted by the student is that most of the teachers are not ICT literate. In addition ICT facilities are not available in the lecture halls, which is crowded with students. The students' attitudes towards computer are negative because most of them do not have the knowledge of computer usage. This is a foundational problem from their primary and secondary school education. In the case of the availability of ICT tools (AICTT), the students state that the use of ICT depends on its' availability therefore, lack of ICT tools affect the use of ICT in education. It is true that the students are aware of ICT (ICTA), however, the teaching methods are not supporting the use of ICT because most of the teachers are not ICT literate.

VII. CONCLUSION

The case of ICT for education and for university undergraduates offering mathematics course in particular are more critical today than ever before since new means of improving instructional methods are triggering a change in the delivery of education. This paper confirms the response of university undergraduate to the role of ICT in education. The summary of the ICT usage shows that majority (32.7%) use technology once or more in a day. The majority of the respondents (35%) said that the greatest barrier to using ICT is technical. The inadequacy of the ICT facilities and infrastructures and lack of opportunities for the university undergraduates for training on how to use ICT for learning and interactions are major problems that must be looked into by the university management. Recommendations were made, that the government should develop ICT policies and guidelines to support both (Primary and Secondary levels), university undergraduates and the academic staff in their academic work. ICT tools should be made more accessible to both academic staff and students.

ACKNOWLEDGEMENT

The authors would like to acknowledge all the authors of articles sited in this paper. In addition, the authors gratefully acknowledge UTM, Research Universiti Malaysia for their support and encouragement.

Appendix A

QUESTIONNAIRE FOR UNIVERSITY STUDENTS

ICT here refers to the application of digital equipments to all aspects of teaching and learning, which encompasses (PC, TV, Radio, Cellular phones, Laptops, overhead projectors, slide projectors, power-point projector, electronic boards, internet, hardware, software, and any technological equipment for teaching and learning).

Please rate each of the following on 1-5 Likert scale. Where (1) “Strongly Disagree,” (2) “Disagree”, (3)” Neither Agree nor Disagree”, (4) “Agree”, and (5) “Strongly Agree”.

ICT in Education:

1. Students are interested to use ICTs in Education. ----------------------------------
2. Our educational system is sufficient to support ICTs to be used in education.--------------
3. By the use of ICTs in education, the education system can be made more effective.--------
4. ICTs have made education easier----------------
5. ICTs is invented to promote education-------------

Use of ICTs

1. ICTs is a necessary part of education--------------
2. Student training is essential, apart from using ICTs in education? -----------
3. Can ICTs assist students in education? -------------
4. ICTs have made students independent and self learners--------------------
5. Most people think that ICTs is limited to computer and internet----------

Difficulties Facing ICTs usage:

1. Teacher factor-------
2. Lecture hall factor-----
3. Energy related problem(Electricity Power supply)--------
4. Inadequate internet connectivity-----
5. Knowledge of computer usage---------
6. Availability of computers-------
7. Students’ attitude towards computer------

Availability of ICT Tools:

1. Use of ICTs depends on the available tools---------------
2. Lack of ICT tools affects the use of ICTs in education---------
3. Availability of on-line courses make education effective----------
4. The availability of ICT tools will ensure the development of ICT infrastructure--------
5. Students with computer knowledge facilitate the use of ICTs in education-------------

ICTs Awareness:

1. Students are well aware of ICTs-------------
2. Existing methods of teaching are enough to support ICTs- -------
3. ICT in education create awareness for teaching and learning------------
4. There is still need to improve the awareness of ICTs among students---------
5. Institutions are trying to create awareness about the importance of using ICTs in education----------

Demographic Information:

1. Gender: 1=Male 2=Female.
2. Age: 1= 16-18years, 2= 19-25 years, 3= 26years and above
3. Is ICT use mandatory or voluntary at your institution?
1= Mandatory 2=Voluntary
4. How often do you use ICT? 1= once or more a day, 2= once a week,
3= twice a month, 4= once a month, 5= Never.
5. What is your level? 1= Level [1-2], 2= Level [3 and above].
6. If you had to pick one issue that is the greatest barrier to using ICT, what would it be?
1= Time, 2= Technical support, 3= Cost, 4= Training, 5= Compensation,

6= Does not fit my program.

REFERENCES

[1] Condie, R. and B. Munro (2007) The Impact of ICT in schools- a landscape review(Report to Becta). ICT Survey and Research: ICT Resources used in Mathematics. (url: accessed 2007). Volume,

[2] Godwin, S. and R. Surtherland, Whole-class technology for learning mathematics: the case of functions and graphs. Education, Communication and Technology, 2004. 4(1): p. 131-152.

[3] Laborde, C., Integration of Technology in Design of Geometry Tasks with Cabri-Geometry. International Journal of Computers for Mathematical Learning., 2002. 6(3): p. 283-317.

[4] Ruthven, K., S. Hennessy, and S. Brindley, Teacher Representation of the successful use of computerbased tools and resources in secondary-school English, Mathematics and Science. Teaching and Teacher Education., 2004. 20(3): p. 259-275.

[5] Ruthven, K. and S. Hennessy, A practitioner modelof the use of computer-based tools and resources to support mathematics teaching and learning. Educational Studies in Mathematics., 2002. 49(1): p. 47-88.

[6] Pulkkinen, J., Cultural globalizationand integration of ICT in education, in Educational Technology: Opportunities and Challenges, K.E. Kumpulainen Editor. 2007, Finland: University of Oulu: Oulu. p. 13-23.

[7] Wood, D., Theory, training, and technology: . Education and Training, 1995. 37(1): p. 12-16.

[8] Pajo and C. Wallace, Barriers to the uptake of web-based technology by university teachers. Journal of Distance Education, 2001. 16: p. 70-84.

[9] Kumpulainen, K., ed. Educational technology: Opportunities and challenges. 2007, University of Oulu: Oulu, Finland.

[10] Usluel, Y.K., P. As_kar, and T. Bas_, A structural equation model for ICT usage in higher education. Educational Technology & Society, 2008. 11: p. 262-273.

[11] Carol, K., Linear and Non-linear Learning Blog. 2007.

[12] UNESCO, UNESCO portal on Higher Education Institutions:. 2009, Higher Education Institutions in Nigeria.

[13] Ololube, N.P. and D.E. Egbezor, Educational technology and flexible education in Nigeria: Meeting the need for effective teacher education. in S. Marshal, W. Kinthia and W. Taylor(Eds), Bridging the Knowledge Divide:. Educational technology for edvelopment, 2009: p. 391-413.

[14] Ololube, N.P. and D.E. Egbezor, Educational technology and Flexible education in Nigeria: Meeting the need for effective teacher education. in S. Marshal, W. Kinthia and W. Taylor (Eds), Bridging the Knowledge Divide:. Educational technology for develpoment, 2009: p. 391-413.

[15] Oye , N.D., N. A.Iahad, and N. Ab.Rahim, Adoption and Acceptance of ICT Innovations in Nigerian Public Universities. International Journal of Computer Science Engineering and Technology, 2011. 1(7): p. 434-443.

[16] Archibong, I.A. and D.O. Effiom, ICT in University Education: Usage and Challenges among Academic Staff. . Africa Research Review, 2009. 3(2): p. 404-414.

[17] Ijeoma, A.A., E.O. Joseph, and A. Franca, ICT Competence among Academic staff in Unversities in Cross Rivers State, Nigeria. . Computer and Information Science, 2010. 3(4): p. 109-115.

Appendix B

TABLE II. ICT IN EDUCATION (ICTEDU)

Correlations

		ICTEDU1	ICTEDU2	ICTEDU3	ICTEDU4	ICTEDU5
ICTEDU1	Pearson Correlation	1	.334**	.367**	.379**	.112
	Sig. (2-tailed)		.000	.000	.000	.172
	N	150	150	150	150	150
ICTEDU2	Pearson Correlation	.334**	1	.117	.081	.166*
	Sig. (2-tailed)	.000		.153	.324	.043
	N	150	150	150	150	150
ICTEDU3	Pearson Correlation	.367**	.117	1	.248**	.284**
	Sig. (2-tailed)	.000	.153		.002	.000
	N	150	150	150	150	150
ICTEDU4	Pearson Correlation	.379**	.081	.248**	1	.320**
	Sig. (2-tailed)	.000	.324	.002		.000
	N	150	150	150	150	150
ICTEDU5	Pearson Correlation	.112	.166*	.284**	.320**	1
	Sig. (2-tailed)	.172	.043	.000	.000	
	N	150	150	150	150	150

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

TABLE III. USE OF ICT (USEICT)

Correlations

		USEICT1	USEICT2	USEICT3	USEICT4	USEICT5
USEICT1	Pearson Correlation	1	.204*	.351**	.265**	.106
	Sig. (2-tailed)		.012	.000	.001	.195
	N	150	150	150	150	150

USEICT2	Pearson Correlation	.204*	1	.187*	.065	.147
	Sig. (2-tailed)	.012		.022	.427	.072
	N	150	150	150	150	150
USEICT3	Pearson Correlation	.351**	.187*	1	.206*	.105
	Sig. (2-tailed)	.000	.022		.012	.199
	N	150	150	150	150	150
USEICT4	Pearson Correlation	.265**	.065	.206*	1	.119
	Sig. (2-tailed)	.001	.427	.012		.147
	N	150	150	150	150	150
USEICT5	Pearson Correlation	.106	.147	.105	.119	1
	Sig. (2-tailed)	.195	.072	.199	.147	
	N	150	150	150	150	150

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

TABLE IV. AVAILABILITY OF ICT TOOLS (AICTT)

Correlations

		AICTT1	AICTT2	AICTT3	AICTT4	AICTT5
AICTT1	Pearson Correlation	1	.324**	.033	.091	-.023
	Sig. (2-tailed)		.000	.692	.268	.784
	N	150	150	150	150	150
AICTT2	Pearson Correlation	.324**	1	.163*	.150	.070
	Sig. (2-tailed)	.000		.047	.068	.397
	N	150	150	150	150	150
AICTT3	Pearson Correlation	.033	.163*	1	.316**	.121
	Sig. (2-tailed)	.692	.047		.000	.140
	N	150	150	150	150	150
AICTT4	Pearson Correlation	.091	.150	.316**	1	.292**
	Sig. (2-tailed)	.268	.068	.000		.000
	N	150	150	150	150	150
AICTT5	Pearson Correlation	-.023	.070	.121	.292**	1
	Sig. (2-tailed)	.784	.397	.140	.000	
	N	150	150	150	150	150

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

TABLE V. ICT AWARENESS (ICTA)

Correlations

		ICTA1	ICTA2	ICTA3	ICTA4	ICTA5
ICTA1	Pearson Correlation	1	.328**	.101	.006	.121
	Sig. (2-tailed)		.000	.219	.941	.141
	N	150	150	150	150	150
ICTA2	Pearson Correlation	.328**	1	.074	-.110	.092
	Sig. (2-tailed)	.000		.371	.179	.263
	N	150	150	150	150	150
ICTA3	Pearson Correlation	.101	.074	1	.351**	.261**
	Sig. (2-tailed)	.219	.371		.000	.001
	N	150	150	150	150	150
ICTA4	Pearson Correlation	.006	-.110	.351**	1	.192*
	Sig. (2-tailed)	.941	.179	.000		.018

	N	150	150	150	150	150
ICTA5	Pearson Correlation	.121	.092	.261**	.192*	1
	Sig. (2-tailed)	.141	.263	.001	.018	
	N	150	150	150	150	150

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

TABLE VI. DIFFICULTIES FACING ICT USAGE (DFICTU):

Correlations

		DFICTU1	DFICTU2	DFICTU3	DFICTU4	DFICTU5	DFICTU6	DFICTU7
DFICTU1	Pearson Correlation	1						
	Sig. (2-tailed)							
	N	150						
DFICTU2	Pearson Correlation	.476**	1					
	Sig. (2-tailed)	.000						
	N	150	150					
DFICTU3	Pearson Correlation	.257**	.307**	1				
	Sig. (2-tailed)	.002	.000					
	N	150	150	150				
DFICTU4	Pearson Correlation	.261**	.298**	.489**	1			
	Sig. (2-tailed)	.001	.000	.000				
	N	150	150	150	150			
DFICTU5	Pearson Correlation	.275**	.331**	.368**	.459**	1		
	Sig. (2-tailed)	.001	.000	.000	.000			
	N	150	150	150	150	150		
DFICTU6	Pearson Correlation	.140	.249**	.307**	.421**	.469**	1	
	Sig. (2-tailed)	.088	.002	.000	.000	.000		
	N	150	150	150	150	150	150	
DFICTU7	Pearson Correlation	.251**	.260**	.149	.132	.388**	.268**	1
	Sig. (2-tailed)	.002	.001	.068	.107	.000	.001	
	N	150	150	150	150	150	150	150

**. Correlation is significant at the 0.01 level (2-tailed).

TABLE VII. SUMMARY OF THE CORRELATION

Significant	R	P-values
ICTEDU1 to ICTEDU2	33.4	.000
ICTEDU1 to ICTEDU3	36.7	.000
ICTEDU1 to ICTEDU4	37.9	.000
ICTEDU4 to ICTEDU5	32.0	.000
USEICT1 to USEICT3	35.1	.000
USEICT1 to USEICT4	26.5	.001
DFICTU1 to DFICTU2	47.6	.000
DFICTU2 to DFICTU5	33.1	.000
DFICTU3 to DFICTU4	48.9	.000
DFICTU3 to DFICTU5	36.8	.000
DFICTU4 to DFICTU5	45.9	.000
DFICTU4 to DFICTU6	42.1	.000
DFICTU5 to DFICTU6	46.9	.000
DFICTU5 to DFICTU7	38.8	.000
AICTT1 to AICTT2	32.4	.000
AICTT3 to AICTT4	31.6	.000
ICTA1 to ICTA2	32.8	.000
ICTA3 to ICTA4	35.1	.000

Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance

Qasem A. Al-Radaideh
Department of Computer Information Systems,
Faculty of Information Technology and Computer Science
Yarmouk University, Irbid 21163, Jordan.

Eman Al Nagi
Department of Computer Science, Faculty of Information Technology, Philadelphia University, Amman, Jordan.

***Abstract*— Human capital is of a high concern for companies' management where their most interest is in hiring the highly qualified personnel which are expected to perform highly as well. Recently, there has been a growing interest in the data mining area, where the objective is the discovery of knowledge that is correct and of high benefit for users. In this paper, data mining techniques were utilized to build a classification model to predict the performance of employees. To build the classification model the CRISP-DM data mining methodology was adopted. Decision tree was the main data mining tool used to build the classification model, where several classification rules were generated. To validate the generated model, several experiments were conducted using real data collected from several companies. The model is intended to be used for predicting new applicants' performance.**

Keywords- Data Mining; Classification; Decision Tree; Job Performance.

I. INTRODUCTION

Human resource has become one of the main concerns of managers in almost all types of businesses which include private companies, educational institutions and governmental organizations. Business Organizations are really interested to settle plans for correctly selecting proper employees. After hiring employees, managements become concerned about the performance of these employees were management build evaluation systems in an attempt to preserve the good-performers of employees (Chein and Chen, 2006).

Data mining is a young and promising field of information and knowledge discovery (Han et al., 2011). It started to be an interest target for information industry, because of the existence of huge data containing large amounts of hidden knowledge. With data mining techniques, such knowledge can be extracted and accessed transforming the databases tasks from storing and retrieval to learning and extracting knowledge.

Data miming consists of a set of techniques that can be used to extract relevant and interesting knowledge from data. Data mining has several tasks such as association rule mining, classification and prediction, and clustering. Classification techniques are supervised learning techniques that classify data item into predefined class label. It is one of the most useful techniques in data mining to build classification models from an input data set. The used classification techniques commonly build models that are used to predict future data trends. There are several algorithms for data classification such as decision tree and Naïve Bayes classifiers. With classification, the generated model will be able to predict a class for given data depending on previously learned information from historical data.

Decision tree is one of the most used techniques, since it creates the decision tree from the data given using simple equations depending mainly on calculation of the gain ratio, which gives automatically some sort of weights to attributes used, and the researcher can implicitly recognize the most effective attributes on the predicted target. As a result of this technique, a decision tree would be built with classification rules generated from it (Han et al., 2011).

Naïve Bayes classifier is another classification technique that is used to predict a target class. It depends in its calculations on probabilities, namely Bayesian theorem. Because of this use, results from this classifier are more accurate and effective, and more sensitive to new data added to the dataset (Han et al., 2011).

Several studies used data mining for extracting rules and predicting certain behaviors in several areas of science, information technology, human resources, education, biology and medicine.

For example, Beikzadeh and Delavari (2004) used data mining techniques for suggesting enhancements on higher educational systems. Al-Radaideh et al. (2006) also used data mining techniques to predict university students' performance. Many medical researchers, on the other hand, used data mining techniques for clinical extraction units using the enormous patients data files and histories, Lavrac (1999) was one of such researchers. Mullins et al. (2006) also worked on patients' data to extract disease association rules using unsupervised methods.

Karatepe et al. (2006) defined the performance of a frontline employee, as his/her productivity comparing with his/her peers. Schwab (1991), on the other hand, described the performance of university teachers included in his study, as the number of researches cited or published. In general, performance is usually measured by the units produced by the employee in his/her job within the given period of time.

Researchers like Chein and Chen (2006) have worked on the improvement of employee selection, by building a model, using data mining techniques, to predict the performance of newly applicants. Depending on attributes selected from their CVs, job applications and interviews. Their performance could be predicted to be a base for decision makers to take their decisions about either employing these applicants or not.

Previous studies specified several attributes affecting the employee performance. Some of these attributes are personal characteristics, others are educational and finally professional attributes were also considered. Chein and Chen (2006) used several attributes to predict the employee performance. They specified age, gender, marital status, experience, education, major subjects and school tires as potential factors that might affect the performance. Then they excluded age, gender and marital status, so that no discrimination would exist in the process of personal selection. As a result for their study, they found that employee performance is highly affected by education degree, the school tire, and the job experience.

Kahya (2007) also searched on certain factors that affect the job performance. The researcher reviewed previous studies, describing the effect of experience, salary, education, working conditions and job satisfaction on the performance. As a result of the research, it has been found that several factors affected the employee's performance. The position or grade of the employee in the company was of high positive effect on his/her performance. Working conditions and environment, on the other hand, had shown both positive and negative relationship on performance. Highly educated and qualified employees showed dissatisfaction of bad working conditions and thus affected their performance negatively. Employees of low qualifications, on the other hand, showed high performance in spite of the bad conditions. In addition, experience showed positive relationship in most cases, while education did not yield clear relationship with the performance.

In their study, Salleh et al. (2011) have tested the influence of motivation on job performance for state government employees in Malaysia. The study showed a positive relationship between affiliation motivation and job performance. As people with higher affiliation motivation and strong interpersonal relationships with colleagues and managers tend to perform much better in their jobs.

Jantan et al. (2010) had discussed in their paper Human Recourses (HR) system architecture to forecast an applicant's talent based on information filled in the HR application and past experience, using Data Mining techniques. The goal of the paper was to find a way to talent prediction in Malaysian higher institutions. So, they have specified certain factors to be considered as attributes of their system, such as, professional qualification, training and social obligation. Then, several data mining techniques (hybrid) where applied to find the prediction rules. ANN, Decision Tree and Rough Set Theory are examples of the selected techniques.

The same authors, Jantan et al. (2010b) have used decision tree C4.5 classification algorithm to predict human talent in HRM, by generating classification rules for the historical HR records, and testing them on unseen data to calculate accuracy. They intend to use these rules in creating a DSS system that can be used by managements to predict employees' performance and potential promotions.

Generally, this paper is a preliminary attempt to use data mining concepts, particularly classification, to help supporting the human resources directors and decision makers by evaluating employees' data to study the main attributes that may affect the employees' performance. The paper applied the data mining concepts to develop a model for supporting the prediction of the employees' performance. In section 2, a complete description of the study is presented, specifying the methodology, the results, discussion of the results.

II. Building The Classification Model

The main objective of the proposed methodology is to build the classification model that tests certain attributes that may affect job performance. To accomplish this, the CRISP-DM methodology (Cross Industry Standard Process for Data Mining) (CRISP-DM, 2007) was used to build a classification model. It consists of five steps which include: Business understanding, data understanding, data preparation, modeling, evaluation and deployment.

A. Data Classification Preliminaries

In general, data classification is a two-step process. In the first step, which is called the learning step, a model that describes a predetermined set of classes or concepts is built by analyzing a set of training database instances. Each instance is assumed to belong to a predefined class. In the second step, the model is tested using a different data set that is used to estimate the classification accuracy of the model. If the accuracy of the model is considered acceptable, the model can be used to classify future data instances for which the class label is not known. At the end, the model acts as a classifier in the decision making process. There are several techniques that can be used for classification such as decision tree, Bayesian methods, rule based algorithms, and Neural Networks.

Decision tree classifiers are quite popular techniques because the construction of tree does not require any domain expert knowledge or parameter setting, and is appropriate for exploratory knowledge discovery. Decision tree can produce a model with rules that are human-readable and interpretable. Decision Tree has the advantages of easy interpretation and understanding for decision makers to compare with their domain knowledge for validation and justify their decision. Some of decision tree classifiers are C4.5/C5.0/J4.8, NBTree, and others.

The C4.5 technique is one of the decision tree families that can produce both decision tree and rule-sets; and construct a tree for the purpose of improving prediction accuracy. The C4.5 / C5.0 / J48 classifier is among the most popular and powerful decision tree classifiers. C4.5 creates an initial tree using the divide-and-conquer algorithm. The full description of the algorithm can be found in any data mining or machine learning books such as (Han et al., 2011) and (Witten et al., 2011).

WEKA toolkit (Witten et al., 2011) is a widely used toolkit for machine learning and data mining originally developed at the University of Waikato in New Zealand. It contains a large collection of state-of-the-art machine learning and data mining algorithms written in Java. WEKA contains tools for regression, classification, clustering, association rules, visualization, and data pre-processing. WEKA has become very popular with academic and industrial researchers, and is also widely used for teaching purposes. WEKA toolkit package has its own version known as J48. J48 is an optimized implementation of C4.5 rev. 8.

B. Data Collection Process and Data Understanding

When the idea of the study came in to mind, it was intended to apply a classification model for predicting performance depending on a dataset from a certain IT company. So that any other factors regarding the working environment, conditions, management and colleagues would have similar effect on all employees, and so the effect of collected attributes would be more apparent and easier to classify. Unfortunately, data collected from the first IT Company was not enough to be the base of such a classification model. In this case, another attempt was taken to collect another group of data from another IT company. In order to collect the required data, a questionnaire was prepared and distributed either by email or manually to the employees of both companies. Then, it was further distributed on the internet, to be filled by employees working in any IT company. The questionnaire was filled by 130 employees, 37 from the first IT Company, 38 from the second one, and the rest from several other companies using the internet questionnaire.

Several attributes have been asked for in the questionnaire that might predict the performance class. The list of the collected attributes is presented in Table 1.

C. Data Preparation

After the questionnaires were collected, the process of preparing the data was accomplished. First, the information in the questionnaires has been transferred to Excel sheets. Then, the types of data has been reviewed and modified. Some attributes like experience years and service period, have been entered in continuous values. So, they were modified to be illustrated by ranges. Other attributes like specialization, job title and rank, have been generalized to include fewer discrete values than they already have. For example, in specialization, there were values like electrical engineering and computer engineering, they have been considered as one value, engineering. MIS and CIS were considered as IT, and so on.

These files are prepared and converted to (arff) format to be compatible with the WEKA data mining toolkit (Witten et al., 2011), which is used in building the model.

As mentioned previously, the data has been divided into three datasets. The first one includes the data of the first IT company employees. The second includes the data of the second IT company employees. The third one includes all data collected from the three sources. Each dataset has two arff files containing its data, with the class attribute (performance). Each of these datasets was used in a separate experiment.

III. Modeling And Experiments

After the data has been prepared, the classification models have been built. Using the decision tree technique, a tree has been built for each of these experiments. In this technique, the gain ratio measure is used to indicate the weight of effectiveness of each attribute on the tested class, and accordingly the ordering of tree nodes is specified. The results are discussed in the following sections.

Referring to the discussion of earlier studies, and as described in Table (1), a group of attributes has been selected to be tested against their effectiveness on the employee performance.

These attributes consist of (1) Personal information such as: age, gender, marital status and number of kids (if any), (2) Education information such as: university type, general specialization, degree and grade, (3) Professional information such as: number of experience years, number of previous companies worked for, job title, rank, service period in the current company, salary, finding the working conditions uncomfortable and dissatisfaction of salary or rank. These attributes were used to predict the employee performance to be *accomplished*, *exceed* or *far-exceed*.

A. First Experiment (E1): Using the whole dataset (130 instances)

Three classification techniques have been applied on the dataset on hand to build the classification model. The techniques are: The decision tree with two versions, ID3 and C4.5 (J4.8 in WEKA), and Naïve Bayes classifier. For each experiment, accuracy was evaluated using 10-folds cross-validation, and hold-out method. Table 2 displays the accuracy percentages for each of these techniques.

The tree generated by ID3 algorithm was very deep, since it started by attribute JobTitle, which has 20 values.

The JobTitle has the maximum gain ratio, which made it the starting node and most effective attribute. Other attributes participated in the decision tree were UnivType, SalRange, ExpYears, Grade, Age, MStatue, Gender, GSpecial and Rank. Other attributes such as: PrevCo, Nkids, uncomworkcond, dissatsalrank and degree appeared in other parts of the decision tree.

TABLE I. FULL DESCRIPTION OF ATTRIBUTES USED FOR PREDICTING THE PERFORMANCE CLASS

Attribute	Description	Possible Values
*Age	Employee's Age	a, b, c, d, e
Gender	Employee's Gender	Male, Female
MStatue	Employee's Marital Status	Single, Married with kids, Married without kids, Other
NKids	No. of Kids	0, 1, 2, 3, 4

UnivType	The type of university of graduation	Public, Private
GSpecial	General Specialization	Business, IT, English Literature, Engineering, CS, Other
Degree	Employee Education Degree	Diploma, Bachelor, High Diploma, Master, PhD
Grade	Employee Graduation Grade	Excellent, Very good, Good, Acceptable, Other
Country	Country of University	This attribute was eliminated, since majority of the values were Jordan
**Expyears	No. of Working Experience Years	a, b, c, d, e
PrevCo	No. of Previous Companies the employee worked for	0,1,2,3,4,5
JobTitle	Employee's Job Title in the current company	Developer, Officer, QA, Data Entry, System Administrator, Office Manager, Technical Writer, Technical Manager, Software Engineering, Accountant, Infrastructure Engineer, Department Manager, software Architect , Analyst, Designer, Trainer, PM, Consultant, Customer Support, GM.
Rank	Employee's Rank in the current company	Junior, Senior, Team Leader, Manager, Architect
***ServPeriod	Service Period in the current company (in years)	a, b, c, d
Working hours	No of working hours	Full Time, Part Time, Other. This attribute was eliminated, since most instances has Full time value.
****SalRange	Employee's Range of Salary	a, b, c, d, e
UncomWorkcond	Working in uncomfortable conditions (in employee's perspective)	Yes, No
Dissatsalrank	Existence of dissatisfaction in either salary or rank	Yes, No
Performance	Employee's performance, either as informed or predicted. This is a class	Accomplish, Exceed, Far Exceed

Notes: * Age range: a) Under 25 b) 25-29 c) 30-34 d) 35-40 e) Over 40
** Experience Years Range: a) None b) Less than a year c) 1-5 d) More than 5 – 10 e) More than 10
*** Service Period Range: a) Less than 1 year b) 1-5 c) More than 5-10 d) More then 10
**** Salary Range: a) Less than 250 b) 250-499 c) 500-999 d) 1000-2000 e) Over 2000

TABLE II. ACCURACY PERCENTAGES FOR PREDICTING PERFORMANCE IN E1

Method	10-Fold Cross Validation	Hold-out (60%)
ID3	36.9%	36.5%
C4.5 (J4.8)	42.3%	48.1%
Naïve Bayes	40.7%	44.2%

The tree indicated that all these attributes have some sort of effect on the employee performance, but the most affective attributes were: JobTitle, UnivType and Age. Other hints could be extracted from the tree indicates that young employees have better performance than older ones. Wherever Gender is taken into consideration, Male employees have higher performance than Female. Moreover, employees with higher graduation grades have higher performance. Finally, employees with higher ranks have less performance giving indication that managers work less than less ranked employees.

The tree generated using the C4.5 algorithm also indicated that the JobTitle attribute is the most affective attribute. The Naïve Bayes classifier does not show the weights of each attribute included in the classification, but it has been used to be compared with the results generated from ID3 and C4.5 as was shown previously in Table 2. It can be noticed that the accuracy percentage ranges from approximately 36% to 45%, which are low percentages.

B. Second Experiment (E2): Using the dataset gathered from the first IT company (37 instances)

By using the same approach as in E1, Table 3 shows the prediction accuracy for each algorithm applied to this dataset.

TABLE III. ACCURACY PERCENTAGES FOR PREDICTING PERFORMANCE IN E2

Algorithm	10-Fold Cross Validation	Hold-out (60%)
ID3	37.8 %	26. 7%
C4.5 (J4.8)	48.6%	53.3%
Naïve Bayes	37.8%	46. 7%

Decision tree built by ID3 algorithm, showed different trend than the one generated for E1, since in this tree, the starting node was PrevCo, and the attributes with highest gain ratio were JobTitle, GSpecial, NKids, and Age, while the less effective attributes were MStatus, Uncomworkcond, UnivType, Grade and Dissatsalwork. The decision tree built using the C4.5 algorithm was so much pruned to consist of only three attributes; with PrevCo as the starting node, and Dissatsalrank and Grade as other attributes.

For the PrevCo attribute, it can be noticed that the employee performance varies from exceed and far exceed if the PrevCo is less than 3, then becomes accomplish, and raises again when the PrevCo more than 3.

In addition, when dissatsalrank is Yes, the employee performs was Accomplish, while the No value indicates the satisfaction of an employee to perform Exceed. This indicates a normal reaction of dissatisfaction in the salary or rank.

Grade attribute has an interesting result, which indicates that employees with grade good, is far exceed in the performance, while other grades are only exceed. This could be indicating that graduates with high grades do not necessarily indicate good productive employees.

As an example of the generated C4.5 tree, Fig. 1 shows the tree generated for E2, and Table 4 shows the generated classification rules with the number of instances that support each rule.

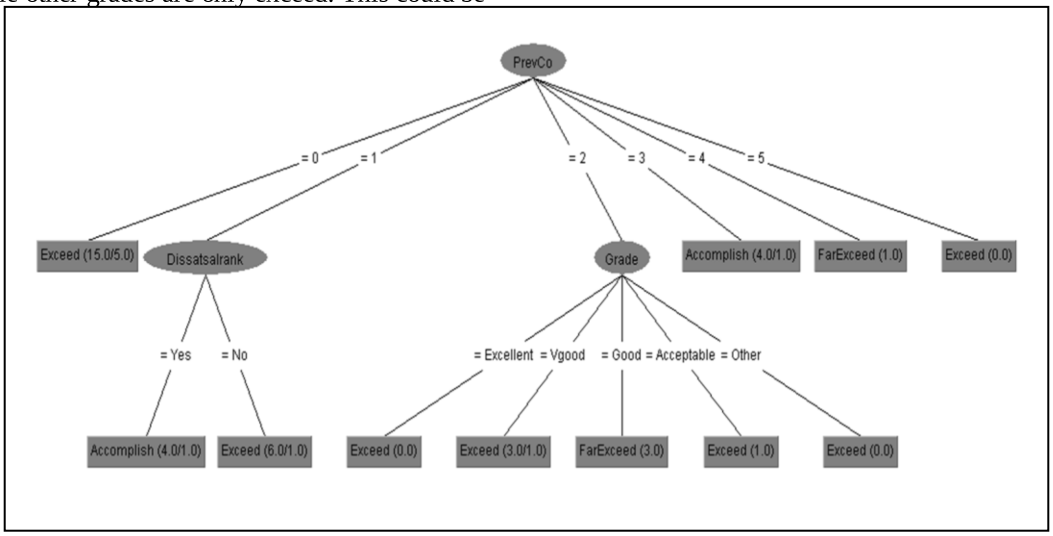


Figure 1. A decision tree generated by C4.5 algorithm for E2 for predicting performance

TABLE IV. CLASSIFICATION RULES GENERATED BY C4.5 ALGORITHM IN E2 FOR PREDICTING PERFORMANCE

Rule #	Rule Antecedent	Performance Decision	# of Instances
1	IF PrevCo = 0 ==>	Exceed	15
2	IF PrevCo = 1 & Dissatsalrank = Yes ==>	Accomplish	4
3	IF PrevCo = 1 & Dissatsalrank = No ==>	Exceed	6
4	IF PrevCo = 2 & Grade <> Good ==>	Exceed	4
5	IF PrevCo = 2 & Grade = Good ==>	FarExceed	3
6	IF PrevCo = 3 ==>	Accomplish	4
7	IF PrevCo = 4 ==>	FarExceed	1

C. Third Experiment (E3): Using the dataset gathered from the second IT company (38 instances)

Table 5 shows the accuracy percentages resulted from applying the algorithms of ID3, C4.5 and Naïve Bayes on the dataset of the second IT Company. Note that the accuracy percentages were increased in this experiment.

The decision tree built using the ID3 algorithm for this experiment has started with JobTitle, as in E1. And then more weight has been given to attributes as GSpecial, Rank, Degree and ExpYears over attributes like SalRange, Grade and PrevCo.

TABLE V. ACCURACY PERCENTAGES FOR PREDICTING PERFORMANCE IN E3

Algorithm	10-Fold Cross Validation	Hold-out (60%)
ID3	50%	43.7%
C4.5 (J4.8)	60.5%	56.2%
Naïve Bayes	65.8%	68.7%

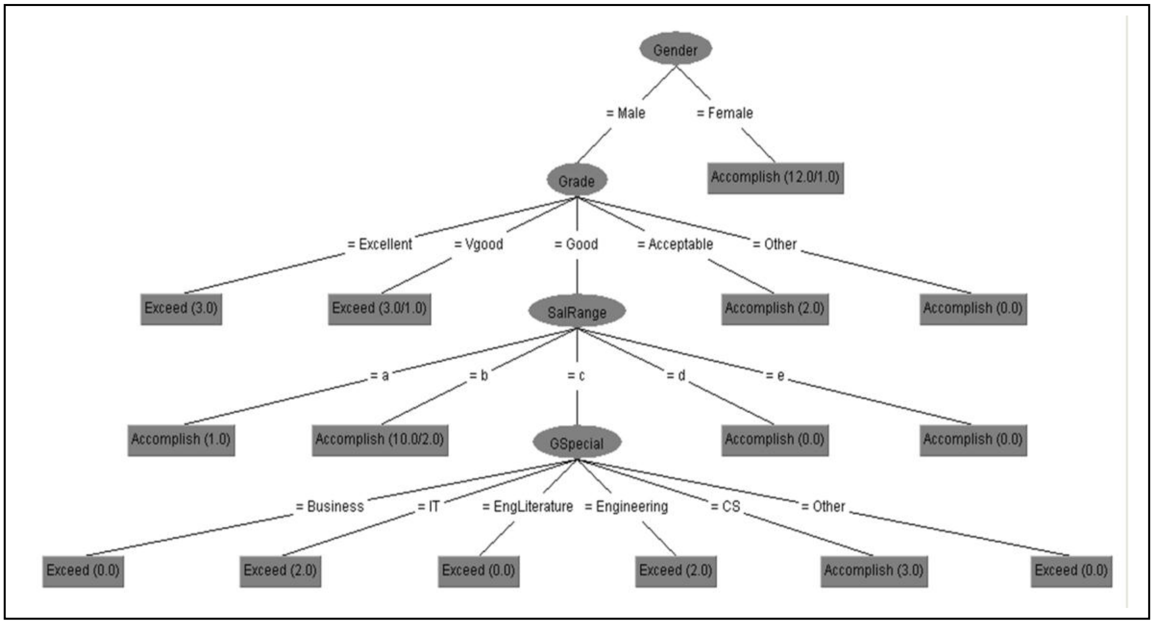


Figure 2. Decision Tree resulted from C4.5 algorithm for E3 to predict performance

TABLE VI. CLASSIFICATION RULES GENERATED BY C4.5 ALGORITHM IN E3 FOR PREDICTING PERFORMANCE

Rule #	Rule Antecedent	Performance Decision	# of Instances
1	IF Gender = Male & Grade = Excellent ==>	Exceed	3
2	IF Gender = Male & Grade = Vgood ==>	Exceed	3
3	IF Gender = Male & Grade = Good & SalRange = a ==> Performance =	Accomplish	1
4	IF Gender = Male & Grade = Good & SalRange = b ==> Performance =	Accomplish	10
5	IF Gender = Male & Grade = Good & SalRange = c & GSpecial = IT ==>	Exceed	2
6	IF Gender = Male & Grade = Good & SalRange = c & GSpecial = Engineering ==>	Exceed	2
7	IF Gender = Male & Grade = Good & SalRange = c & GSpecial = CS ==>	Accomplish	3
8	IF Gender = Male & Grade = Acceptable ==>	Accomplish	2
9	IF Gender = Female ==>	Accomplish	12

While using the C4.5 algorithm, the gain ratio of the Gender was the highest to start the tree with, and then comes Grade, SalRange and GSpecial. Fig. 2 shows the tree generated by WEKA for E3 and Table 6 shows the generated classification rules.

IV. RESULTS AND DISCUSSION

The study has found that several factors might have a great effect on employee performance. One of the most effective factors is the job title.

The trend of effectiveness of the job title is not much clear in the results, since there are about 20 job titles studied, but it can be related to the type of job complexity and the responsibilities related to the title. High responsibilities sometimes affect the employee's motivation and therefore performance in a positive way.

The university type attribute, in the three experiments, has positively affected the performance when the employee was graduated from a public university rather than a private one. This could be due to the fact that public universities accept, in most cases, students with high grades in high school comparing to private universities.

Other educational factors like degree and grade have slightly affected the performance, but not with clear trend, it might depend on other factors depending on the employee personality, which are not considered in this study.

Such, personality factors can be recognized by decision makers in interviews, so that they can complete their knowledge about the applicant. The university general specialization has a very close effect to performance as the job title. This could be due to the relationship between these two factors.

Some personal information like age, marital status and gender also affects the performance. Nevertheless, the age has not clear effect on the performance, since sometimes the performance increases with age, which adds the experience factor, other times, it decreases showing the highest motivation with the younger employees. Marital status, on the other hand, is clearer in its effect, since single employees in all experiments have shown better performance from married employees and even much better than married with kids employees.

But, surprisingly, in experiment E2, a strange trend appeared regarding number of kids, which indicated that the higher number of kids leads to a higher performance. This could be a coincidence outlier, since E2 dataset is not large enough to confirm this rule. Gender on the other hand, has no effect at all on experiments E1 and E2, since the female proportion in both datasets is not significant. But in experiment E3, it indicated a higher performance for male employees than female.

Several professional factors also appeared to affect the performance. Salary, is one of the most positive factors on performance, this effects has been shown in experiments E1 and E3, while in E2 it was not significant. Number of previous companies in E1, showed both positive and negative relationship with the performance. This could be due to newly working employees who do not have experience working in other companies; they do their best to obtain better positions. On the other hand, employees worked in some previous companies may have much experience that would influence their performance. In E2 and E3 experiments only the positive relationship is observed. As for experience years, it affected the performance positively in E1 and E2, while in E3 it was not of much significance.

The Rank attribute has shown an interesting influence on performance, especially in E1 and E3. As in E2 it was not included as an effective factor. It was noticed in E1 and E3 experiments, the performance of senior employees is more than juniors. This is natural because of the experience of the employees that affect the Rank. But, surprisingly, team leaders and managers tend to have less performance than senior employees. This confirms the claims against highly positioned employees, that they do not work much.

Finally, job satisfaction and comfortable working environment has a slight effect on performance. For E3 they were not included as effective factors; while in E1 and E2 they were considered as low weighted effective factors on performance. This could be interpreted that the company in E3 has a more satisfactory conditions than the company in E2.

As a final remark on the accuracy of the classification models built for the three experiments, it can be noticed that for the different algorithms used, the classification accuracy was much more in experiments E2 and E3 than in E1. This might be because of the different companies of employees, included in E1, which created different factors affecting the classes in the experiments. While in E2 and E3, in spite of their small datasets, but the employees under study has the same working conditions, working environment, management and colleagues that made the study more focused on the measurable attributes on hand.

V. Conclusion And Future Work

This paper has concentrated on the possibility of building a classification model for predicting the employees' performance. On working on performance, many attributes have been tested, and some of them are found effective on the performance prediction. The job title was the strongest attribute, then the university type, with slight effect of degree and grade.

The age attribute did not show any clear effect while the marital status and gender have shown some effect in some of the experiments for predicting the performance. Salary, number of pervious companies, experiment years and job satisfaction, each had a degree of effect on predicting the performance.

For companies managements and human resources departments, this model, or an enhanced one, can be used in predicting the newly applicant personnel performance. Several actions can be taken in this case to avoid any risk related to hiring poorly performed employee.

As future work, it is recommended to collect more proper data from several companies. Databases for current employees and even previous ones can be used, to have a correct performance rate for each one of them.

When the appropriate model is generated, software could be developed to be used by the HR including the rules generated for predicting performance of employees.

References

[1] Allen, M.W., Armstrong, D.J., Reid, M.F., Riemenschneider, C.K. (2009). "IT Employee Retention: Employee Expectations and Workplace Environments", SIGMIS-CPR'09 , May 2009, Limerick, Ireland.

[2] Al-Radaideh, Q. A., Al-Shawakfa, E.M., Al-Najjar, M.I. (2006). "Mining Student Data Using Decision Trees", International Arab Conference on Information Technology (ACIT 2006), Dec 2006, Jordan.

[3] Chein, C., Chen, L. (2006) "Data mining to improve personnel selection and enhance human capital: A case study in high technology industry", Expert Systems with Applications, In Press.

[4] Cho, S., Johanson, M.M., Guchait, P. (2009). "Employees intent to leave: A comparison of determinants of intent to leave versus intent to stay", International Journal of Hospitality Management, 28, pp374-381.

[5] CRISP-DM, (2007). Cross Industry Standard Process for Data Mining: Process Model. http://www.crisp-dm.org/process/index.htm. Accessed 10th May 2007.

[6] Delavari, N., PHON-AMNUAISUK S., (2008). Data Mining Application in Higher Learning

[7] Dreher, G.F. (1982) "The Role of Performance in the Turnover Process", Academy of Management Journal, 25(1), pp. 137-147.

[8] Han, J., Kamber, M., Jian P. (2011). Data Mining Concepts and Techniques. San Francisco, CA: Morgan Kaufmann Publishers.

[9] Ibrahim, M.E., Al-Sejini, S., Al-Qassimi, O.A. (2004). "Job Satisfaction and Performance of Government Employees in UAE", Journal of Management Research, 4(1), pp. 1-12.

[10] Jantan, H., Hamdan, A.R. and Othman, Z.A. (2010a) "Knowledge Discovery Techniques for Talent Forecasting in Human Resource Application", International Journal of Humanities and Social Science, 5(11), pp. 694-702.

[11] Jantan, H., Hamdan, A.R. and Othman, Z.A. (2010b). "Human Talent Prediction in HRM using C4.5 Classification Algorithm", International Journal on Computer Science and Engineering, 2(08-2010), pp. 2526-2534.

[12] Karatepe, O.M., Uludag, O., Menevis, I., Hadzimehmedagic, L., Baddar, L. (2006). "The Effects of Selected Individual Characteristics on Frontline Employee Performance and Job Satisfaction", Tourism Management, 27 (2006), pp. 547-560.

[13] Kayha, E. (2007). "The Effects of Job Characteristics and Working Conditions on Job Performance", International Journal of Industrial Ergonomics, In Press.

[14] Khilji, S., Wang, X. (2006). "New evidence in an old debate: Investigating the relationship between HR satisfaction and turnover", International Business Review, In Press.

[15] Lavrac, N. (1999). "Selected Techniques for Data Mining in Medicine", Artificial Intelligence in Medicine, 16(1999), pp. 3-23.

[16] Mullins, I., Siadaty, M., Lyman, J., Scully, K., Garrett, C., Millar, W., Mullar, R., Robson, B., Apte, C., Weiss, S., Rigoutsos, I., Platt, D., Cohen, S., Knaus, W. (2006). "Data Mining and Clinical Data Repositories: Insights from 667,000 Patient Data Set", Computers in Biology and Medicine, 36, pp. 1351-1377.

[17] Russ, F.A., McNeilly, K.M. (1995). "Links among Satisfaction, Commitment, and Turnover Intentions: The Moderating Effect of Experience, Gender and Performance", Journal of Business Research, 34, pp. 57-65.

[18] Salleh, F., Dzulkifli, Z., Abdullah, W.A. and Yaakob, N. (2011). "The Effect of Motivation on Job Performance of State Government Employees in Malaysia", International Journal of Humanities and Social Science, 1(4), pp. 147-154.

[19] Schwab, D. (1991). "Contextual variables in employee performance-turnover relationships", Academy of Management Journal, 34 (4), pp. 966-975.

[20] Sexton, R.S., McMurtrey, S., Michalopoulos, J.O., Smith, A.M. (2005). "Employee Turnover: A Neural Network Solution", Computer & Operations Research, 32, pp. 2635-2651.

[21] Witten I. Frank E., and Hall M. (2011). Data Mining: Practical Machine Learning Tools and Techniques 3rd Edition, Morgan Kaufmann Publishers.

AUTHORS PROFILE

Qasem A. Al-Radaideh is an Assistant Professor of Computer Information Systems at Yarmouk University. He recieved his Ph.D. in Data Mining field from the University Putra Malaysia in 2005. His research interests include: Data Mining and Knowledge Discovery in Database, Natural Language Processing, Arabic Language Computation, Information Retrieval, and Websites evaluation. He has several publications in the areas of Data Mining and Arabic Language Computation. He is curently the national advisor of Microsoft students partners program (MSPs) and the MS-Tech Clubs in Jordan.

Eman Al Nagi is a lecturer in the department of computer science at the Philadelphia University in Jordan. She received her MSc. In Information Technology Management from the University of Sunderland, UK on 2009. She has about 6 years experience in Software development and design

Web 2.0 Technologies and Social Networking Security Fears in Enterprises

Fernando Almeida
INESC TEC, Faculty of Engineering, University of Porto
Rua Dr. Roberto Frias, 378, 4200-465 Porto, Portugal

***Abstract*— Web 2.0 systems have drawn the attention of corporation, many of which now seek to adopt Web 2.0 technologies and transfer its benefits to their organizations. However, with the number of different social networking platforms appearing, privacy and security continuously has to be taken into account and looked at from different perspectives. This paper presents the most common security risks faced by the major Web 2.0 applications. Additionally, it introduces the most relevant paths and best practices to avoid these identified security risks in a corporate environment.**

***Keywords-Web 2.0; security, social networking; management risks*.**

I. Introduction

Applications are the lifeblood of today's organization as they allow employers to perform crucial business tasks. When granted access to enterprise networks and the Internet, applications can enable sharing of information within workgroups, throughout an enterprise and externally with partners and customers. Until recent years, when applications were launched only from desktop computers and servers inside the corporate network, data security policies were relatively easy to enforce. However, today's organizations are grappling with a new generation of security threats. Consumer-driven technology has unleashed a new wave of Internet-based applications that can easily penetrate and circumvent traditional network security barriers.

The Web 2.0 introduces the idea of a Web as a platform. The concept was such that instead of thinking of the Web as a place where browsers viewed data through small windows on the readers' screens, the Web was actually the platform that allowed people to get things done. Currently this initial concept has gained a new dimension and is really starting to mean a combination of the technology allowing customers to interact with the information [1].

Social-networking Web sites, such as Facebook and MySpace, now attract more than 100 million visitors a month [2]. As the popularity of Web 2.0 has grown, companies have noted the intense consumer engagement and creativity surrounding these technologies. Many organizations, keen to harness Web 2.0 internally, are experimenting with the tools or deploying them on a trial basis.

Reference [3] admits that Web 2.0 could have a more far-reaching organizational impact than technologies adopted in the 1990s (e.g., enterprise resource planning (ERP), customer relationship management (CRM), and the supply chain management (SCM)). The organizational of these new collaborative platforms are illustrated in figure 1. The latest Web tools have a strong bottom-up element and engage a broad base of workers. They also demand a mind-set different from that of earlier IT programs, which were instituted primarily by edicts from senior managers.

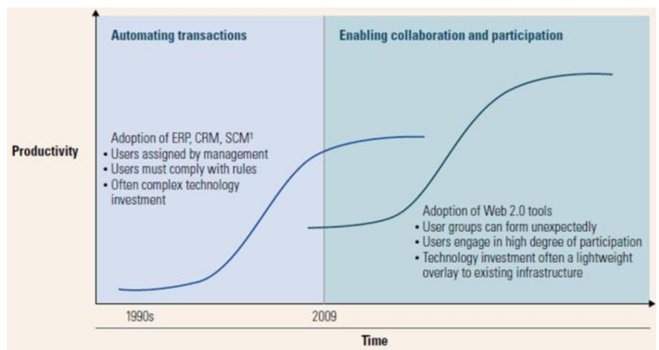


Figure 1. Adoption of corporate technologies. [4]

Web 2.0 covers a wide range of technologies. The most widely used are blogs, wikis, podcasts, information tagging, prediction markets, and social networks. A short description of these technologies potentialities is given in table 1.

TABLE 1. Web 2.0 Technologies [4]

Web 2.0 technologies	Description	Category of technology
Wikis, shared workspaces	Facilitates co-creation of contents across large and distributed set of participants.	Broad collaboration.
Blogs, podcasts, videocasts	Offers individuals a way to communicate and share information with other people.	Broad communication.
Prediction markets, polling	Harnesses the power of community and generates a collectively derived answer.	Collective estimation.
Tagging, user tracking, ratings, RSS	Add additional information to primary content to prioritize information.	Metadata creation.
Social networking, network mapping	Leverages connections between people to offer new applications	Social graphing.

New technologies are constantly appearing as the Internet continues to evolve. What distinguishes them from previous technologies is the high degree of participation they require to be effective [5]. Unlike ERP and CRM, where most users either simply process information in the form of reports or use the technology to execute transactions (such as issuing payments or entering customer orders), Web 2.0 technologies are interactive and require users to generate new information and content or to edit the work of other participants.

These new Internet-based communications tools such as Facebook, Twitter and Skype have already achieved widespread penetration inside organizations [6]. Inevitably, these new Internet-based technologies and applications have spawned a new set of challenges for enterprises seeking to secure their networks against malicious threats and data loss. Allowing employees to access Web 2.0 applications has made enforcing data security policies a far more complex problem. Even worse, many businesses have no way to detect, much less control these new applications, increasing the potential for intentional or accidental misappropriating of confidential information.

II. Web 2.0 Adoption In Organizations

Web 2.0 solutions are used for a variety of business purposes. According to survey study conducted by Gartner [6], about half of the organizations employ Web 2.0 solutions for IT functions, and roughly a third of organizations use them for marketing, sales or customer service. One in five organizations reported using Web 2.0 for public relations or human resources, particularly in the recruitment field [7]. The same study also establishes that by 2014, social networking services will replace e-mail as the primary vehicle for interpersonal communications, for 20 percent of the business users.

Another study conducted by [8] on the end of 2010 reports that Web 2.0 continues to grow, showing significant increases in the percentage of companies using social networking (40 percent) and blogs (38 percent). Furthermore, this survey shows that the number of employees using the dozen Web 2.0 technologies continues to increase. On the same way, nearly two-thirds of respondents at companies using Web 2.0 say they will increase future investments in these technologies, compared with just over half in 2009 [8].

The most common business benefits from using Web 2.0 based on the literature revision includes the increasing speed of access to knowledge, reducing communication costs, increasing effectiveness of marketing, increasing customer satisfaction, increase brand reputation, increasing speed of access to knowledge and reducing communication costs [9] [10]. Different types of networked organizations can achieve different benefits, namely:

- Internally networked organizations – some companies are achieving benefits from using Web 2.0 primarily within their own corporate walls. In this case, Web 2.0 is integrated tightly into their workflows and promotes significantly more flexible processes;
- Externally networked organizations – other companies achieve substantial benefits from interactions that spread beyond corporate borders by using Web 2.0 technologies to interact customers and business partners;
- Fully networked organizations – finally, some companies use Web 2.0 in revolutionary ways. They derive very high levels of benefits from Web 2.0's widespread use, involving employees, customers, and business partners.

III. Web 2.0 Security Risks

The collaborative, interactive nature of Web 2.0 has great appeal for business from a marketing and productivity point of view. Companies of all sizes and vertical markets are currently taking full advantage of social networking sites such as Facebook, Twitter and LinkedIn to connect with colleagues, peers and customers. In fact, not only these technologies are useful, but companies that don't adapt could well find themselves left behind the social revolution [11].

Companies are leveraging these sites for more than just communicating. Through Web 2.0 and social networking areas, enterprises are exchanging media, sharing documents, distributing and receiving resumes, developing and sharing custom applications, leveraging open source solutions, and providing forums for customers and partners [12].

While all this interactivity is exciting and motivating, there is an enterprise triple threat found in Web 2.0: losses in productivity, vulnerabilities to data leaks, and inherent increased security risks.

There are certain organizations that embrace Web 2.0 usage by employees, but the majority of them follow a different approach. Reference [6] shows that eighty-one percent of organizations restrict the use of at least one Web 2.0 tool because they are concerned about security. Therefore, organizations restrict social media usage through policy, technology and controlling the use of user-owned devices. While blocking access to social media provides better security, it is widely accepted that it is never feasible nor sustainable in the face of emerging use in the 21st century. Instead, we are living in a future where organizations must plan and design environments with less control of employee activities.

As referred previously in this paper, the primary concern that organizations have about employee usage of Web 2.0 technologies is security. Figure 2 illustrates the top concerns perceived by companies about the use of Web 2.0 technologies.

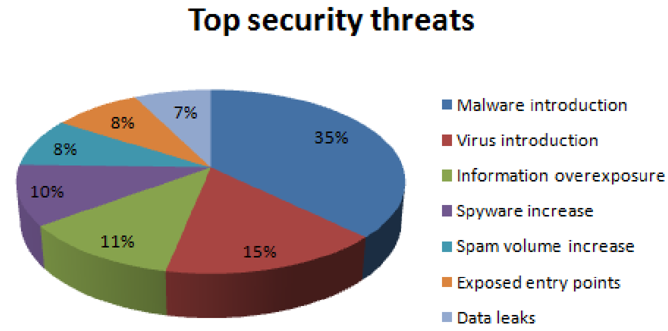


Figure 2. Top security threats of Web 2.0 usage. [7]

The security concern is a specific obstacle to adoption and integration of social media in organizations. The top four perceived threats from employees' use of Web 2.0 are malicious software (35 percent), viruses (15 percent), overexposure of information (11 percent) and spyware (10 percent).

As with any evolution of a product or service, the old ways of performing a task or providing a solution simply may not work. This is also true in reducing and mitigating Web 2.0 threats. Time tested security solutions are no longer the key defense in guarding against attacks and data loss. Some characteristics of 2.0 securities that are being discussed in the literature are:

- Traditional Web filtering is no longer adequate;
- New protocols of AJAX, SAML and XML create problems for detection;
- RSS and rich Internet applications can enter directly into networks;
- Non-static Web content makes identification difficult;
- High bandwidth use can hinder availability;
- User-generated content is difficult to contain.

Security teams of a company must be aware of the need to address Web 2.0 threat in their desktop clients, protocols and transmissions, information sources and structures, and server weakness. In fact, none of these attack vectors are new, but the way to respond to them may be. Many of the threats are obvious associated with Internet use, but there are others, particularly the ones that can lead to confidential data loss, that can be addressed and mitigated by enterprises.

Direct posting of company data to Web 2.0 technologies and communities is the most common. No vulnerability need to be exploited or malicious code injected when employees (whether as part of their responsibilities or not) simply post protected or restricted information on blogs, wikis, or social networking sites. According to many security companies, the attacks on these technologies are on the rise. Many of these attacks also come via malicious payloads, which are downloaded when spam and phishing scams are utilized. According to Sophos [13], 57% (an increase of over 70% from the previous year) of people who use social networks report receiving spam and phishing messages.

Some security concerns are specific to Web 2.0 tools used by employees. For example, technologies that are perceived to facilitate work productivity, such as webmail, collaborative platforms and content sharing applications, are less likely to raise concern than the mainstream social media tools such as Facebook, LinkedIn, YouTube and Twitter, which are typically not allowed by companies.

There are both real and perceived consequences of inappropriate Web 2.0 and social media use:

- The financial consequence for security incidents (including downtime, information and revenue loss);
- The inappropriate use of social media may loss company reputation, brand or client confidence;
- Additional unplanned investments necessary for implementing social media in their organizations;
- Litigation of legal threats caused by employees disclosing confidential or sensitive information.

Organizational leaders are facing real consequences when adopting Web 2.0 technologies, but they recognize a growing demand for employee usage. The CIOs must find the delicate balance between security and the business need for these tools, and enable their use in such a way that reduces the risk for data loss or reputational harm to the corporate brand. While a sound security policy is a necessity in proactively responding to Web 2.0, policies must be enforced by technology.

IV. Paths And Best Practices To Avoid Security Risks

Most enterprises already have a form of an acceptable use policy, which should govern the use of all resources in the enterprise computing environment. The Web 2.0 application evaluation should form part of the organization’s risk management process. Organizations can implement policies restricting employees’ use of Web 2.0 applications. The following guidelines should be followed when formulating the policy:

- The policy should be created after consultation with all stakeholders;
- The policy should be based on principles, but should be detailed enough to be enforceable;
- The policy must be effectively communicated;
- Policies should be aligned with those already in operation relating to, for example, e-mails.

Some best practices should be taken into account when implementing the policy. First, responsibility of implementing and enforcing the policy should be shared and delegated to the various departments. Second, a compliance officer should be made accountable for the oversight and co-ordination function. Finally, all users must acknowledge the policy in writing. We must always consider that a policy is only effective if it is known and understood by all users.

The users should be aware and educated on the safeguards and policies. Therefore, users must be trained to identify Web 2.0 applications and understand the risks, as well as stay informed about the latest news on fraudulent Internet activities. Employees should understand and implement the security feature, which these websites provide.

Critically read the current policy in a context of 2.0 technologies, and identify gaps that need to be addresses, is a fundamental task for a CIO. For instance, because the risks and the inherent difficulty managing the use of social networking applications, many enterprises have made the decision to not allow access to social networking services and Web 2.0 powered sites from inside the corporate perimeter. Of greatest importance is a clear and unambiguous warning in the policy about sharing confidential corporate information.

Enforcement of the policy can be made through analysis of Web logs for use during business time, or through automated searches of websites for corporate information. According to Gartner [4], many organizations have already included Web 2.0 and data protection sections to their training on protecting corporate information.

In the following, we will present some IT policies that should be included to allow a safe inclusion of 2.0 technologies in the enterprise environment.

A. Application control list

Network traffic and applications are generally controlled at the firewall by tracking the ports used, source and destination addresses, and traffic volume. However, these methods may not be sufficient to precisely define or control the traffic from Internet-based applications. To address this problem, we must use protocol decoders to decrypt and examine network traffic for signatures unique to an application. In this way, even when applications attempt to hide by using non-standard ports and protocols, they can still be discovered. In addition, protocol decoders enable decryption and examination of encrypted network traffic. This allows application control to be applied to IPSec and SSL-encrypted VPN traffic, including HTTPS, POP3S, SMTPS and IMAPS protocols.

Applications that need to be explicitly managed are entered into an application control list in the firewall policy. Administrators can create multiple application control lists, each configured to allow, block, monitor or shape network traffic from a unique list of applications. An application "whitelist" is appropriate for use in a high security network as it allows only traffic from listed applications to pass through the gateway. On the other hand, an application "blacklist" allows all unlisted application traffic to pass. Applications can be controlled individually or separated into categories and controlled as groups.

B. Application traffic shaping

Application traffic shaping allows administrators to limit or guarantee the network bandwidth available to all applications or individual applications specified in an application list entry. For example, a business could limit the bandwidth used by Skype and Facebook chat to no more than 100 kilobytes per second, or restrict YouTube traffic to reserve network bandwidth for mission critical applications. Traffic shaping can also be configured on a time-sensitive basis to restrict user access or bandwidth available to applications during certain times of the day. Traffic shaping policies must be created independently of firewall policies and application control lists so that administrators can reuse them in multiple policies and list entries. Shared traffic shaping policies can be applied to individual firewalls or across all firewalls.

C. Monitoring and review

Extensive logs and audit trails should be maintained. These should be regularly reviewed, with a reporting system implemented. The application monitoring and reporting feature may collects application traffic information and displays it using visual trend charts, giving administrators a quick way to gain insight into application usage on their networks.

The most relevant trend charts (e.g., top blocked websites, top ten applications by bandwidth) may be generated for each firewall policy that has application monitoring enabled. Using the knowledge gained from application trend charts, administrator can quickly optimize the use of applications considering the organization's security policy and worker needs.

D. Browser settings

Browser and security settings should be customized to its highest level. Alternatively, non-standard browsers (such as Mozilla Firefox) that allow for anonymous surfing and which are equipped with advanced functions can be used. Users should note whether the security features (such as HTTPS) on the browser are operating.

E. Anti-malware software

Installing anti-virus, anti-spyware programs and spam-filters for both inbound and outbound traffic should be implemented at the gateway and desktop levels. Anti-malware software, with the following functionalities, should be implemented:

- Messages should be deep-scanned, searching for signature patterns, placing reliance heuristic or behavioral based protection [14];
- Virus scanners should be able to detect any treat and update the network and firewall rules immediately;
- Utilize software that decomposes all container file types its underlying parts, analyzing the underlying parts for embedded malware.

F. Authentication

Strong or non-password based authentication should be deployed and used for access to sensitive information and resources. Web 2.0 applications usually employ weak authentication, and are targets for a chain of penetration and social engineering attacks that can compromise valuable resources. Requiring appropriate token-based or biometric authentication at key points can help to prevent incidents.

G. Avoid clickjacking

Currently, clickjacking is considered one of the most dangerous and troubling security problems on the Web [15]. In this attack, two layers appear on a site, one visible, one transparent, and users inadvertently interact with the transparent layer that has malicious intent. New countermeasures, such as NoScript with ClearClick reduce the clickjacking risk. Additionally, users can take other countermeasures to limit clickjacking risk, such as minimizing cookie persistence by logging out of applications and using a dedicated browser for each website visited.

H. Data loss protection

Data infiltration is a continuing challenge of organizations participating in the Web 2.0 environment. Protecting the integrity and confidentiality of organizational information from theft and inadvertent loss is a key issue today. The cost of dealing with a data breach continues to rise each year. In September of 2011, Reference [16] conducted a study that reveals that the average cost to an enterprise from a data breach rose from $7.12 million in 2010 to $7.25 million in 2011.

The implementation of a data loss protection (DLP) solution may be integrated with 2.0 technologies. The DLP, based on central policies, will be responsible to identify, monitor and protect data at rest, in motion, and in use, through deep content analysis.

The DLP solutions both protect sensitive data and provide insight into the use of content within the enterprise. Few enterprises classify data beyond public vs. everything else [17]. Therefore, DLP helps organizations better understanding their data, and improves their ability to classify and manage.

V. Conclusion

As we enter the second decade of the 21st century, the landscape of communication, information and organizational technologies continues to reflect emerging technological capabilities, as well as, changing user demands and needs. The Web 2.0 is a typical term used to describe these social technologies that influence the way people interact. Simultaneously, these Web 2.0 technologies are coming to the enterprise, radically transforming the way employees, customers and applications communicate and collaborate.

These technological advancements will continue to bring new opportunities and threats, thus requiring agility and continued evolution of resources. Successful organizations will be those that determine where and how to embrace these emergent tools to add new value and agility to their organizations. Success will require careful, on-going efforts to safeguard assets, including infrastructure, data, and employees, along with measured and educated adoption of new cyber technologies.

A comprehensive security program should be adopted by companies to deal with the introduction of Web 2.0 technologies in a corporate environment. As a first step, a Web 2.0 policy should be formulated implemented and the compliance with the policy should be monitored. The policy should be easy to understand, implemented and monitored, yet, detailed enough to be enforceable and be used to hold users accountable. Users should be trained on acceptable Web 2.0 practices and security features.

Besides that, the company shall adopt concrete IT policies to allow a safe inclusion of Web 2.0 technologies in the enterprise environment. A security solution that provides complete content protection, including application detection, monitoring and control is needed to discover threats embedded in Internet-based application traffic, and also to protect against data loss resulting from inappropriate use of social media applications. In addition, the content-based security enforcement is essential to mitigate these threats when they are discovered and to provide compete protection and threat elimination. Finally, other IT initiatives can be adopted such as high customized browse settings, installation of anti-malware software, adoption of strong authentication mechanisms and establishment of a data loss protection solution.

References

[1] G. Cormode and B. Krishnamurthy, “Key differences between Web 1.0 and Web 2.0”, 2008, available on http://www2.research.att.com/~bala/papers/web1v2.pdf.

[2] A. Lipsman, “The network effect: facebook, linkedin, twitter & tumbir reach new heights in May”, 2011, available on http://blog.comscore.com/2011/06/facebook_linkedin_twitter_tumblr.html.

[3] J. Sena, “The impact of Web 2.0 on technology”, International Journal of Computer Science and Network Security, Vol. 9, No. 2, pp. 378-385, 2009.

[4] M. Chui, A. Miller and R. Roberts, “Six ways to make Web 2.0 work”, McKinsey Quarterly, Vol. 1, pp. 2-7, 2009.

[5] O. Mahmood and S. Selvadurai, “Modelling web of trust with web 2.0”, World Academy of Science, Engineering and Technology, Vol. 18, pp. 122-127, 2006.

[6] A. Jaokan and C. Sharma, “Mobile VoIP – approaching the tripping point”, 2010, available on http://www.futuretext.com/downloads/mobile_voip.pdf.

[7] L. Kisselburgh, E. Spafford and M. Vorvoreanu, “Web 2.0: a complex balancing act”, McAfee White Papers, pp. 2-13, 2010.

[8] J. Bughin and M. Chui, “The rise of the networked enterprise: Web 2.0 finds its payday”, McKinsey Quarterly, Vol. IV, pp. 3-8, 2010.

[9] D. Rowe and C. Drew, “The impact of Web 2.0 on enterprise strategy”, Journal of Information Technology Management, Vol. 19, No. 10, pp. 6-13, 2006.

[10] K. Yukihiro, “In-house use of Web 2.0: entrprise 2.0”, NEC Technical Journal, Vol. 2, No. 2, pp. 46-49, 2007.

[11] S. Wright and J. Zdinak, “New communication behaviours in a Web 2.0 world”, Alcatel Lucent White Papers, pp. 5-23, 2009.

[12] H. Mjemanze and D, Skeeles, “Expert guide to Web 2.0 threats: how to prevent an attack”, ArcSight White Papers, pp. 3-14, 2010.

[13] G.Cluley and C. Theriault, “Security threat report: a look at the challenges ahead”, Sophos White Papers, pp. 2-13, 2009.

[14] S. Pruitt, “Firewalls, the future and Web 2.0”, 2007, available on http://www.networksecurityjournal.com/features/firewalls-the-future-web-2-061107/.

[15] G. Rydstedt, E. Bursztein, D. Boneh and C. Jackson, “Busting frame busting: a study of clickjacking vulnerabilities on popular sites”, Proceedings of the IEEE Symposium on Security and Privacy, pp. 2-13, 2010.

[16] B. Wedge, “Data loss prevention: keeping your sensitive data out of the public domain”, Insights on IT Risk Magazine, Vol. 9, pp. 2-23, 2011.

[17] S. Al-Fedaghi, “A conceptual foundation for data loss prevention”, International Journal of Digital Content Technology and its Applications, Vol. 5, No. 3, pp. 293-303, 2011.

Authors Profile

Fernando Almeida is a pos-doc researcher at INESC TEC. He received the PhD. in Computer Engineering in 2010 from Faculty of Engineering of University of Porto. His research interests are in information security, Web 2.0 business models and enterprise architectures.

A Knowledge-Based Educational Module for Object-Oriented Programming & The Efficacy of Web Based e-Learning

Ahmad S. Mashhour, PhD
IS Dept., University of Bahrain
Manama, Bahrain

Abstract—**The purpose of this research was to explore the effectiveness of using computer-aided learning methods in teaching compared to traditional instructions. Moreover, the research proposed an intelligent-based educational module for teaching object oriented languages. The results provide teachers positive outcomes of using computers technology in teaching. Sample of more than one hundred undergraduate students from two universities located in the Middle East participated in this empirical study. The findings of this study indicated that students using e-learning style perform more efficient in terms of understandability than traditional face-to-face learning approach. The research also indicated that female students' performance is equally likely to male students.**

Keywords_ Computer-Aided Learning (CAL); Educational module; e-Learning; Knowledge-Base; Object-Oriented Programming (OOP); Student performance; Traditional Learning.

I. INTRODUCTION

Students learn through many different mechanisms that provide the best long-term retention rates including lectures, books, demonstrations, and experimentation. Reference [1] explained variety of learning styles that include:

- Verbal style: learning materials from word-based interfaces such as books and lectures.
- Visual style: learning materials from visually oriented stimuli such as pictures, graphs, and movies.
- Sequential style: learning materials in discrete steps reaping partial understanding from incomplete instructions.
- Global Style: learning materials throughout top-down design by understanding facts of a topic before understanding its component's structures.

The use of intelligent tools and multimedia modules in education is a rapidly growing trend. Students can access their university's computer rooms and retrieve information in a more entertaining and easily remembered format than purchasing packs of extra readings, sample problems, and old exams. The writer have recently developed a knowledge-based module that incorporate intelligence in browsing materials, displaying items of information, self-correcting faults, and measuring learner progresses. But before explaining the structure of the module, let's address the available learning styles.

Many researchers have studied traditional learning style effectiveness, and they have concluded that only a small portion of student population matches the optimal learning styles. The majority of students are then left with teaching mechanisms that are not suitable for their optimal learning mode [2-4]. Currently, most universities around the world offer Internet/Intranet-based education. One of the most obvious infrastructure attributes of e-learning depends upon technology for its implementation. This includes browser technology, platform-independent transmission protocols, and media-capable features, e.g., Java-enabled client/server interactivity [5].

The aim of this research is to compare traditional learning and e-learning techniques, and the impact they have on student's performance and to gain a better understanding to factors that influence student performance. A knowledge-based educational module is developed, and students' performance is compared with traditional teaching. The research was carried through a survey of a sample of 110 respondents who are taking both tradition and online courses. A standard questionnaire was distributed via face to face interviews to secure high response rate. The collected data was analyzed and tested using SPSS and Excel programs.

The proposed educational module is a visual environment that appeals to both visual and traditional learning styles. It has many features that meet the needs of those students who are not well served by traditional learning styles; the module is designed for teaching Object Oriented Programming (OOP) concepts. The principles that characterize OOP are abstraction, encapsulation, inheritance, and polymorphism. There are many Languages that support object-oriented mechanism, include Ada95, C++, C#, Visual Basic.net, Java, and Small talk [6-7].

II. RESEARCH QUESTIONS

The aim of the research is to study the viability of using computer-aided learning methods in teaching.

This study sought to answer the following questions:

1. What models may be used in developing computer-aided learning courseware.
2. How acceptable is the online course compared to traditional learning and whether the learning outcome differs by gender.

The structure of this paper is as follows: In the next section, readers are furnished with the background materials related to computer-aided learning. The third section introduces the knowledge-base model. In the fourth section, a number of experiments have been performed to demonstrate the viability of the model. Conclusions and future works are presented within the last section.

III. Theoretical Perspective

Advancements in technology and changes in dynamics in the world have been the twin drivers of the shift from classroom based traditional learning to e-learning. Reference [8] indicated that many years ago, the development of computer software and hardware were directed towards education. During this period, higher education has witnessed fundamental changes from courses delivered in the traditional face-to-face method to those delivered via video cassette and television, to a proliferation of courses and course content delivered via computer technologies. The use of Internet resources in course and curriculum development has made a significant impact on teaching and learning. The application of the Internet has evolved from the display of static and lifeless information to a rich multimedia environment that is interactive, dynamic, and user friendly [9]. In fact, the Internet has become an important component in the teaching and learning process, and according to [10], Web 2.0 technologies have the potential to shape both the way instructors teach and the way students learn. As a result, the use of the Internet in higher education settings has become more accepted and widely used tool in academia [11-13].

Moreover, the development and refinement of university commercial developed Course Management Systems (CMS) like Blackboard and WebCT have resulted in the proliferation of web utilization in higher education [13-14]. CMS have shown significant increase of students' involvement in multiple aspects of courses [15], and while these tools were initially developed for use in distance education pedagogies, their use in on-campus classroom settings to compliment traditional courses is now considered a viable and often preferred option. In short, these technologies have made it possible to easily and efficiently distribute course information and materials to students via the Internet/Intranet, and allow for greater online communication and interaction [15-16].

Numerous studies have been conducted on computer-aided learning. Many of these studies focused on technical topics [17-18], case studies [19], and the use of educational packages [20]. Nowadays, many computer-aided learning packages were implemented in most of higher education institutions around the world [21-22], and learning from a distance continues to gain popularity. Reference [23] detailed the success of a computer-mediated asynchronous learning (CMAL) program of graduate studies in Educational Leadership and Higher Education that was offered through the University of Nebraska – Lincoln. The study explained the evolution of the concept focusing on an integrated sequence of high-quality learning to: (a) enhance student learning experiences; (b) provide greater accessibility by removing barriers of time and space; (c) deliver learning opportunities to participants around the world on a conventional university semester schedule; (d) develop learning cohorts representing many cultures and nationalities; (e) foster active and substantial participation in the learning process; (f) provide multiple pathways to learning; and (g) facilitate the development of a world-wide community of learners. The Program allowed for asynchronous interactions, and enabled students to access to the contributions of all other participants. Additionally, there were opportunities for real-time technology-based collaboration between and among participants.

The U.S. Department of Education [24] reviewed empirical studies of online learning literature from 1996 through July 2008, the key findings include:

- Blended and purely online learning conditions generally result in similar student learning outcomes.
- Elements such as video or online quizzes do not appear to influence the amount that students learn in online classes.
- Online learning can be enhanced by giving learners control of their interactions with media and prompting learner reflection.
- When groups of students are learning together online, support mechanisms such as guiding questions generally influence the way students interact, but not the amount they learn.

IV. Effect Of Learning Method On Student Performance

Although job markets worldwide are not accepting e-learning degrees, the literature continues to claim that e-learning (Web, wireless, etc.) knowledge delivery methods can be as effective as traditional (face-to-face traditional ones). References [25-28], showed no significant difference in achievement between online and face-to-face traditional delivery modes. Several other studies have shown that e-learning can improve learning and achievement for students and employees [29-31] among other. Reference [32] noted that at the University of Phoenix, standardized test scores of its online graduates were 5 to 10 percent higher than graduates' on-campus programs at three Arizona public universities. Another study at California State, Northridge looked at two groups of statistics courses and found that online students consistently scored better than their traditional counterparts by an average of 20 percent [33].

Some studies have found poorer learning outcomes associated with e-learning. One study [34] found that higher course failure rates in a Web section compared with a classroom section of an introductory psychology course. Reference [35] found poorer course grades and standardized achievement test scores following Web-based rather than classroom-based instruction in a more advanced psychology course. Similarly, other researchers found lower final exam scores for students in the instructor's Web-based Psychology Research Methods course compared with the classroom section [36]. Reference [37] summarized the failure of e-learning educational software platforms that did not perform well for users because this type of software is not based on sound educational principles.

Reference [38] investigated the issue of time through the use of a detailed comparison of the time required to prepare and teach a traditional course and the same course presented in an online format. He found that teaching in an online format can be more time-consuming than teaching in a traditional in-class format. The additional time required by the online format is found to result largely from increased student contact and individualized instruction and not from the use of technology per se.

Therefore, it is difficult to conclude that e-learning is more, less, or equally effective at the learning level than traditional classroom-based education/training. It is not surprising that results differ so widely as the studies themselves differ in many ways from each other (e.g., the content, duration, and goals of the training; the quality of the research design used).

Since student performance is a major concern of many parties including students, educational institutions, government, and parents. With the emergence of a new mean for learning, i.e., e-learning, there are many factors that can be used to decide the best learning methodology that influence the student performance. After reviewing and studying the literature I come up with seven important dimensions, including:

- Teaching materials [39],
- tutorial supporting [40-41],
- learning Environment [42-43],
- assessment techniques [44],
- collaboration with peers or groups [45-47],
- mutual communication [48], and
- teaching styles [49-51].

These dimensions are presented in (figure 1) below. An empirical test had been performed on a sample of students in a university to validate these variables. The results of testing confirmed the positive impact of these variables on the learning process. Detailed discussions of these factors are outside the scope of this research. References provided can be used as a guide for more explanation.

V. THE PROPOSED EDUCATIONAL MODEL

Object oriented programming (OOP) has become the dominant programming style in both the software industry and education over last two decades or so. It is now widely accepted that the object oriented style offers better tools for teaching and software development because they facilitate the creation of complex software products [52].

A lot of research is being conducted with regards to the effective instructional methods for the Object Oriented design and programming. Many related studies have shown that students face various learning difficulties with this paradigm [53-54]. Such studies propose the utilization of educational tools for helping students overcome learning difficulties.

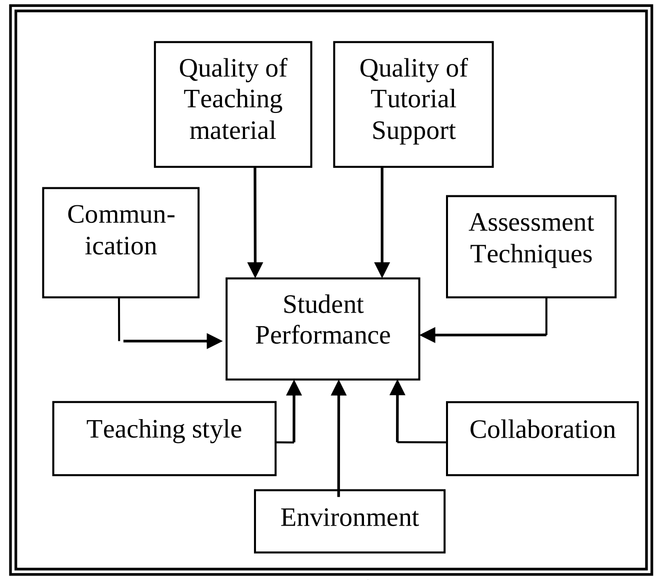


Figure 1. Student Performance Model

As mentioned above, the proposed knowledge-based module is a visual system that can be used for on-line/off-line multi-functional activities to present and consult educational knowledge. Using the computers for that type of knowledge will provide:

- Easy learner's access to many forms of educational knowledge (for on-line/off-line activities)
- Easy achievements of teachers' training objectives with separate tasks of educational knowledge
- Opportunities of electronic educational material for authors to design a proper educational environment
- An opportunity to use tools or develop new ones
- Enhancement of problem solving skills

A. *The Methodology*

The design strategy of this study is based on the top down planning process, where the highest level is the strategic planning of the problem solving mechanisms. This highest level provides learners with an overall language top-down constructs. At the next level, learners construct more detailed plans for parts of the problem.

In this level, the language specifics are not yet tied up together. The next levels require from the learners to fill the gap by the chosen implementation language (See Figure 2).

The first and second levels introduced learners to the concepts of object oriented programming principles, as more lines of code need to be organized, managed, and maintained to enhance the functionality of computing systems. This requirement coined the need for object oriented programming that will provide better paradigms and tools for:

- Modeling the real world as close to user's perspective as possible

- Easily interacting with computational environments using familiar metaphors.
- Constructing reusable software components and easily extensible libraries of software modules.
- Easily modifying and extending implementations of components without having to record every item from scratch [55-56].

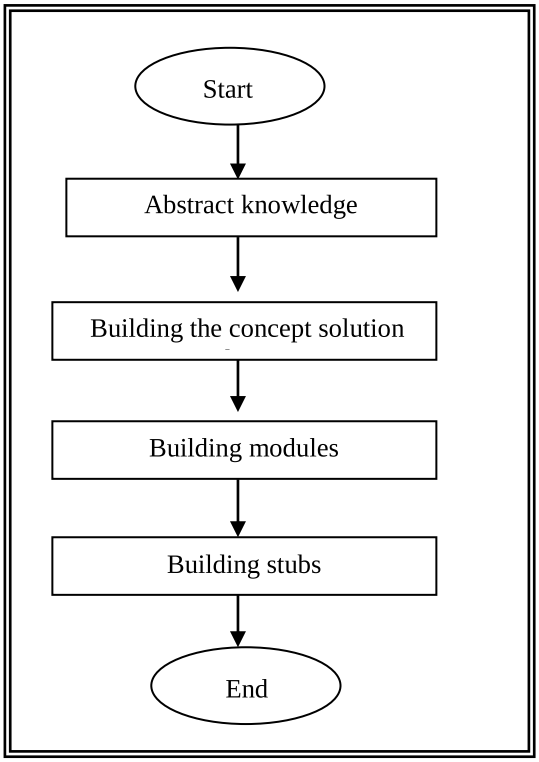


Figure 2. Abstraction Levels of Learning

The idea behind OOP was to break a program down into small and independent components. Each component could be developed independently of any other component, and then it could be combined with other component in order to form a larger application.

In OOP there are a number of basic constructs and concepts that are used in order to develop an object-oriented program [57]:

a) *Class:* A small independent program that can be combined with other similar programs in order to provide a larger application. A class is made up of attributes and methods (encapsulation);

b) Attribute: Usually a simple data element that is part of a class. It is usually a variable or a constant;

c) *Method:* Small independent part of a class's code. Similar to the notion of a function or a procedure in linear programming;

d) *Object:* A class acts only as a template and consequently in order to be used in a program an instance of this class has to be created. This instance is called an object. Any number of objects can be instantiated by a single class.

e) *Inheritance:* Ability to create a new class (subclass) from an existing class (superclass). The subclass will inherit all of the characteristics of the superclass (attributes and methods). The subclass can extend the superclass by defining its own attributes and methods.

f) *Aggregation:* The ability to join a number of instances of one class in order to create a new class (e.g. if square is a class, then joining 64 instances of this class will give a new class called chess board).

The syntax for objects in OOP can be defined as:

< Object oriented >::= <object> <classes>
<Classes>::= <Abstract Data Type> <inheritance>
<polymorphism><overloading>

Where objects communicate or invoke with one another through sending messages to each other. Collections of objects that respond to the same message are implemented through classes where a class describes and implements all the methods that capture the behavior of its instances.

Classes contribute greatly to the modeling of the real world, software extensibility, and software reusability. A class is similar to a module. Modules, in object oriented concepts, can be used in different applications that can extend or specialize a class. This can be accomplished through the use of inheritance.

The low levels of learning introduce the building block of a language implementation. It starts by demonstrating the basic constructs such as classes. The proposed system displays the grammar rule as follows:

Class < <Class-name>
{
    Private:
        <Private-member-declarations>
    Public:
        <public-member-declarations>
    }

The construct <Class-name> identifies the name of the new class definition.

The following syntax can be used to determine a newly derived class definition that inherits constructs from several basic classes:

Class <new-derived-class>:
    [virtual] [public| private] <base-class-1>
        { , [virtual] [public| private] <base-class-2 . . . > }
    {
    . . .
    }

where < base-class-1>, <base-class-2>, and so on, are existing class definitions.

First, the system introduces the basic concepts and asks students to write his/her own construct implementation. This implementation is evaluated according to its syntax structures. When students' writes constructs have not complied with the syntax, a message is issued to students and language. Syntax is displayed and the rest that does not comply is identified and a message is issued to explain to students the next steps that need to be followed.

The new model helps students by offering a wealth of information that goes to students rather than trudge to repositories of information like classrooms and libraries.

The proposed system automates testing and instructions whose purpose is to increase the level of individualization in educational process and presents more precise record-keeping on the abilities of students.

B. Test evaluation

Although performance evaluations are unique, techniques used for one problem generally cannot be used for the next problem. Nevertheless, there are common steps to all testing evaluation including:

- The definition of the system under investigation,
- selection of evaluation criteria,
- determining factors required to be analyzed,
- selection of evaluation techniques,
- designing experiments,
- analyzing and interpreting data, and
- Presenting findings.

C. Investigation of the New Model Enhancement

In order to estimate the optimal learning mode, a number of experiments have been performed to demonstrate the model viability. Table 1 presents the results of experimentation.

D. Evaluation of Learners' Performance

The teaching module must contain model of the learner performance. This can be achieved by using a specialized model called "Student Model". The student model is a file containing a learner's history and his/her present performance. The model is considered important because it measures the student understandability of taught materials, and ability to learn. The student-modeling module consists of three parts.

o Compute the learning rate and learner's skill level by:

Where S is the score domain and S(i) is the score of task number i . The t is the time domain, and t(i) is the time spent on task i. The skill level indicates the level at which students grasp concepts. This is divided into four levels giving scores from 1 to 4.

These scores are translated into qualitative term as follows:

o Update the items of information in the students' record.

o Analyze the information that was collected during the teaching process.

E. Experiment

The project proceeds as follows: A sample of 110 students was selected with 66 females' students and 44 males. From equation (1), the collected data is presented in table 1.

From Table 1, sufficient items of information cannot be drawn concerning the significance of analysis. There were needs to some statistical techniques to guide through analysis. A comparative test was used to sample means (see Table 2).

To have a better visualization, the means for the two samples were drawn (see figure 3).

Figure 3 show that the knowledge-based module has better performance on student than the traditional teaching approach. Also it looks that female may achieve better performance than male using the new approach. This visualization may sometimes be misleading. So more thorough analysis need to be done using advanced descriptive and parametric statistical techniques.

F. Analysis of Results

Researchers need to make a decision about the truth or falsity of statistical hypothesis. A statistical hypothesis is presumed to be either true or false. Using such methods would enable a decision to be made within a certain margin of errors as whether the statistical hypothesis is enabled or must be rejected as false.

$$Learning\ Rate = \frac{\sum_{i \in S} s(i)}{\sum_{i \in T} t(i)} \qquad (1)$$

Table I. STUDENT LEARNING MEASUREMENTS

Gender	measurement	Performance	Understandability	Consistency	Moral	Learning mode
female	**Un**	4	2	1	0	1
	Us	8	12	10	8	10
	Li	15	12	19	18	23
	Ve	39	40	26	40	30
male	**Un**	2	2	3	2	5
	Us	5	9	7	6	9
	Li	10	10	6	9	8
	Ve	27	23	28	27	22

Key: Un: Unlikely Us: Usually Li: Likely Ve: Very likely

Table II. COMPARATIVE TEST

Qualitative measurement	Avg-Male	Avg-Female
Un	2.8	1.6
Us	7.2	9.6
Li	8.6	17.4
Ve	25.4	35

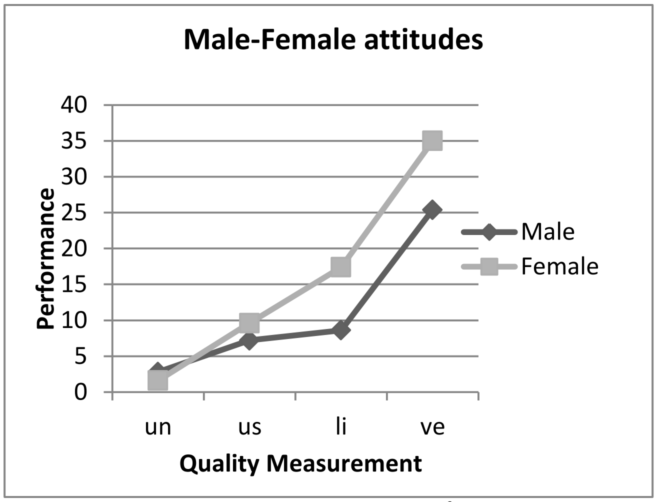


Figure 3. Comparative test of means

Four steps are required for testing of any statistical hypothesis:

- State the statistical hypothesis H0
- Specify the degree of risks.
- Assuming H0 to be correct, determine the probability (p) of obtaining a sample mean that differs from the population mean by an amount larger or equal the observed mean.
- Make a decision regarding H0 – reject it or not reject (accept it).

The hypothesis was tested to see if there is no difference between the means of male and female learning. A z-test was used to determine whether the two means of the two samples are equal. Examining the output (see Table 3), the mean data for males was 2.8 whereas the mean for female was 1.6. We have to answer the question is this difference in the mean statistically significant? Because we had no hypothesis before the study that male would learn better than female, it is appropriate to use a “two-tail” test [58].

From table below we see that the critical value of z for a two-tail test with these number (6) degrees of freedom is 1.959961. The value of z calculated for these tests is 0.56391, which is less than 1.959961. Thus the difference between the two sets of data is not significant at the 5% level.

This means that there is no difference in perception of programming by female- students from that of male students.

Table III. TEST- TWO SAMPLE FOR MEANS

	male	*female*
Mean	11	15.9
Known Variance	98.26667	203.7467
Observations	4	4
Hypothesized Mean Difference	0	
Z	-0.56391	
P(Z<=z) one-tail	0.286406	
z Critical one-tail	1.644853	
P(Z<=z) two-tail	0.572812	
z Critical two-tail	1.959961	

To get a better understanding of the difference between group means, the concepts of confidence intervals (CI) was used. The CI for the difference between □□ and □2 is outlined as follows. The two samples, male and female, were treated as one sample of n pairs.

For each pair, the difference in average performance was computed and CI was constructed (See Table 4).

Table IV. DESCRIPTIVE STATISTICS

Mean	-4.9
Standard Error	2.594224
Median	-5.6
Standard Deviation	5.188449
Sample Variance	26.92
Kurtosis	-3.64305
Skewness	0.379286
Range	10.8
Minimum	-9.6
Maximum	1.2
Sum	-19.6
Count	4
Confidence Level(95.0%)	8.255987

The 100(1- α)% confidence interval is given by:

$$\left(\bar{x} - t_{[1-\alpha/2;\, n-1]}\, s/\sqrt{n}\,,\ \bar{x} + t_{[1-\alpha/2;\, n-1]}\, s/\sqrt{n}\right)$$

The $t_{[1-\alpha/2;\, n-1]}$ is the (1- α/2-quantile of a t-variate with n-1 degrees of freedom. In this case, from table IV, the CI is 8.255987 and the 95% confidence interval is:

-4.9 ± 8.255987 = (-13.155987, 3. 355987)

The confidence interval includes zero. Therefore, the two samples are not different.

In empirical research the hypotheses of greatest interest usually pertain to means, proportions, and /or correlation coefficients. At times, however, there are interests in questions regarding variances such as:

Are the individual differences in learning among male greater than among female? Assuming that the two sample variances are equal, a t-test can be used and result presented in table 5.

From table 5, one can see that there is no much evidence to reject the assumption. Furthermore, there are no much differences between the two-sample variances, i.e. inter-variance for gender is not significantly different.

Table V. t-TEST - TWO-SAMPLE ASSUMING EQUAL VARIANCES

Measures	*male*	*female*
Mean	11	15.9
Variance	98.26667	203.7467
Observations	4	4
Pooled Variance	151.0067	
Hypothesized Mean Difference	0	
Df	6	
t Stat	-0.56391	
P(T<=t) one-tail	0.296624	
t Critical one-tail	1.943181	
P(T<=t) two-tail	0.593248	
t Critical two-tail	2.446914	

VI. CONCLUSIONS

This study shows that computer-aided learning (CAL) is offering a viable educational alternative to traditional education, while at the same time there is an increasing pressure on educational institutions to move towards this type of education. The effectiveness and the efficiency of the educational technology tools are becoming overwhelming in the learning process. The use of the Information Technology educational tools such as WebCT, Blackboard, the Internet and other software tools has become more accepted and widely used tools in higher education due to the development of the interactive features of Web 2.0.

The benefits achieved include the quality of skills earned by the learner, the reduction in the cost of the learning process, and the globe reach of the education. The technology have also made it possible to easily and efficiently deliver and manage course material (and other related aspects such as fees and online exams) to student via the Internet, and many academic departments are now struggling to support face-to-face courses by creating course websites to post lecture notes, syllabi, assignments and other course-work.

The research in this paper is based on teaching object oriented programming to university students.

This type of learning was approached because OOP seems to have less complexity in the development of large applications (consequently faster, easier and cost effective development), easier in detecting errors, easier in updating of application, ability to extend software applications, the efficiency in creating new classes from existing ones (inheritance), and the ability to store and reuse classes whenever they are needed (reusability). In this research, a model course was designed for learning based on object oriented programming.

The proposed course was implemented on two groups of male and female students at a Persian Gulf State university. The results were achieved by statistical analysis and testing using SPSS, and have showed better response for the new proposed course in comparison to traditional teaching. In addition, male and female students reacted equally to the new discipline of teaching.

The main limitation in this study is that the number of students surveyed does not express concrete opinion. Thus survey more students in more number of universities is needed.

REFERENCES

[1] J. T. Bell, and H. S. Fogler, “Vicher: a virtual reality based educational module for chemical reaction engineering”, Computer Applications in Engineering Education, vol. 4, No. 4, pp. 285-296. 1996.

[2] A. Dovgiallo, V. Gritsenko, V. Bykov, V. Kolos, S. Kudrjavtseava, Y. Tsybenko and N. Vlasenko “Intelligence in educational technologies for traditional and distance case: trends and the need for cooperation”, Proceedings of UNESCO Second International Congress, 1996.

[3] E. Chrysler, “Measuring the effect of redesigning an introductory MIS course”, Journal of Computer Information Systems, vol. 36, number 2, pp. 30-36, Winter 1995-1996.

[4] L. A. Harris, and B. A. Weaver, “A comparison of information systems ethics and attitudes among college students, Journal of Computer Information Systems, Winter (1994-1995).

[5] A. L. Ellis, E.D. Wagner, and W.R. Longmire, Managing Web-based Training, ASTD, 1999

[6] J. Farrell, Microsoft visual C#, Course Technology, Cengage Learning, 4th edition, 2010

[7] S. Khosh, and R. Abnous, “Object Oriented: concepts, languages, databases, and user”, Proceeding of The International Conference on Intelligent Information Systems, S. Mani, R. Shankle, M. Dick, M. Pazzani Eds., Two-stage Machine learning model for guideline development artificial intelligence in medicine, 1999, pp.16, 51-80,

[8] C. F. Harrington, S. A. Gordon, and J. S. Timothy, “Course management system utilization and implications for practice: a national survey of department chairpersons”, Online Journal of Distance Learning Administration, volume VII, number IV, Winter 2004, retrieved from http://www.westga.edu/~distance/ojdla/winter74/harrington74.

[9] P. Powel, and C. Gill, “Web content management systems in higher education”, Educause Quarterly,2003, 2:43-50.

[10] P. Daniels, “Course management systems and implications for practice”, International Journal of Emerging Technologies & Society, vol. 7, no. 2, 2009, pp: 97–108, retrieved from http://www.swinburne.edu.au/hosting/ijets/journal/V7N2/pdf/Article3-Daniels.pdf.

[11] R. Glahn, and R. Glen, “Progenies in education: the evolution of internet teaching”, Community College Journal of Research and Practice. 26: 777-785, 2002.

[12] R. Katz, “Balancing technology and tradition: the example of course management systems”, Educause Quarterly, July-August, 2003.

[13] J. Angelo, “New lessons in course management”, University Business, 2004, retrieved from

http://www.universitybusiness.com/page.cfm?p=616.

[14] G. Morgan, "Course management system use in the University of Wisconsin system", Educause Center for Applied Research (ECAR), May, 2003.

[15] B. Stith, "Web-Enhanced Lecture Course Scores Big with Student and Faculty, Syllabus", T.H.E. Journal, 2000, 27 (8), pp. 20-28, retrieved from www.editlib.org/d/14856/proceeding_14856.pdf.

[16] S. Gray, "Collaboration tools". Syllabus,1999, vol. 12, No. 5

[17] Z. Al-Khanjari, F. Masoud, S. Kutti, and M.Hatam,"Plugging software tools in the existing e-learning portals", in the International Arab Conference on Information Technology (ACIT), 2002.

[18] H. Alfawareh, "Rebuilding object oriented software architecture tool for understanding purposes", in the proceedings of the International Arab Conference on Information Technology (ACIT), 2002.

[19] H. Chorfi, and M. Jemni,, "e-learning at Esstt, a case study and perspectives", in the proceedings of the International Arab Conference on Information Technology (ACIT), 2002.

[20] S. Mani, R. Shankle, M. Dick, and M. Pazzani, Two-stage machine learning model, 1999.

[21] C. P. Chrysostomou, and G. Papadopoulos, "An evaluation of e-learning technologies: establishing an object-oriented approach to learning object design and development", the proceeding of the first International Conference on E-business and e-Learning, Amman- Jordan, 2005.

[22] [22] M. Biggs, "Comparison of student perceptions of classroom instruction: traditional, hybrid, and distance education", Turkish Online Journal of Distance Education, 7(2), 2006, pp. 46-51.

[23] S. L. Stick, and N.V. Ivankova, "A decade of innovation and success in virtual learning: a world-wide asynchronous graduate program in educational leadership and higher education", Online Journal of Distance Learning Administration, volume VII, number IV, Winter 2004.

[24] B. Means, Y. Toyama, R. Murphy, M. Bakia, and K. Jones, "Evaluation of evidence-based practices in online learning", A meta-analysis and review of online-learning studies, 2009, Washington, D.C.: U.S. Department of Education, retrieved from http://ctl.sri.com/people/displayPerson.jsp?Nick=bmeans.

[25] S. Sawa, "Asynchronous streaming video from the classroom to the remote student's desktop: preliminary results of a distance education project", in T. Reeves & S. Yamashita (Eds.), Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, 2006, pp. 898-903, Chesapeake, VA: AACE, retrieved from http://www.editlib.org/p/23813.

[26] K. Machtmes, J. W. Asher, "A meta-analysis of the effectiveness of tele-courses in distance education", The American Journal of Distance Education, 2000, 14(1), 27 – 46.

[27] T. L. Russell, "The no significant difference phenomenon: a comparative research annotated bibliography on technology for distance education", Raleigh: Office of Instructional Telecommunication, North Carolina State University, 1999.

[28] J. T. Socks, and P. P. Freddolino,"Evaluation of a world wide Web-based graduate social work research methods course", Computers in Human Services, 1998, pp. 15: 51-69.

[29] M. Brown, "Handhelds in the classroom", Education World, May, 2001

[30] D. L. Whetzel, D. B. Felker, and K. M. Williams, "A real world comparison of the effectiveness of satellite training and classroom training", Educational Technology Research and Development, 1996, 44: 5-18.

[31] M. Allen, E. Mabry, M. Mattrey, J. Bourhis, S. Titsworth, and N. Burrell, "Evaluating the effectiveness of distance learning: a comparison using meta-analysis", Journal of Communication, 2004, pp. 54: 402-420.

[32] L. Gubernick, and A. Ebeling, "I got my degree through e-mail", Forbes, 159, pp. 84-92, 1999.

[33] K. McCollum, "A professor divides his class in two to test value of online instruction", Chronicle of Higher Education, 1997, 43, p. 23.

[34] S. B. Waschull, "The online delivery of psychology courses: attrition, performance, and evaluation", Teaching of Psychology, 2001 28: 143-146.

[35] K. E. Mottarella, B. A. Fritzsche, and T. J. Parrish, "Who learns more? Achievement scores following Web-based versus classroom instruction in psychology courses", Psychology of Learning and Teaching, 2004, 4: 50-53.

[36] M. Newlin, A. Wang, and M. Kosarzycki, "Who succeeds in Web-courses", in the International Conference on Intelligent Information Systems1998, EDUCOM, Orlando, FL.

[37] [37] D. Bunis, Commentary, College and University, Fall, v.78, n.2, 2003.

[38] J. Cavanaugh, "Teaching online - a time comparison", Online Journal of Distance Learning Administration, volume VIII, Number I, Spring, 2005.

[39] A. Rogers, Teaching Adults, Open University Press, 2002, third edition. Berkshire, UK.

[40] F. Heidari, and K. Galvin, "Action learning groups: can they help students develop their knowledge and skill?", Nurse Education in Practice, 2003, 3, pp. 49-55 retrieved fromhttp://www.educause.edu/Resources/EvaluationofEvidenceBasedPract/174235.

[41] H. Ashton, and C. Shuldham, " An exploratory study of student's perceptions of the benefits of regular small group tutorials in a two year data release course", Journal of Advanced Nursing, 1994, pp. 20, 925–934.

[42] N. Zhou, and D. Agosto, "The collaborative information behavior of middle school students in online learning environments", presented at the Panel of Behaviors and Preferences of Digital Natives: Informing a Research Agenda at the annual conference of American Society of Information Science and Technology (ASIST) 2007, Milwaukee, WI, USA.

[43] N. Zhou, "Interactions of education and culture for economic and human development: an Asian perspective", in Delors et al., "Learning: the treasure within", Paris: UNESCO Press, 1996.

[44] R. Stiggins, J. Arter, J. Chappuis, and S. Chappuis, Classroom Assessment for Student Learning: Doing it Right-Using it Well, Allyn & Bacon; 1st edition, 2009.

[45] B. Barros, and M. F. Verdejo, "An approach to analyze collaboration when shared structured workspaces are used for carrying out group learning processes", AIED'99, Lemans, July, 1999.

[46] C. Crook, "Computers and the collaborative experience of learning", routledge, Newyork,1996, p.79.

[47] G. Stahl, T. Koschmann, and D. D. Suthers, Computer-supported collaborative learning: An historical perspective, Cambridge handbook of the learning sciences, 2006.

[48] R. Fenton, N. O'Leary, "Improving student achievement through enhancing the instructional communication competence of teachers", paper presented at: Annual meeting of the communication and instruction division of the western states communication association", February 1990.

[49] F. C Worrell, L. D. Kuterbach, "The use of student ratings of teacher behaviors with academically talented high school students", The Journal of Secondary Gifted Education, 2001, pp. 14, 236-247.

[50] N. A. Buang, A. Abd Samd, "Is there a relationship between business studies teachers' teaching styles and students' motivation to learn?", the soc. Sci., 6(5): pp. 357-360, 2011 (Medwell Journal).

[51] M.S. Hass, "The influence of teaching methods on student achievement on Virginia's end of course standards of learning test for Algebra I", unpublished Ph.D. Dissertation submitted to Faculty of Virginia Polytechnic Institute and State University, Sep., 2002.

[52] M. C. Okur, "Teaching object oriented programming at the introductory level", Journal of Yasar University, 1(2), 2006, pp. 149-157, http://joy.yasar.edu.tr/makale/no2_nol1/05_m_cudi_okur.pdf.

[53] S. Georagantaki, and S. Retails, "Using educational tools for teaching object oriented design and programming", journal of Information Technology Impact, Vol. 7, No., 2, 2007, pp. 111-130.

[54] E. Lahtinen, K. Ala-Mutka, and H. M. Jarvinen, "A study of the difficulties of novice programmer", in the proceeding of the 10th annual SIGCST conference on innovation and technology in computer science education, 14-18 June, 2005, pp. 27-29.

[55] John Vinsonhaler, Jean Vinsonhaler, and C. Wagner, “PERIS: a knowledge based system for academic productivity”, Journal of Computer Information Systems, spring, 1996.

[56] V. Shankaraman, and B. S. Lee, ”Knowledge - based safety training system - a prototype implementation”, Computer in Industry, 25, 1994, PP. 145-157.

[57] J. Farrell, An Object-Oriented Approach to Programming Logic and Design, 3rd ed., Course Technology, Cingage Learning, 2010.

[58] Gips, James (1997). Mastering Excel, A problem Solving Approach. John Wiley & Sons.

AUTHOR PROFILE

Ahmad Mashhour received his PhD degree in Information Systems from the University of London (LSE), U.K in 1989, and his Masters and BS degrees from the University of Pittsburgh, PA., USA in the years 1983, 1981 respectively. He is currently Associate Professor in Information Systems at the University of Bahrain. His research interests include e-Learning, Information security, and Simulation Modeling & Analysis.

Clustering as a Data Mining Technique in Health Hazards of High levels of Fluoride in Potable Water

T..Balasubramanian
Department of Computer Science,
Sri Vidya Mandir Arts and Science college, Uthangarai(PO),
Krishnagiri(Dt), Tamilnadu, India.

R.Umarani
Department of Computer Science,
Sri Saradha College for Women,
Salem, Tamilnadu, India

Abstract— **This article explores data mining techniques in health care. In particular, it discusses data mining and its application in areas where people are affected severely by using the underground drinking water which consist of high levels of fluoride in Krishnagiri District, Tamil Nadu State, India. This paper identifies the risk factors associated with the high level of fluoride content in water, using clustering algorithms and finds meaningful hidden patterns which gives meaningful decision making to this socio-economic real world health hazard. [2]**

Keywords-Data mining, Fluoride affected people, Clustering, K-means, Skeletal.

I. INTRODUCTION

A. Data Mining

Data Mining is the process of extracting information from large data sets through using algorithms and Techniques drawn from the field of Statistics, Machine Learning and Data Base Management Systems. Traditional data analysis methods often involve manual work and interpretation of data which is slow, expensive and highly subjective Data Mining, popularly called as knowledge discovery in large data[1], enables firms and organizations to make calculated decisions by assembling, accumulating, analyzing and accessing corporate data. It uses variety of tools like query and reporting tools, analytical processing tools, and Decision Support System. [5][8].

B. Fluoride as a Health Hazard

Fluoride ion in drinking water ingestion is useful for Bone and Teeth development, but excessive ingestion causes a disease known as Fluorosis. The prevalence of Fluorosis is mainly due to the consumption of more Fluoride through drinking water. Different forms of Fluoride exposure are of importance and have shown to affect the body's Fluoride content and thus increasing the risks of Fluoride-prone diseases. [10]Fluorosis was considered to be a problem related to Teeth only. But it now has turned up to be a serious health hazard. It seriously affects Bones and problems like Joint pain, Muscular Pain, etc. which are its well-known manifestations. It not only affects the body of a person but also renders them socially and culturally crippled.

The goal of this paper by using the clustering algorithms as a tool of data mining technique to find out the volume of people affected by the high fluoride content of potable water.

II. MATERIALS AND METHODS

A. Literature Survey of The Problem

To understand the health hazards of fluoride content on living beings, discussions were made with medical practitioners and specialists like General Dental, Neuro surgeons and Ortho specialists. We have also gathered details about the impact of high fluoride content water from World Wide Web [9]. By analyzing all these we came to know that the increased fluoride level in ground water creates dental, skeletal and neuro problems. In this analysis we focus only on skeletal hazards by high fluoride level in drinking water. Level of fluoride content in water in different regions of Krishnagiri District was obtained from Water Analyst . Based on the recommendations of WHO which released a water table, Tamil Nadu Water And Drainage Board (TWAD) suggested that the level of fluoride content in drinking water should not exceed 1.5 mg/L.[7]

The water table also shows the minerals content level and associated health hazards. We found out that Krishnagiri District of Tamil Nadu in India is most affected by fluoride level in water by naturally surrounded hills in the District. They have analyzed the sample ground potable water from various regions of Krishnagiri District and maintained a table of High level fluoride (1.6mg/L to 2.4mg/L) contaminated ground drinking water of panchayats and villages list in this District. We conclude that in Krishnagiri district, many people in the villages and panchayats are severely affected by ground potable water. So we decided to make a survey and found out the combination of diseases which are possibly affected mostly by high fluoride content in water.

Figure 1. Skeketal Osteoroposis by Fluoride

TABLE 1. CLASSIFICATION OF SYMPTOMS OF DISEASES

Neck pain	Joint pain	Body Pain	Foot Neck Pain	Class
Low	Low	--	--	Mild Skeletal
Low	Low	Low	--	Mild Skeletal
Low	Low	Low	Low	Mile to Moderate Skeletal
Low	Medium	Low	Medium	Moderate Skeletal
Low	Medium	Low	High	Moderate Skeletal
Low	Medium	Medium	-Medium	Osteoporosis

B. Data Preparation

Based on the information from various physicians and water analyst, we have prepared questionnaires to get raw data from the various fluoride impacted villages and panchayats, having fluoride level in water from 1.6mg/L to 2.4mg/L. People of different age groups with different ailments were interviewed with the help of questionnaires prepared in our mother tongue, Tamil since the people in and around the district are illiterate.

Total data collected from Villages and Panchayats

$$\left.\begin{array}{ll}\text{Men} & 251\ (48\%)\\ \text{Women} & 269\ (52\%)\end{array}\right\}\ 520$$

Based on the medical practitioner's advice, while classifying the data, the degrees of symptoms are placed in several compartments as follows:

Mild Skeletal Victims

Moderate Skeletal Victims

Osteoporosis Victims

With the following classification,

Those who are found with one to three low symptoms are grouped as Mild victim of skeletal disease.

Those who are found with four low symptoms or one to three medium and one high symptom are grouped as Moderate victims of skeletal disease.

Those who are found with more than two medium symptoms are grouped as osteoporosis victims of skeletal disease.

C. Clustering as the Data mining application

Clustering is one of the central concepts in the field of unsupervised data analysis, it is also a very controversial issue, and the very meaning of the concept "clustering" may vary a great deal between different scientific disciplines [1]. However, a common goal in all cases is that the objective is to find a structural representation of data by grouping (in some sense) similar data items together. A cluster has high similarity in comparison to one another but is very dissimilar to objects in other clusters.

D. Weka as a data miner tool

In this paper we have used WEKA (to find interesting patterns in the selected dataset), a Data Mining tool for clustering techniques.. The selected software is able to provide the required data mining functions and methodologies. The suitable data format for WEKA data mining software are MS Excel and ARFF formats respectively. Scalability-Maximum number of columns and rows the software can efficiently handle. However, in the selected data set, the number of columns and the number of records were reduced. WEKA is developed at the University of Waikato in New Zealand. "WEKA" stands for the Waikato Environment of Knowledge Analysis. The system is written in Java, an object-oriented programming language that is widely available for all major computer platforms, and WEKA has been tested under Linux, Windows, and Macintosh operating systems. Java allows us to provide a uniform interface to many different learning algorithms, along with methods for pre and post processing and for evaluating the result of learning schemes on any given dataset. WEKA expects the data to be fed into be in ARFF format (Attribution Relation File Format)[12].

WEKA has two primary modes: experiment mode and exploration mode .The exploration mode allows easy access to all of WEKA's data preprocessing, learning, data processing, attribute selection and data visualization modules in an environment that encourages initial exploration of data. The experiment mode allows larger-scale experiments to be run with results stored in a database for retrieval and analysis.

E. Clustering in WEKA

The classification is based on supervised algorithms. This algorithm is applicable for the input data. The process of grouping a set of physical or abstract objects into classes of similar objects is called clustering.. The Cluster tab is also supported which shows the list of machine learning tools. These tools in general operate on a clustering algorithm and run it multiple times to manipulating algorithm parameters or input data weight to increase the accuracy of the classifier. Two learning performance evaluators are included with WEKA [6].

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

F. Experimental Setup

The data mining method used to build the model is cluster. The data analysis is processed using WEKA data mining tool for exploratory data analysis, machine learning and statistical learning algorithms. The training data set consists of 520 instances with 15 different attributes. The instances in the dataset are representing the results of different types of testing to predict the accuracy of fluoride affected persons. According to the attributes the dataset is divided into two parts that is 70% of the data are used for training and 30% are used for testing. [11]

G. Learning Algorithm

This paper consists of an unsupervised machine learning algorithm for clustering derived from the WEKA data mining tool. Which include:

- K-Means

The above clustering model was used to cluster the group of people who are affected by skeletal fluorosis at different skeletal disease levels and to cluster the different water sources using by the people which are causes for skeletal fluorosis in krishnagiri district.

III. Discussion And Result

A. Attributes selection

First of all, we have to find the correlated attributes for finding the hidden pattern for the problem stated. The WEKA data miner tool has supported many in built learning algorithms for correlated attributes. There are many filtered tools for this analysis but we have selected one among them by trial.[5]

Totally there are 520 records of data base which have been created in Excel 2007 and saved in the format of CSV (Comma Separated Value format) that converted to the WEKA accepted of ARFF by using command line premier of WEKA.

The records of data base consist of 15 attributes, from which 10 attributes were selected based on attribute selection in explorer mode of WEKA 3.6.4. (Fig 2)

S.NO.	Attributes	Data Type
01.	Name	Text
02.	Age	Numeric(Integer)
03.	Education	Text
04.	Sex	Character
05.	Fluoride Level	Numeric(Real)
06.	Profession	Text
07.	Praganancy status	Boolean
08.	Drinking water	Text
09.	Duration	Numeric(Integer/Real)
10.	Known status of fluoride	Boolean
11.	Neck Pain	Numeric(Binary)
12.	Joint Pain	Numeric(Binary)
13.	Body Pain	Numeric(Binary)
14.	Foot Neck Pain	Numeric(Binary)
15.	Disease Level	Text

TABLE 2. CLASSIFICATION OF ATTRIBUTES

We have chosen Symmetrical random filter tester for attribute selection in WEKA attribute selector. It listed 14 selected attributes, but from which we have taken only 8 attributes. The other attributes are omitted for the convenience of analysis of finding impaction among peoples in the district

S.NO.	Attributes	Data Type
01.	Age	Numeric(Integer)
02.	Education	Text
03.	Fluoride Level	Numeric(Real)
04.	Drinking water	Text
05.	Duration	Numeric(Integer/Real)
06.	Neck Pain	Numeric(Binary)
07.	Joint Pain	Numeric(Binary)
08.	Body Pain	Numeric(Binary)
09.	Foot Neck Pain	Numeric(Binary)
10.	Disease Level	Text

TABLE 3. SELECTED ATTRIBUTES FOR ANALYSIS

B. K-Means Metho

The k-Means algorithm takes the input parameter, k, and partitions a set of n objects into k clusters so that the resulting intracluster similarity is high but the intercluster similarity is low. Cluster similarity is measured in regard to the mean value of the objects in a cluster, which can be viewed the cluster's centroid or center of gravity.

The k –Means algorithm proceeds as follows

First , it randomly selects k of the objects, each of which initially represents a cluster mean or center. For each of the remaining objects, an object is assigned to the cluster to which it is the most similar, based on the distance between the object and the cluster mean. It then computes the new mean for each cluster. This process iterated until the criterion function converges. Typically, the square-error criterion is used, defined as [2] [3] [4]

$$E=\sum_{I=1}^{K} \sum_{P\in C_i} |p-m_i|^2$$

Where E is the sum of the square error for all objects in the data set; p is the point in space representing a given object; and mi is the mean of cluster Ci . In other words, for each object in each cluster, the distance from the object to its cluster center is squared, and the distances are summed. This criterion tries to make the resulting k clusters as compact and as separate as possible

1) K-Means algorithm

Input;

= k:the number of clusters,

= D:a data set containing n objects

Output: A set of k clusters.

Method:

(1) arbitrarily choose k objects from from D as the initial cluster centers;

(2) (re)assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster;
(3) Update the cluster means, i.e., calculate the mean value of the objects for each cluster;
(4) until no change;

```
=== Run information ===

Evaluator:   weka.attributeSelection.SymmetricalUncertAttributeEval
Search:      weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation:    FORMAT OF 1-520 SKELETAL-weka.filters.unsupervised.attribute.Remove-R1
Instances:   520
Attributes:  15
             Name
             Age
             Education
             Sex
             FL
             Profession
             Pragnancy status while interview
             Drinking water type
             Duration of  drinking water used in years
             Known status of fluoride impact
             Neck Pain
             Joint Pain
             Body Pain
             Food Neck Pain
             Disease Level
Evaluation mode:    evaluate on all training data



=== Attribute Selection on all input data ===

Search Method:
	Attribute ranking.

Attribute   Class (nominal): 15 Disease Level):
	Symmetrical Uncertainty Ranking Filter

Ranked attributes:
 0.42554  13 Body Pain
 0.39888  12 Joint Pain
 0.37011  11 Neck Pain
 0.29908   1 Name
 0.24185  14 Food Neck Pain
 0.11147   2 Age
 0.09357   6 Profession
 0.09249   3 Education
 0.07813   9 Duration of  drinking water used in years
 0.01282   7 Pragnancy status while interview
 0.01263  10 Known status of fluoride impact
 0.01133   8 Drinking water type
 0.00667   4 Sex
 0          5 FL

Selected a    12    11    1    14    2    6    3    9    7    10    8    4    5:14
```

Figure 2. Attribute selection in WEKA Explorer

Suppose that there is a set of objects located in space as depicted in the rectangle shown in fig (a) Let k = 3; i.e., the user would like the objects to be partitioned into three clusters.

According to the algorithm above we arbitrarily choose three objects as the three initial cluster centers, where cluster centers are marked by a "+". Each objects is distributed to a cluster based on the cluster center to which it is the nearest. Such a distribution forms encircled by dotted curves as show in fig (a)

Next, the cluster centers are updated. That is the mean value of each cluster which is recalculated based on the current objects in the cluster. Using the new cluster centers, the objects are redistributed to the clusters based on which cluster center is the nearest. Such a redistribution forms new encircled by dashed curves, as shown in fig (b).

This process iterates, leading to fig (c). The process of iteratively reassigning objects to clusters to improve the partitioning is referred to as iterative relocation. Eventually, no redistribution of the objects in any cluster occurs, and so the process terminates. The resulting cluster is returned by the clustering process.

C. K-Means in WEKA

The learning algorithm k-Means in WEKA 3.6.4 accepts the training data base in the format of ARFF.

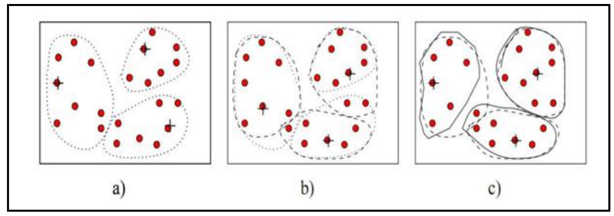


Figure 3. Clustering of a set of objects based on k-means method

It accepts the nominal data and binary sets. So our attributes selected in nominal and binary formats naturally. So there is no need of preprocessing for further process.

We have trained the training data by using the 10 Fold Cross Validated testing which used our trained data set as one third of the data for training and remaining for testing.

After training and testing this gives the following results.

```
=== Run information ===

Scheme:      weka.clusterers.SimpleKMeans -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -S
Relation:    bbb
Instances:   520
Attributes:  9
             Age
             FL
             Drinking water type
             Duration of  drinking water used in years
             Neck Pain
             Joint Pain
             Body Pain
             Foot Neck Pain
Ignored:
             Disease Level
Test mode:   Classes to clusters evaluation on training data
=== Model and evaluation on training set ===


kMeans
======

Number of iterations: 5
Within cluster sum of squared errors: 936.7893495509218
Missing values globally replaced with mean/mode

Cluster centroids:
                                          Cluster#
Attribute                                Full Data       0        1
                                           (520)      (387)    (133)
=========================================================================
Age                                       33.8635   35.7984  28.2331
FL                                         1.7781    1.8171   1.6647
Drinking water type                    bore water bore water well water
Duration of  drinking water used in years    10.0      10.0     20.0
Neck Pain                                  0.3596    0.385    0.2857
Joint Pain                                 0.4115    0.4574   0.2782
Body Pain                                  0.3       0.3256   0.2256
Foot Neck Pain                             0.2173    0.2351   0.1654


Clustered Instances

0      387 ( 74%)
1      133 ( 26%)


Class attribute: Disease Level
Classes to Clusters:

  0   1  <-- assigned to cluster
 139  69 | None
  94  25 | Mild Skeletal
  57  17 | Moderate Skeletal
  97  22 | Osteoporosis

Cluster 0 <-- Osteoporosis
Cluster 1 <-- None
```

Figure 4. K-means in weka based on diseases symptoms

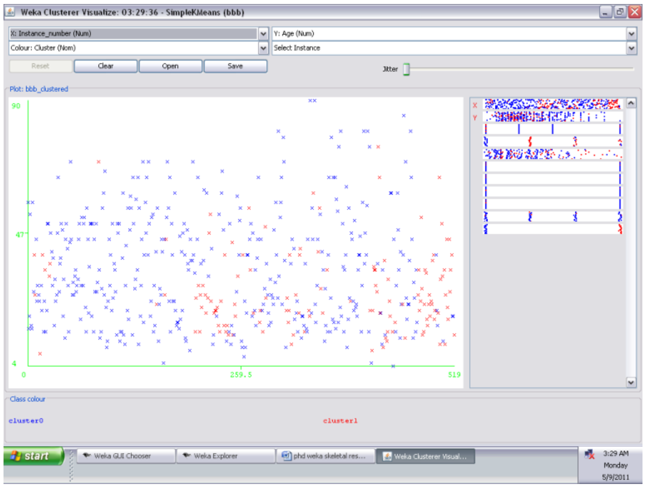


Figure 5. Disease symptoms in clusters of kmeans in weka

1) *Euclidean distance*

K-means cluster analysis supports various data types such as Quantitative, binary, nominal or ordinal, but do not support categorical data. Cluster analysis is based on measuring similarity between objects by computing the distance between each pair. There are a number of methods for computing distance in a multidimensional environment.

Distance is a well understood concept that has a number of simple properties.

- Distance is always positive
- Distance from point x to itself is always zero
- Distance from point x to point y cannot be greater than the sum of the distance from x to some other point z and distance from z to y.
- Distance from x to y is always the same as from y to x.

It is possible to assign weights to all attributes indicating their importance. There are number of distance measures such as Euclidean distance, Manhattan distance and Chebychev distance. But in this analysis Weka tool used Euclidean distance. Euclidean distance of the difference vector is most commonly used to compute distances and has an intuitive appeal but the largest valued attribute may dominate the distance. It is therefore essential that the attributaes are properly scaled.

Let the distance between two points x and y be D(x,y).

$$D(x,y) = (\sum(x_i-y_i)^2)^{1/2}$$

2) *Clustering of Disease Symptoms*

The collected disease symptoms such as Neck pain, Joint pain, Body \pain, Foot Neck as raw data, supplied to kmeans method is being carried out in weka using Euclidean distance method to measure cluster centroids. The result is obtained in iteration 12 after clustered. The centroid cluster points are measured based on the diseases symptoms and the water they are drinking. Based on the diseases symptoms in raw data the kmeans clustered two main clustering units. From the confusion matrix above we came to know that the district mainly impacted by skeletal osteoporosis. (Fig 3)

IV. CONCLUSION

Data mining applied in health care domain, by which the people get beneficial for their lives. As the analog of this research we found out that the meaningful hidden pattern from the real data set collected the people impacted in Krishnagiri district is by drinking high fluoride content of potable water. By which we can easily know that the people do not get awareness among themselves about the fluoride impaction. If it continues in this way, it may lead to some primary health hazards like Kidney failure, mental disability, Thyroid deficiency and Heart disease.

However the Primary Health hazards of fluoride are Dental and Bone diseases which disturbed their daily 000000 life. It is primary duty of the Government to providing good hygienic drinking water to the people and reduces the fluoride content potable water with the latest technologies and creating awareness among the people in some way like medical camps and taking documentary films. Through this research the problem of fluoride in krishnagiri come to light. It is a big social relevant problem. Pharmaceutical industries also can identify the location to develop their business by providing good medicine among people with service motto.

REFERENCE

[1] Jain, M. Murty, and P. Flynn, "Data clustering: A review," ACM Computing Surveys, vol. 31, no. 3, pp. 264–323, 1999.
[2] Jiawei Han and Micheline Kamber – Data mining concepts and Techniques. -Second Edition –Morgan Kaufmann Publishers
[3] Arun K.Pujari –Datamining Techniques – University Press.
[4] Introduction to Datamining with case studies - G.K.Gupta PHI.
[5] Berrry Mj Linoff G Data mining Techniques: for Marketing, Sales and Customer support USA.Wiley,1997.
[6] Weka3.6.4 data miner manual.
[7] Water Quality for Better Health – TWAD Released Water book.
[8] Data mining Learning models and Algorithms for medical applications – White paper - Plamena Andreeva, Maya Dimibova, Petra Radeve
[9] Elementary Fuzzy Matrix Theory and Fuzzy Models for Social Scientists - W.B.Vasantha Kandasamy (e-book :http:mit.iitm.ac.in)
[10] Professionals statement calling for an End to water Fluoridation – Conference Report (www.fluoridealert.org)
[11] Analysis of Liver Disorder Using Data mining algorithms - Global Journal of computer science and Technology l.10 issue 14 (ver1.0) November 2010 page 48.
[12] The WEKA Data Mining Software: An Update, Peter Reutemann, Ian H. Witten, Pentaho Corporation, Department of Computer Science

AUTHOR'S PROFILE

Dr.R.Uma Rani received her Ph.D., Degree from Periyar University, Salem in the year 2006. She is a rank holder in M.C.A., from NIT, Trichy. She has published around 40 papers in reputed journals and national and international conferences. She has received the best paper award from VIT, Vellore , Tamil Nadu in an international conference. She was the PI for MRP funded by UGC. She has acted as resource person in various national and international conferences. She is currently guiding 5 Ph.D., scholars. She has guided 20 M.Phil., scholars and currently guiding 4 M.Phil., Scholars. Her areas of interest include information security, data mining, fuzzy logic and mobile computing.

T.Balasubramanian received his M.Sc computer Science in Jamal Mohamed College, Trichy under Bharathidasan university and Mphil Degree from Periyar University. Now persuing his Ph.D research under Bharathiar University, Coimbatore. Doing research under health care domain in Datamining applications. He published 6 research papers in various National, International conferences and 4 papers in various International journals.

GSM-Based Wireless Database Access For Food And Drug Administration And Control

Engr. Prof Hyacinth C. Inyiama
Electronic and Computer Engineering Dept
Nnamdi Azikiwe University
Awka, Nigeria

Engr. Mrs Lois Nwobodo
Electronic and Computer Engineering Dept
Nnamdi Azikiwe University
Awka, Nigeria

Engr. Dr. Mrs. Christiana C. Okezie
Electronic and Computer Engineering Dept
Nnamdi Azikiwe University
Awka, Nigeria

Engr. Mrs. Nkolika O. Nwazor
Research and Development Department
Electronic Development Institute
Awka, Nigeria

***Abstract*— GSM (Global system for mobile communication) based wireless database access for food and drug administration and control is a system that enables one to send a query to the database using the short messaging system (SMS) for information about a particular food or drug. It works in such a way that a user needs only send an SMS in order to obtain information about a particular drug produced by a pharmaceutical industry. The system then receives the SMS, interprets it and uses its contents to query the database to obtain the required information. The system then sends back the required information to the user if available; otherwise it informs the user of the unavailability of the required information. The application database is accessed by the mobile device through a serial port using AT commands. This system will be of great help to the National Agency for Food and Drug Administration and Control (NAFDAC), Nigeria in food and drug administration and control. It will help give every consumer access to information about any product he/she wants to purchase, and it will also help in curbing counterfeiting. NAFDAC can also use this system to send SMS alerts on banned products to users of mobile phones. Its application can be extended to match the needs of any envisaged user in closely related applications involving wireless access to a remote database.**

Keywords- Web; GSM; SMS; AT commands; Database; NAFDAC.

I. INTRODUCTION

The National Agency for Food and Drug Administration and Control (NAFDAC), established by Decree No. 15 of 1993 as amended is a Parastatal of the Federal Ministry of Health, with the mandate to regulate and control quality standards for Foods, Drugs, Cosmetics, Medical Devices, Chemicals, Detergents and packaged water imported, manufactured locally and distributed in Nigeria[1]. The importance of food and drugs to man and animal is very obvious. They need healthy food in order to grow and sustain life. The organs of the body may not function properly if exposed to unhealthy food. These situations of ill health provide the compelling need for drugs in order to modify the functioning of the body and restore it to a healthy state. To be acceptable, the drug must not be deleterious to the body but should rather lead to the restoration of normal life. In like manner, cosmetics should have no harmful effect on the body to which they are applied. It is the duty of all government to protect the health of the citizens, and in Nigeria, this is the responsibility of the Federal Ministry of Health [2].

The NAFDAC management has made intensive efforts towards making positive impacts in the lives of stakeholders (consumers and Dealers) through public enlightenment campaigns, education, persuasions and prosecution of defaulters.

II. PROBLEM STATEMENT

Counterfeiting is a global problem. Many goods moving through international commerce are counterfeited. Industry data show that 5-7% of world trade, valued at about US$280 billion is lost to counterfeiting. The pharmaceutical industry and the personal care products industry in Nigeria are riddled with counterfeits. Millions of dollars of counterfeit pharmaceuticals and personal care products are reported to move through various authorized and unauthorized channels. These (authorized and unauthorized) channels make it possible for counterfeits, expired, repackaged and relabeled products to be shipped internationally. Several criminal networks involved in drug faking and counterfeiting have evolved over the years. They include manufacturers, importers, distributors and retailers. Other collaborators are inspection agents, shipping and clearing agents and corrupt government officials of drug regulatory agencies, customs and police [3].

NAFDAC is established to do the following among other functions:

- Regulate and control the importation, exportation, manufacture, advertisement, distribution, sale and use of drugs, cosmetics, medical devices, bottled water and chemicals
- Conduct appropriate tests and ensure compliance with standard specifications designated and approved by the council for the effective control of quality of food,

drugs, cosmetics, medical devices, bottled water and chemicals and their raw materials as well as their production processes in factories and other establishments.

- Undertake inspection of imported food, drugs, cosmetics, medical devices, bottled water and chemicals and establish relevant quality assurance system, including certification of the production sites and of the regulated products.
- Undertake the registration of food, drugs, medical devices, bottled water and chemicals
- Collaborate with the National Law Enforcement Agency in measures to eradicate drug abuse in Nigeria

NAFDAC have succeeded in registering most food and certified okay by it. These products are issued registration numbers for easy identification. This has helped in reducing counterfeits. They have also gone a long way in sensitizing most Nigerians on the need for a product to be certified okay by them, but there is still the challenge of some products bearing fake NAFDAC registration number. Information about fake or substandard products banned from the Nigerian market can easily be accessed at the organization's web site. The awareness about such products is also created over the media. But there is the challenge of making this information available to those in areas without internet access or coverage. Moreover, there is the problem of low internet literacy. The problem of poor energy supply being faced by Nigeria makes it difficult for a greater proportion of the populace to have access to this information dispersed over the media.

Telecommunication has turned the world into a global village. It is a very cheap means of communication and is available in most parts of Nigeria. Technology is made more easily affordable if what is available is in a society is used to meet the immediate needs of that society. The challenges faced by other means of communication like the satellite (via internet), television, etc., led us to the introduction of this system that makes use of short messaging system (SMS) which is very cheap. Moreover, the GSM network over which it is sent is easily available in many parts of the country.

III. Schematic Overview Of The Main Components In A Gsm Network

The block diagram overview of a GSM network is shown in figure1.The GSM network consists mainly of the following functional parts [4]:

- MSC – the mobile service switching centre (MSC) is the core switching entity in the network.
- VLR – the visitor location register (VLR) contains subscriber data for subscribers registered in an MSC. Every MSC contains a VLR. Although MSC and VLR are individually addressable, they are always contained in one integrated node.
- GMSC – the gateway MSC (GMSC) is the switching entity that controls mobile terminating calls. When a call is established towards a GSM subscriber, a GMSC contacts the HLR of that subscriber, to obtain the address of the MSC where the subscriber is currently registered. That MSC address is used to route the call to the subscriber.
- HLR – the home location register (HLR) is the database that contains a subscription record for each subscriber of the network. A GSM subscriber is normally associated with a particular HLR. The HLR is responsible for the sending of subscription data to the VLR (during registration) or GMSC (during mobile terminating call handling).
- CN – the core network (CN) consists of, amongst other things, MSC(s), GMSC(s) and HLR(s). These entities are the main components for call handling and subscriber management.[5]
- BSS – the base station system (BSS) is composed of one or more base station controllers (BSC) and one or more base transceiver stations (BTS).
- MOBILE STATION (MS) – The mobile station is the user equipment in GSM. The MS is what the user can see of the GSM system. The station consists of two entities, the Mobile Equipment (the phone itself), and the Subscriber Identity Module (SIM), in form of a smart card contained inside the phone.

A. Signaling in GSM

The various entities in the GSM network are connected to one another through signaling networks. Signaling is used for subscriber mobility, subscriber registration, call establishment, etc. The connections to the various entities are known as 'reference points'. Examples as seen from figure1 include the following [6]:

- A interface – the connection between MSC and BSC
- Abis interface – the connection between BSC and BTS
- C interface- the connection between GMSC and HLR
- D interface – the connection between MSC and HLR
- Um interface – the radio connection between MS and BTS
- ISDN user part (ISUP) – ISUP is the protocol for establishing and releasing circuit switched calls.

B. Mobile station

The Mobile Station, i.e. the GSM handset, is logically built up from the following components:

- Mobile equipment (ME) – this is the GSM terminal, excluding the SIM card;
- Subscriber identification module (SIM) – this is the chip embedded in the SIM card that identifies a subscriber of a GSM network; the SIM is embedded in the SIM card.

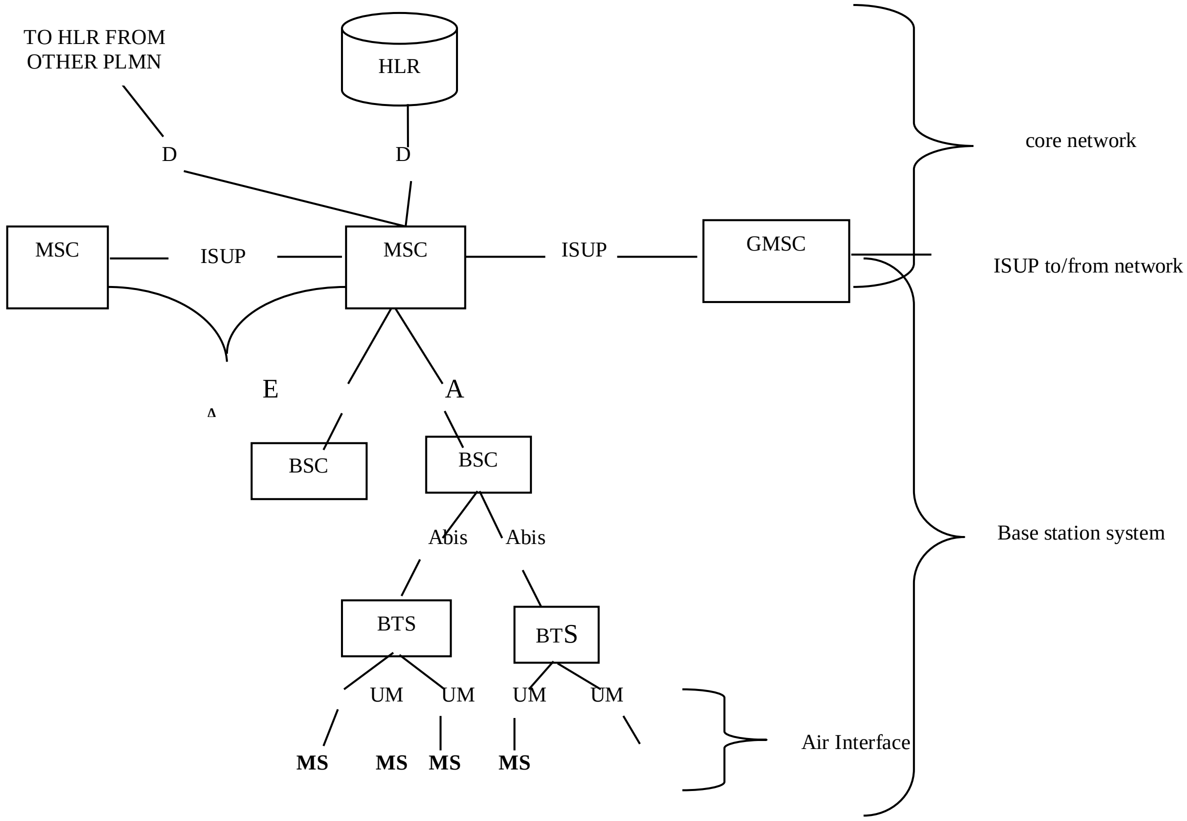


Figure 1. Overview of the GSM network

When the SIM card is inserted in the ME, the subscriber may register with a GSM network. The ME is now effectively personalized for this GSM subscriber.

C. Short message service system (SMS)

SMS is an acronym of Short Message Service, it is a communication service component a mobile communication systems, usually in a text form,[4] it uses standard rules which allows the exchange of short text messages between mobile phone devices, it could be from computer to phone or vice versa. SMS has used on modern handsets originated from radio telegraphy in radio memo pagers using standardized phone protocols. It was later defined as part of the Global System for Mobile Communications (GSM) series of standards in 1985 as a means of sending messages not more than 160 characters in a phone page, to and from GSM mobile handsets. GSM defines the Short Message Service - Cell Broadcast (SMS-CB), which allows messages (advertising, public information, etc.) to be broadcast to all mobile users in a specified geographical area. Messages are sent to a Short message service center (SMSC) which provides a "store and forward" mechanism. It attempts to send messages to the SMSC's recipients. If a recipient is not reachable, the SMSC queues the message for later retry. Some SMSCs also provide a "forward and forget" option where transmission is tried only once.

Both mobile terminated (MT, for messages sent to a mobile handset) and mobile originating (MO, for those sent from the mobile handset) operations are supported. Message delivery is usually "best effort", so there are no guarantees that a message will actually be delivered to its recipient, but delay or complete loss of a message is uncommon

IV. OPERATION OF THE EXISTING SYSTEM

It is the duty of NAFDAC to register new drug, cosmetics, bottled water or drinks produced in Nigeria or imported from other countries which has successfully passed its tests. When these products are registered, a NAFDAC registration number is issued for the product which is written on the product's pack. Some manufacturers push their substandard products to the market without going through this process. NAFDAC officials physically visit some markets and pharmaceutical stores to ensure that the products being sold to consumers are approved by it. They take the record of these products and go back to their office to crosscheck if it is in their records.

V. THE PROPOSED SYSTEM

The proposed system is a system that will help users to access information on an item effectively, without wasting energy and time as far as there is a GSM network covering that area. The proposed system is aimed at achieving many things. It is an easy way of detecting fake products, or unregistered product. It increases productivity and ensures adequate management, ease of update and maintenance of operation, speed optimization and reduces the use of manual processing. It is reliable. System design in most cases are based on modularization, be it software or hardware systems. The project design would be based on modularization which is the divide and conquer method.

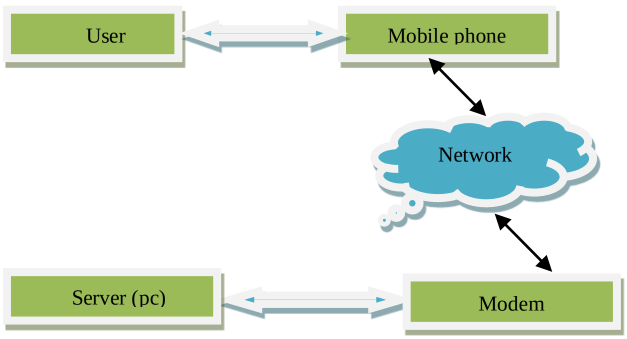


Figure 2. Diagram of the new system

It allows for easy and effective design as the whole system is broken down into smaller units or modules, and design for each module is carried out independently. Modularization facilitates easy troubleshooting and correctness or easy debugging (in the case of software development).

VI. System Specification

This is the specification for the entire system. That is the various components integrated in the construction of a project as shown in Figure2. The processes of operation of the various parts of the system are as described hereunder.

The components of the system are:

- User's phone and internet modem.
- A server (Personal Computer) running on Windows Operating System
- Protocol distribution unit (PDU)
- Database management system and Visual Basic Programming Environment

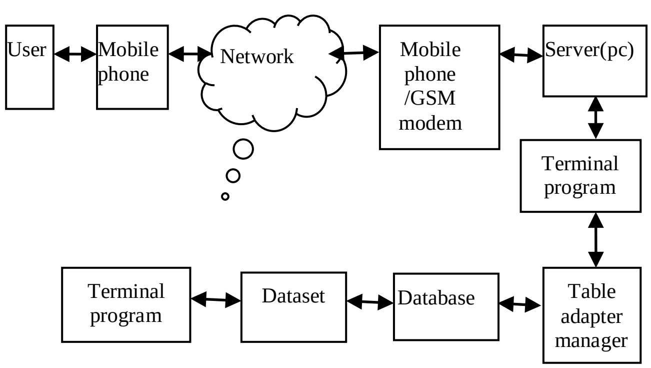


Figure 3. Block diagram representation of the operation of the system

Figure 3 shows the user who keys in the message and the message destination in a chosen format. The message format is of the form:

nafdac-regno-04-0437

Note:

"nafdac" is the name of the table the system will query

"regno" is a search criteria

"04-0437" is ''the search with''.

The message passes through the network, an object of the network known as the short message service system sends it to the owner, this object does the store and forward process, it stores the message if the recipient is not available on the network and forwards it if the recipient is available. The recipient in this system is the modem.

A. Modem

This is a device that modulates and demodulates, converting from analogue signal to digital signal, and vice versa. It facilitates easy transmission of signals from the mobile device to the server.

B. Server

It houses almost all the software components used in this project.

C. Terminal program

This instructs the computer on the next action to take when a message is received; it can be written in languages like Visual Basic, C#, C++ and many others. For this system, visual basic.net was used.

D. Table Adapter Manage

Table Adapters provide communication between your application and a database. More specifically, a Table Adapter connects to a database, executes queries or stored procedures, and either returns a new data table populated with the returned data or fills an existing Data Table with the returned data. Table Adapters are also used to send updated data from your application back to the database.

E. Database

A database consists of an organized collection of data for one or more multiple uses.[7]

F. Dataset

The Data Set, which is an in-memory cache of data retrieved from a data source, it is a major component of the ADO.net architecture. The dataset is a memory where the Table Adapter keeps the retrieved data from the data source, when instructed by the terminal program.

VII. Software Subsystems Design

A. AT (Attention) Command

The system uses AT COMMAND to establish communication with the modem. The AT commands that were used in this project are;

- AT+CGMS - Sends SMS Messages.
- AT+CMGD – to clear the SMS receiving memory location in the GSM modem or message storage.

- ATEO – makes the phone not to echo commands.
- AT+CPMS - used to assign the message storage to read from, delete and write to.
- AT+CMGR – read SMS messages from storage.
- AT+CMGF – tells the modem to operate in test mode.

B. Syntax for Sending SMS

The syntax for sending SMS to NAFDAC is shown below

nafdac-drugname-puriton
nafdac-sachetwater-aquarapha
nafdac-regno-04-0437

C. The SQL Command of NAFDAC Table

FillGetdata( )

a) fillBy- drug name,GetDataby-drug name(drug name)
then the SQL command is thus
select drug name
from nafdac
where drug name =?

b) fillBy-sachet water,GetDataby- sachet water (sachet water)

then the SQL command is thus
select sachet water
from nafdac
where satchet water =?

c) fillBy- reg no,GetDataby-reg no(reg no)

then the SQL command is thus
select reg no
from library
where reg no =?

VIII. CONCLUSION

The SMS based NAFDAC Registration Exercise is a new innovation that unveils the features of SMS in GSM applications. This work is a new development that will help curb the menace of counterfeit drugs in our society. This work can be easily adapted to match the need of any envisaged user.

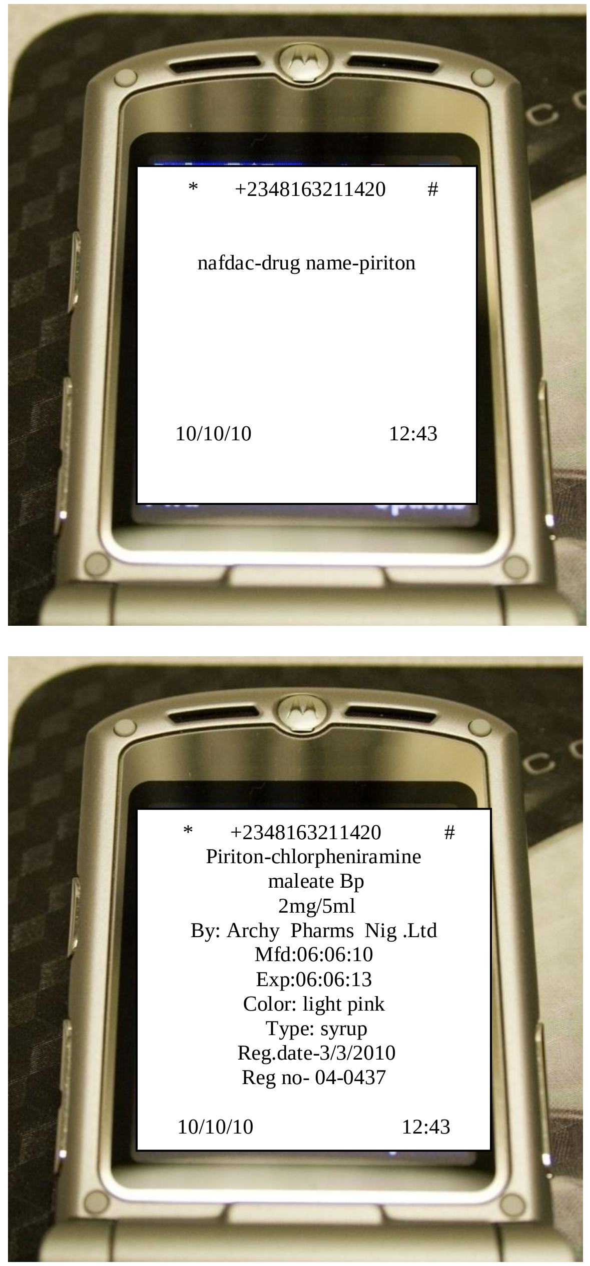


Figure 4. The Project Input and the Output

REFERENCES

[1] http://www.nafdacnigeria.org. retrieved 22nd February 2011

[2] About NAFDAC. http://www.nafdacnigeria.org/about.html. retrieved 22nd february 2011

[3] Global Trends. http://www.nafdacnigeria.org/globaltrends.htm retrieved 22nd february 2011.

[4] S. Guthery, M. Cronin, Mobile Application Development with SMS and the SIM Toolkit, McGraw-Hill Professional,1997. Inc(2009) pp.5-10

[5] F. Hillebrand, F. Trosby, K. Holley & I. Harrris, Short Message Service, Trafford Publishing, 2010, Page 19.

[6] J.Eberspaecher, H. J. Voegel, C. Bettstetter, GSM Switching, Services, And Protocols, John Wiley & Sons,2001, Pg 254-267

[7] C.J Date,An Introduction to database systems, Wiley, 1998, pg 1.

Novel Techniques for Fair Rate Control in Wireless Mesh Networks

Mohiuddin Ahmed
Computer Science and Engineering Department
Islamic University of Technology
Gazipur, Bangladesh

K.M.Arifur Rahman
Electrical & Electronic Engineering Department
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh

***Abstract*— IEEE 802.11 based wireless mesh networks can exhibit severe fairness problem by distributing throughput among different flows originated from different nodes. Congestion control, Throughput, Fairness are the important factors to be considered in any wireless network. Flows originating from nodes that directly communicate with the gateway get higher throughput. On the other hand, flows originating from two or more hops away get very little throughput. For this reason a distributed fair scheduling is an ideal candidate for fair utilization of gateway's resources (i.e., bandwidth, airtime) and thereby achieving fairness among contending flows in WMNs. There are numerous solution for aforementioned factors in wireless mesh network. We figured out some problems of few existing solutions and integrated to give the solution for those problems. We considered neighborhood phenomenon, airtime allocation and elastic rate control to design a novel technique to achieve fair rate control in wireless mesh network. And finally we introduce distributed fair scheduling to get fairness in mesh network.**

Keywords-Wireless mesh network; Network Throughput; Congestion Control; Fairness; Airtime Allocation; Neighbourhood Phenomenon; Gateway. (key words)

I. INTRODUCTION

Wireless Mesh Networks [1] (WMNs) are envisioned to replace the wired backbone by multi hop wireless network. The shared nature of the wireless medium, unpredictable link quality, and existence of hidden terminals pose a number challenges for WMNs. The poor performance of TCP in wireless networks demands efficient congestion control in WMNs. Furthermore, efficient rate control is required for fair bandwidth sharing and optimum network throughput. We work toward designing a distributed fair rate control mechanism in WMNs.

As various wireless networks evolve into the next generation to provide better services, a key technology, wireless mesh networks (WMNs), has emerged recently. In WMNs, nodes are comprised of mesh routers and mesh clients. Each node operates not only as a host but also as a router, forwarding packets on behalf of other nodes that may not be within direct wireless transmission range of their destinations. A WMN is dynamically self- organized and self-configured, with the nodes in the network automatically establishing and maintaining mesh connectivity among. This feature brings many advantages to WMNs such as low up-front cost, easy network maintenance, robustness, and reliable service coverage.

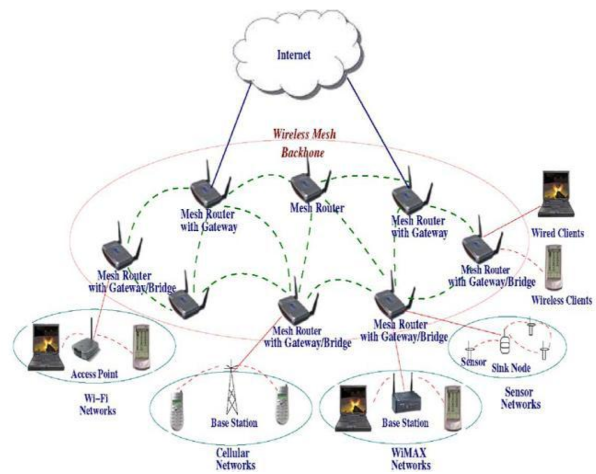


Figure 1. Wireless Mesh Network

WMNs consist of two types of nodes: mesh routers and mesh clients.Fig.1 shows a basic wireless mesh network provided by Other than the routing capability for gateway/repeater functions as in a conventional wireless router, a wireless mesh router contains additional routing functions to support mesh networking. To further improve the flexibility of mesh networking, a mesh router is usually equipped with multiple wireless interfaces. Compared with a conventional wireless router, a wireless mesh router can achieve the same coverage with much lower transmission power through multi hop communications. Optionally, the medium access control (MAC) protocol in a mesh router is enhanced with better scalability in a multi-hop mesh environment.

A major challenge of IEEE 802.11 based wireless mesh network is to provide fair rate allocation. In multi-hop WMN nodes that are directly connected to the gateway get higher priority for their aggregate flow than the nodes that are two or more hops away from the gateway. This results in severe unfairness and causes starvation of flows travelling multiple hops. So, a fair rate control mechanism is needed to control the high priority traffic and to allow multi hop nodes to increase their rate.

Unfairness is considered as the consequence of one of the pressing issues of hidden node problem. Hidden terminal interference is caused by the simultaneous transmission of two node stations that cannot hear each other, but are both received by the same destination station. It is said that RTS/CTS can solve the problem but still there exists the problems of low throughput and increased average packet latency and poor fairness.

In this paper, we analyzed the solutions from three research papers [2,3,4] and then provided our approaches to eliminate those problems. Our contribution includes the elimination of stealing effect, maximum airtime allocation and mini backup. And thus our approaches are capable of ensuring fair rate control mechanism in wireless mesh network.

II. Related Work

Wireless mesh network has become an attractive research area now-a-days. Significant research has been done to investigate the rate and congestion control of wireless multi hop networks.

In GAP [2] there is a rate control mechanism which limits rate of single hop nodes by enforcing a utilization threshold.

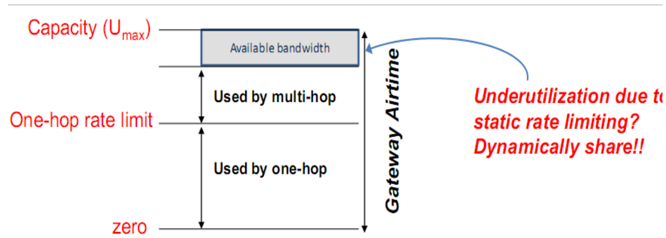


Figure 2. Gateway Airtime Partitioning

They considered traffic only to and from gateway .Thus their proposed algorithm only tries to provide fair share for all spatially disadvantaged nodes around the gateway, not in anywhere else. So this algorithm is not well suited in such network where local traffic exists.

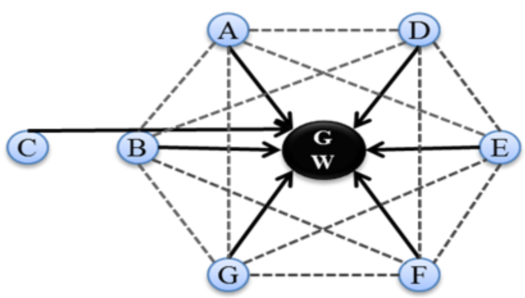


Figure 3. Multiple Competing Flows

Their considered network holds single hop node as leaf node. That creates a hidden terminal problem.

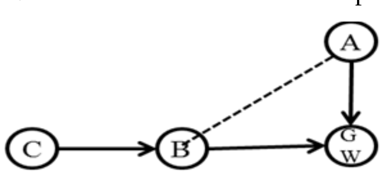


Figure 4. Single competing flow

Considering Figure. 4, whenever A gains the network then B will not get chance and thus nodes connected with B will be congested and A will get high priority always among its one hop competing flows. As usual data transmission will require more time and thus the congested nodes data transmission will lead to a worse situation.

In addition, IFA [5] and [6] proposed link-layer rate control mechanism to solve TCP fairness. In IFA the authors studied per-TAP fairness and end-to-end performance in WMNs (Multi-hop wireless backhaul networks).They propose an inter-TAP fairness algorithm that aims to achieve per-TAP fairness without modifying the TCP protocol. Here, nodes explicitly exchange information and all nodes calculate their fair shares using that information. Also algorithm proposed by Raniwala et al. [6] works away from the unstable MAC layer.

Alternatively, another class of work to rate and congestion control is done by Rangwala et al. where wireless congestion is considered as neighborhood phenomenon [4] instead of considering as per flow problem.

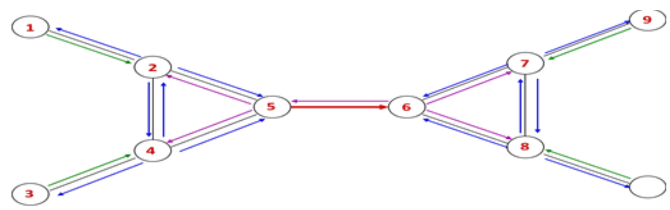


Figure 5. Neighborhood Phenomenon

According to these schemes when a link is congested then all the correspondent neighbor link has to reduce their transmission rate to solve the congestion of that link. As a consequence the neighbor link's neighbor also has to limit their rate although those links are not supposed to reduce their rate because their transmission doesn't affect the root congested link. And thus the network utilization and throughput is reduced. They consider mesh network as independent network and Converging & diverging of flows are absent here.

Furthermore, Ramesh Govindan et al. focused on airtime allocation [3] to solve the drawbacks of [4]. In [3] they assigned minimum airtime limit to each of the active link in the network. The calculation of the assigned airtime limit is based on the number of active neighborhood link. These assigned airtime limit converges very slowly from minimum to optimal. Moreover, the calculation of the airtime limit is sophisticated due to its control overhead.

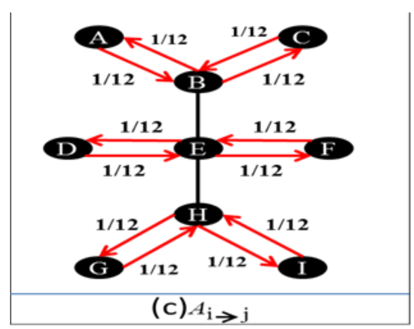


Figure 6. Airtime Allocation

In MLM-FQ [8] local schedulers self-coordinate their scheduling decisions and collectively achieve fair bandwidth sharing. But it does not allow spatial channel reuse of network resources. Same authors then proposed EMLM-FQ to further improve the spatial channel reuse. Moreover both the scheduling techniques are sender initiated thus have good probability of collisions.

III. Problem Description

A. *Gateway Airtime Partitioning[2]*

Define abbreviations and acronyms the first time they are used in the text, even in this paper [2] they considered traffic only to and from gateway & thus their proposed algorithm only tries to provide fair share for all spatially disadvantaged nodes around the gateway, not in anywhere else. So this algorithm is not well suited in such network where local traffic exists.

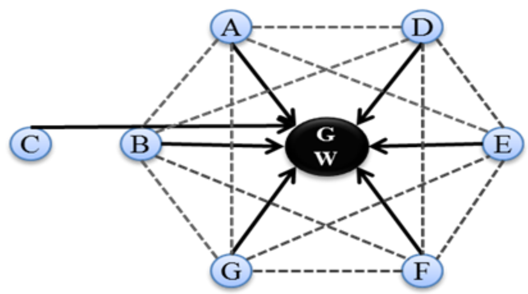


Figure 7. Two-hop node competing with single-hop node & Dotted lines indicate that connected nodes are within sensing range

Their considered network holds single hop node as leaf node. That creates a hidden terminal problem (Stealing Effect). Previous papers bypassed the stealing effect scenario & no solution was provided.

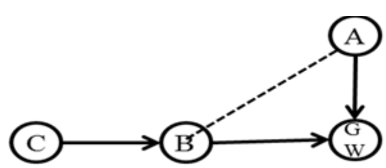


Figure 8. Oversimplified Network Model

The single hope leaf node connected to gateway doesn't sense the other node as it is out of the interference range of each other so one node will be able to transmit and other will get chance when one finishes.

Considering figure.8, whenever A gains the network then B will not get chance and thus nodes connected with B will be congested and A will get high priority always among its one hop competing flows. As usually data transmission will require more time and thus the congested nodes data transmission will lead to a worse situation. Their considered network model is oversimplified as single hop nodes are leaf node. It is very usual that there might be some other scenarios of network where single hop node is not the leaf node. If we consider the different scenarios like mentioned then the oversimplified problem can be solved.

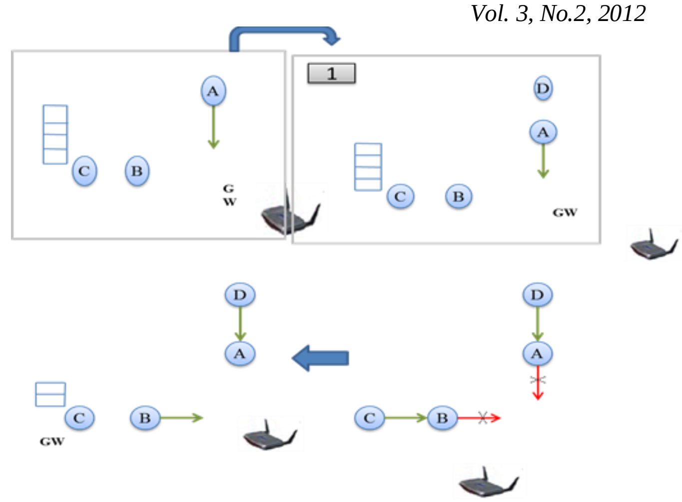


Figure 9. Modified Topology

Now we see that in the above figure 9. Whenever D will gain the network A have to stop sending data to gateway and thus creates an opportunity for other competing flows who were congested earlier. Once the disadvantaged nodes get chance others will not be able gain the network resource unless they finish. Thus short term fairness might not be possible but long term fairness is easily achieved. But the scenario is not beyond Stealing Effect problem. Stealing effect might occur in any depth of the network among two hop away nodes.

B. *Transparent Airtime Allocation[3]*

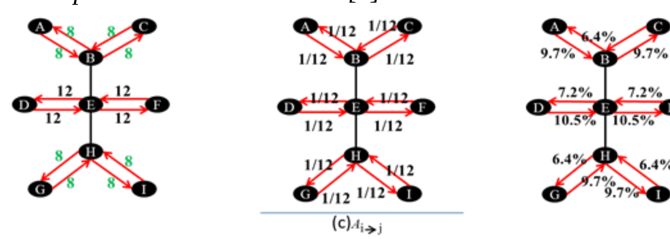


Figure 10. Airtime Allocation

In this paper [3] the writer focused on airtime allocation to solve the drawbacks of the previous paper e.g. Network Underutilization. They assign minimum airtime limit to each of the active link in the network. The calculation of the assigned airtime limit is based on the number of active neighborhood link. These assigned airtime limit converges very slowly from minimum to optimal. Moreover the calculation of the airtime limit is sophisticated due to its control overhead.

C. *Understanding Congestion Control[4] Understanding Congestion Control[4]*

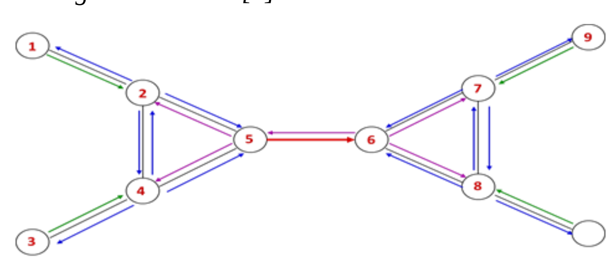


Figure 11. Considered Network Topology.

They considered congestion control as neighborhood phenomenon. When a link is congested then all the correspondent neighbor link has to reduce their transmission rate to solve the congestion of that link.

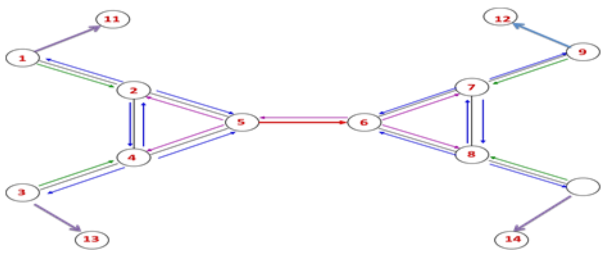


Figure 12. Different Network Topology

As a consequence the neighbor link's neighbor also has to limit their rate although those links are not supposed to reduce their rate because their transmission doesn't affect the root congested link. And thus the network utilization and throughput is reduced. They consider mesh network as independent network and Converging & diverging of flows are absent here. Gist is they fail to provide proper network utilization.

IV. Proposed Techniques

Our solution is comprised of three main components. We first find the maximum air time limit by constructing the compatibility graph, which contains the links that can be activated at the same time and define different possible cliques (see Subsection A),then we use a distributed fair scheduling mechanism by Start Time Fair Queuing (see Subsection B). Finally, we solve the stealing effect by receiver initiated CTS mechanism (see Subsection C).

A. *Assignment of Maximum Airtime Limit*

In paper [3], airtime was assigned to each link dynamically from minimum to maximum (discussed earlier).Our challenge is to assign maximum airtime limit to each active link of the network. We define a procedure to compute the airtime-limit, by constructing compatibility graph depending on the neighborhood scenario of the network and find different possible cliques.

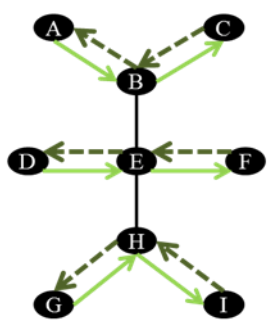


Figure 13. Topology of the network

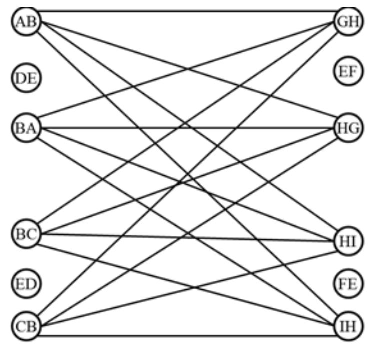


Figure 14. Compatibility graph of the network presented in figure 13

Here in the compatibility graph, each vertex represents single active link {i, j} of the network and the edge between two vertices tell that these two links can be activated at the same time.

Now from the above compatibility graph we can construct the set of all possible cliques for the corresponding network, where a clique is a set of all links that can be activated simultaneously. A clique is a complete sub graph. From the figure 14 we get 8 sets of cliques like below-

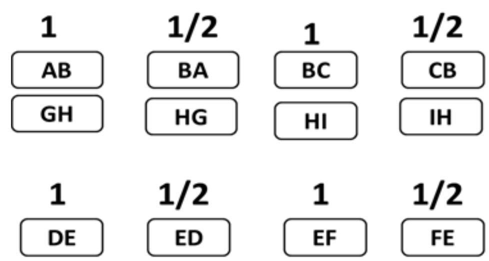


Figure 15. Clique sets for figure 14

- Weight of Data traffic = 1
- Weight of ACK traffic = 1/2
- 1+1/2+1+1/2+1+1/2+1+1/2 = 6
- So, 6 different time slots are required

So, from the above scenario it is clear that, maximum 1/6 of the total airtime limit will be assigned to each clique set and we can show it like below-

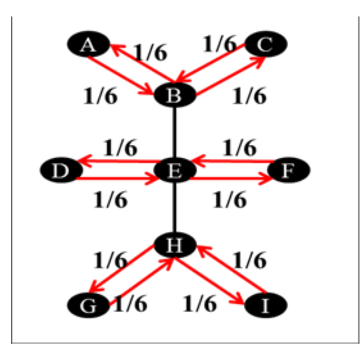


Figure 16. Maximum Airtime Limit to each flow

Thus our second challenge is solved by assigning maximum ¼ of the airtime limit to each flow of the network.

B. Start Time Fair Queuing

We propose a distributed fair scheduling mechanism- Start Time Fair Queueing to achieve fairness and to minimize collision in the mesh network. Here each node will maintain a table where there are two fields - start tag and finish tag. Every node shares their start tag. When a node transmits it include the finish tag of next packet in its header.

For example- node 1 has following start tags and finish tags for their packets-

Start Tag	Finish Tag
100	130
130	150
150	190

TABLE 1. START TAG AND FINISH TAG FOR A PACKET

Here, first packet has start tag 100, packet size 30. So its finish tag is 130. Second packet has start tag 130, packet size 20. So its finish tag is 150. Third packet has start tag 150, packet size 40. So its finish tag is 190. When node 1 transmits packet 1 to other node (i.e., node 2) it include the finish tag of packet 2 (i.e., here 150) in its header. All other nodes of node 1's transmission range will overhear this information and know its next packet's finish time. In this way each node will know the finish time of next packet of all other nodes belong to its transmission range. This information will be in the sorted order in the table.

Node with the lowest finish time above in the table will transmit next. Whenever any node finishes its transmission its finish time will be higher immediately and it will not transmit any packet for long time. Next node with the lowest finish time above in the table will transmit and update its finish tag value. This process goes on until all the packets are transmitted.

One problem of this mechanism is that –if finish tags of two node become equal, then collision occurs. To solve this problem we propose another mechanism – Mini Backup, based on maximum airtime limit.

In mini backup we prioritize a flow based on the maximum airtime limit that we proposed in previous section 4.1. Here the flow that has already done more airtime utilization will get less priority and flow that has done less airtime utilization will get higher priority. But still each node may experience collision due to stealing effect caused by hidden node and may starve for long time. For this we need to solve the stealing effect.

C. Elimination of the Stealing Effect

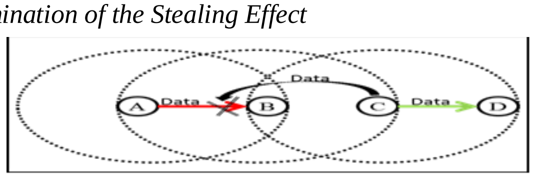


Figure 17. Hidden Node Problem

Here, node A will try several times to send its data and each time the data packets sent by C could collide with the data packets sent by A. Collision occur due to contention with hidden nodes(detail of this problem is discussed in section) and node A will starve.

We introduce receiver initiated CTS mechanism here to solve the above problem. In this mechanism, receiver (i.e., here node B) takes the responsibility to send CTS periodically to all the nodes of its transmission range (i.e., here node A and C) when it senses a collision within its transmission range.

So, after getting CTS node C will stop transmitting data to node D and node A will get the chance to transmit data. Thus starvation of node A will be removed.

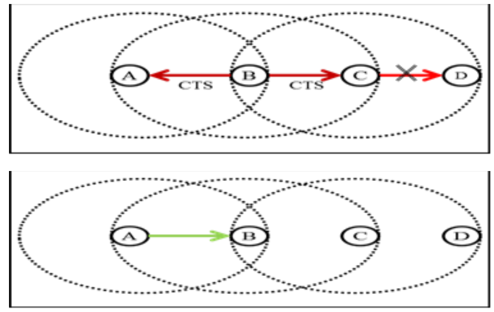


Figure 18. Solved scenario of Hidden Node Problem

V. CONCLUSION

In our research paper, we integrated the ideas from different research works and proposed some new approaches for fair rate control. Basically, we proposed new techniques for maximum airtime allocation, elimination of stealing effect and start time fair queuing. All these approaches are efficient enough to ensure fairness in wireless mesh network. In future, we will work extensively on these issues and would like to find out more efficient solutions along with real time simulation results.

REFERENCES

[1] Ian F. Akyildiz , Xudong Wang , Weilin Wang, Wireless mesh networks: a survey, Computer Networks (2005)

[2] Vincenzo Mancuso, Omer Gurewitz, Ahmed Khattab and Edward W. Knightly, Elastic Rate Limiting for Spatially Biased Wireless Mesh Networks, ACM MobiCom (2010).

[3] Ki-Young Jang, Konstantinos Psounis, Ramesh Govindan, Simple Yet Efficient, Transparent Airtime Allocation for TCP in Wireless Mesh Networks,CoNEXT (2010)

[4] RANGWALA, S., JINDAL, A., JANG, K.-Y., PSOUNIS, K.,AND GOVINDAN,R.Understanding congestion control inmulti-hop wireless mesh networks. In Proc. of ACM MobiCom (2008).

[5] V. Gambiroza, B. Sadeghi, and E. W. Knightly, "End-to-End Performance and Fairness in Multi-hop Wireless Backhaul Networks," in Proceedings of ACM MOBICOM, Philadelphia, PA, USA, Sept. 2004.

[6] A. Raniwala, P. De, S. Sharma, R. Krishnan, and T. c. Chiueh, "End-to-end flow fairness over ieee 802.11-based wireless mesh networks," in Proceedings of IEEE INFOCOM,2007

[7] Haiyun Luo, Jerry Cheng, and Songwu Lu, "Self-Coordinating Localized Fair Queueing in Wireless Ad Hoc Networks",IEEE ToM,04

AUTHORS PROFILE

Mohiuddin Ahmed has achieved Bachelor of Science degree in Computer Science and Information Technology from Islamic University of Technology. Research interest includes Human Computer Interaction, Cloud Computing, Wireless Networks, Artifical Intelligence.

K.M.Arifur Rahman is a final year undergraduate student in Electrical and Electronic Engineering department in Bangladesh University of Engineering and Technology. Research interest includes Genetic Algorithm, Ubiquitous Computing, Wireless Network and Image Processing.